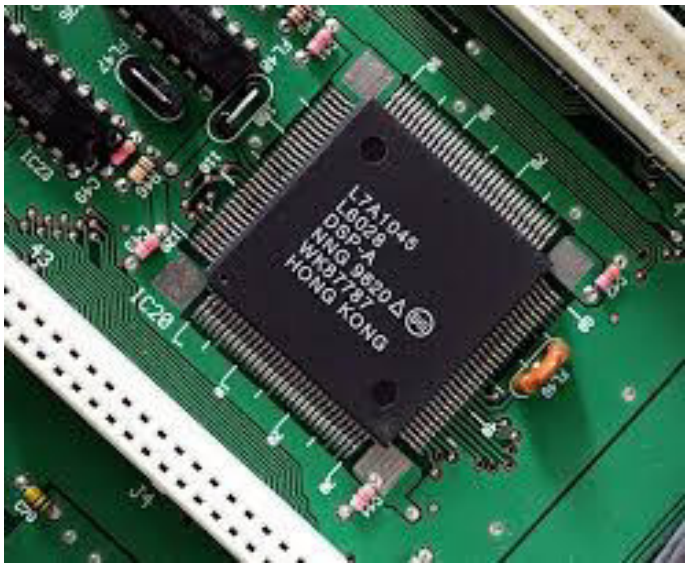
الجمهوريــة الجزائريــة الديمقراطيــة الشعبيــة

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

Faculty of Technology-University M'Hamed Bougara of Boumerdes, Algeria

ncasee2024@univ-boumerdes.dz
https://ncasee24.univ-boumerdes.dz/

NCASEE'24-Conference



Proceedings of the

1st National Conference of Advanced Systems in Electrical Engineering (NCASEE'24)

EDITORS

Aimad BOUDOUDA, Univ. Boumerdes, Algeria	**Abderrezak AIBECHE**, Univ. Boumerdes, Algeria
Hamza AKROUM, Univ. Boumerdes, Algeria	**Samia BELKACEM**, Univ. Boumerdes, Algeria
Karim BAICHE, Univ. Boumerdes, Algeria	**Youcef TABET**, Univ. Boumerdes, Algeria
Noureddine MESSAOUDI, Univ. Boumerdes, Algeria	**Mohammed AMMAR**, Univ. Boumerdes, Algeria
Yassine MERAIHI, Univ. Boumerdes, Algeria	**Khaled HARRAR**, Univ. Boumerdes, Algeria
Aimad AHRICHE, Univ. Boumerdes, Algeria	**Dalila Acheli**, Univ. Boumerdes, Algeria

ISBN: 978-9969-9733-5-8

9 789969 973358

Dépôt légal: 9901-2024

Edition



Faculty of Technology-University M'Hamed Bougara of Boumerdes, Algeria

The 1st National Conference of Advanced Systems in Electrical Engineering (NCASEE'24)

December 9th, 2024

ISBN: 978-9969-9733-5-8

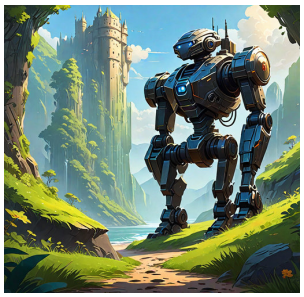
Preface NCASEE'24

The **1st National Conference on Advanced Systems in Electrical Engineering (NCASEE24)** was held on December 9, 2024, at the Faculty of Technology, University M'Hamed Bougara of Boumerdes, Algeria. This conference aimed to showcase new advancements and research findings in the fields of electrical engineering and its applications.

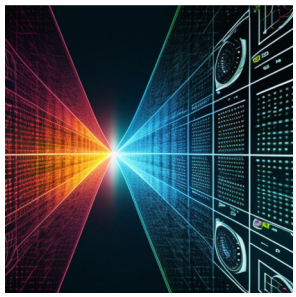
NCASEE24 provided a valuable platform for researchers, engineers, and scientists to interact and present their cutting-edge research. It offered a unique opportunity for participants to exchange ideas, demonstrate innovations, and foster scientific collaboration. Over 170 papers were submitted to the conference's scientific committee, of which 144 papers and posters were selected for presentation during two oral sessions, two poster sessions, two online sessions and two plenary sessions. The presentations were conducted in both in-person and online formats.

This conference served as a significant venue for sharing state-of-the-art research, facilitating technical discussions among attendees, and highlighting advancements across all areas of electrical engineering.

NCASEE24 covered a broad range of topics within the conference tracks, including, but not limited to:



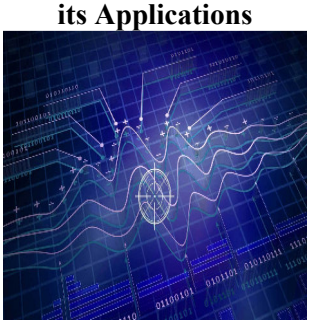
T1 : Artificial Intelligence and its Applications



T2 : Pattern Recognition & Computer Vision



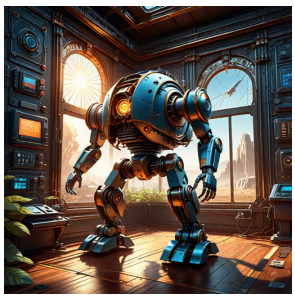
T3 : Radar and Remote Sensing



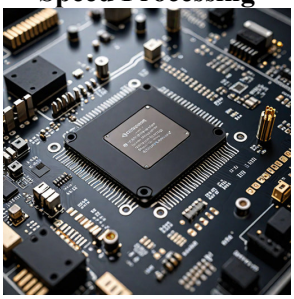
T4 : Signal, Image, Video, and Speed Processing



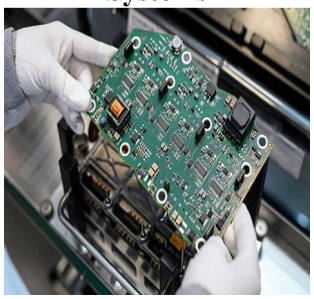
T5 : Telecommunication Systems



T6 : Advanced Control Systems, Automation & Robotics



T7 : Circuit and Embedded Systems



T8 : Power Electronics and Electrical Drives



T9 : Smart Grids Technologies & Applications



T10 : Transmission and Distribution



T11 : Antenna and Wave Propagation



T12 : E-Health Applications and Systems



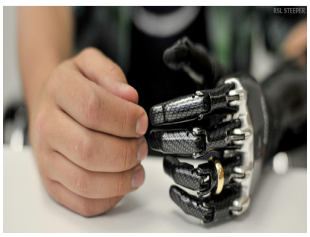
T13 : Wireless Communication



T14 : Bioinformatics and Instrumentation



T15 : Embedded and IoT Systems



T16 : Biomedical Engineering



T17 : Electrical Vehicls



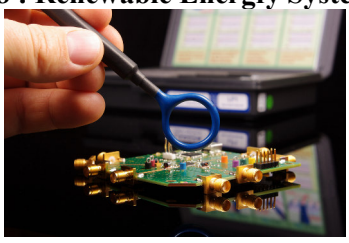
T18 : Renewable Energfy Systems



T19 : Performance and QoS



T20 : Wireless Power Transfer



T21 : Electromagnetic Compatibility-EMC

We take this opportunity to express our gratitude to the organizing and scientific committees, as well as our dedicated volunteers, for their exceptional efforts in making this event a success. We also extend our sincere appreciation to the Faculty of Technology, University M'Hamed Bougara of Boumerdes, for their invaluable support in hosting and facilitating this conference. A special thanks goes to **Professor SAIDI Mohamed** for his steadfast backing of this conference.

Our heartfelt thanks go to the invited speakers, **Dr. Moussa HAMADACHE** and **Professor Madjid KIDOUCHE**, for their outstanding contributions and remarkable expertise in their respective fields of research.

Finally, we are deeply grateful to all participants who enriched the event through their active involvement, stimulating discussions, and fruitful exchanges, which benefited everyone involved.

Thank you for being a part of **NCASEE24!**

Conference General Chairs
Dr. Aimad BOUDOUDA
Dr. Abderrezak AIBECHE

Honorary General Chairs

Pr. Mostepha YAHI, Rector of Univ. Boumerdes, Algeria
Pr. Mohamed SAIDI, Dean of the Faculty of Technology, Univ. Boumerdes, Algeria

President of the scientific committee

Dr. BOUDOUDA Aimad, Univ. Boumerdes, Algeria

Conference General Chairs

Dr. BOUDOUDA Aimad, Univ. Boumerdes, Algeria
Dr. AIBECHE Abderezzak, Univ. Boumerdes, Algeria
Pr. AHRICHE Aimad, Univ. Boumerdes, Algeria.
Pr. AMMAR Mohammed, Univ. Boumerdes, Algeria
Pr. MESSAOUDI Noureddine, Univ. Boumerdes, Algeria

Conference General Co-Chairs

Dr. BELKACEM Samia, Univ. Boumerdes, Algeria
Dr. TABET Youcef, Univ. Boumerdes, Algeria
Pr. HARRAR Khaled, Univ. Boumerdes, Algeria
Pr. FELLAG Sid Ali, Univ. Boumerdes, Algeria

Organizing Committee Chairs

Dr. AKROUM Hamza, Univ. Boumerdes, Algeria
Dr. BAICHE Karim, Univ. Boumerdes, Algeria
Pr. MERAIHI Yassine, Univ. Boumerdes, Algeria

Organizing Committee

Miss. Sihem TIRECHE, Univ. Boumerdes, Algeria
Dr. Amel BOUSTIL, Univ. Boumerdes, Algeria
Dr. Dahbia AMROUCHE, Univ. Boumerdes, Algeria
Dr. Belynda FARES, Univ. Boumerdes, Algeria
Dr. Fatima BENNIA, Univ. Boumerdes, Algeria
Miss. Faiza IBAOUNI, Univ. Boumerdes, Algeria
M. Mohamed Yanis HIOU, Univ. Boumerdes, Algeria
M. Ibrahim HAMIDAT, Univ. Boumerdes, Algeria
Miss. Sofia AZZOUGUI, Univ. Boumerdes, Algeria
Miss. Siham MOUSSAOUI, Univ. Boumerdes, Algeria

Scientific Committee

Pr. ACHELI Dalila, Univ. Boumerdes, Algeria
Pr. AHRICHE Aimad, Univ. Boumerdes, Algeria
Pr. AMMAR Mohammed, Univ. Boumerdes, Algeria
Pr. BENLATRECHE Mohamed Salah, Univ. Mila, Algeria
Pr. BENMOUSSA Yahia, Univ. Boumerdes, Algeria
Pr. BENTARZI Hamid, Univ. Boumerdes, Algeria
Pr. BOUDJERDA Nasserdine, Univ. Jijel, Algeria
Pr. BOUKRA Abdelmadjid, Univ. Bab Ezzouar, Algeria
Pr. BOUSHAKI Razika, Univ. Boumerdes, Algeria
Pr. BOUZIDA Ahcene, Univ Bouira, Algeria
Pr. CHALLAL Mouloud, Univ. Boumerdes, Algeria
Pr. CHELIHI Abdelghani, Univ. Constantine, Algeria

Pr. CHELLIL Ahmed, Univ. Boumerdes, Algeria
Pr. CHERIFI Dalila, Univ. Boumerdes, Algeria
Pr. DAAMOUCHE Abdelhamid, Univ. Boumerdes, Algeria
Pr. DAOUI Abdelhakim, Univ. Boumerdes, Algeria
Pr. FELLAG Sid Ali, Univ. Boumerdes, Algeria
Pr. FERGUENE Farid, Univ. Bab Ezzouar, Algeria
Pr. HAMADOUCHE M'Hamed, Univ. Boumerdes, Algeria
Pr. HARKAT Mohammed Faouzi, Univ. Annaba, Algeria
Pr. HARRAR Khaled, Univ. Boumerdes, Algeria
Pr. HENTOUT Abdelfettah, ENSTA, Algeria
Pr. HIDOUCI Walid Khaled, ESI, Algeria
Pr. HOCINI Abdesselem, Univ. M'sila, Algeria
Pr. IDIR Abdelhakim, Univ. M'sila, Algeria
Pr. KHELDOUN Aissa, Univ. Boumerdes, Algeria
Pr. KHOUAS Abdelhakim, Univ. Boumerdes, Algeria
Pr. KIDOUCHE Madjid, Univ. Boumerdes, Algeria
Pr. KOUADRI Abdelmalek, Univ. Boumerdes, Algeria
Pr. KOUINI Benalia, Univ. Boumerdes, Algeria
Pr. LACHEHEB Samir, Univ. M'sila, Algeria
Pr. LACHOURI Abderezzak, Univ. Skikda, Algeria
Pr. MELLIT Adel, Univ. Jijel, Algeria
Pr. MELLIT Mohamed, Univ. Jijel, Algeria
Pr. MERAIHI Yassine, Univ. Boumerdes, Algeria
Pr. MESSAOUDI Noureddine, Univ. Boumerdes, Algeria
Pr. ABDELLI Radia, Univ. Béjaïa, Algeria
Pr. RAHMOUNE Fayçal, Univ. Boumerdes, Algeria
Pr. RECIOUI Abdelmadjid, Univ. Boumerdes, Algeria
Pr. RIAHLA Mohamed Amine, Univ. Boumerdes, Algeria
Pr. SAFI Brahim, Univ. Boumerdes, Algeria
Dr. AIBECHE Abderrezak, Univ. Boumerdes, Algeria
Dr. ABDESLAME Djamila, Univ. Boumerdes, Algeria
Dr. AKLI Isma, CDTA, Algeria
Dr. AKROUM Hamza, Univ. Boumerdes, Algeria
Dr. AMMAR Abdelkarim, Univ. Boumerdes, Algeria
Dr. BAICHE Karim, Univ. Boumerdes, Algeria
Dr. BELAIDI Hadjira, Univ. Boumerdes, Algeria
Dr. BELKACEM Samia, Univ. Boumerdes, Algeria
Dr. BENSAFI Noureddine, Univ. Boumerdes, Algeria
Dr. BOUDOUDA Aimad, Univ. Boumerdes, Algeria
Dr. BOURKACHE Noureddine, Univ. Tizi Ouzou, Algeria
Dr. BOUSTIL Amel, Univ. Boumerdes, Algeria
Dr. BOUTELLAA El-Hocine, Univ. Boumerdes, Algeria
Dr. BOUZID-DAHO Abdellatif, Univ. Tizi Ouzou, Algeria
Dr. CHERRAT Nidhal, Univ. Boumerdes, Algeria
Dr. DAFEUR Abdelhakim, Univ. Boumerdes, Algeria
Dr. DEHMAS Mokrane Univ. Boumerdes, Algeria
Dr. DEKHANDJI Fatma Zohra, Univ. Boumerdes, Algeria
Dr. DICHOU Karima, Univ. Boumerdes, Algeria
Dr. DJAFRI Kahina, Univ. Boumerdes, Algeria
Dr. DJEKOUNE A. Oualid, CDTA, Algeria
Dr. DOGHMANE Mohamed ZINELABIDINE, Univ. Boumerdes, Algeria
Dr. FLITTI Abdelmadjid, Univ. Mostaghanem, Algeria
Dr. GANA Massine, Univ. Boumerdes, Algeria
Dr. GARMAT Abdelkade, Univ. Boumerdes, Algeria
Dr. GUERBAI Yasmine, Univ. Boumerdes, Algeria
Dr. HEDIR Mahdia, Univ. Boumerdes, Algeria

Dr. HAMADACHE Moussa, Univ. Boumerdes, Algeria
Dr. HIMEUR Yacine, CDTA, Algeria
Dr. HOCINE Faiza, Univ. Boumerdes, Algeria
Dr. IKHLEF Boualem, Univ. Boumerdes, Algeria
Dr. KHELOUAT Lila, Univ. Boumerdes, Algeria
Dr. MAHDI Ismahane, Univ. Boumerdes, Algeria
Dr. MAHSEUR Mohammed, Univ. Alger1, Algeria
Dr. MAOUCHE Riadh, Univ. Boumerdes, Algeria
Dr. MASSOUT Samia, Univ. Boumerdes, Algeria
Dr. MECHAKRA Hamza, Univ. Boumerdes, Algeria
Dr. MEKHMOUKH TALEB Sylia, Univ. Boumerdes, Algeria
Dr. MERAH MOSTEFA, Univ. Mostaghanem, Algeria
Dr. MILOUDI Lalia, Univ. Boumerdes, Algeria
Dr. MOUATSI Abdelmalek, Univ. Constantine, Algeria
Dr. MOUHOUCHE Faiza, Univ. Boumerdes, Algeria
Dr. NAAMANE Rachid, Univ. Boumerdes, Algeria
Dr. NEKAA Messaouda, Univ. Boumerdes, Algeria
Dr. OMARI Tahar, Univ. Boumerdes, Algeria
Dr. OULEFKI Adel, CDTA, Algeria
Dr. OURARI Samia, Univ. Boumerdes, Algeria
Dr. RIACHE Samah, Univ. Boumerdes, Algeria
Dr. SEDJELMACI Ibticeme, Univ. Boumerdes, Algeria
Dr. SMAANI Billel, Univ. Mila, Algeria
Dr. TABET Youcef, Univ. Boumerdes, Algeria
Dr. TADJER Sid Ahmed, Univ. Boumerdes, Algeria
Dr. TIBERKAK Allal, Univ. Medea, Algeria
Dr. TOUBAL Maamar-Alla Eddine, Univ. Boumerdes, Algeria
Dr. TRIKI Yacine, Univ. Boumerdes, Algeria
Dr. YSBAA Saadia, Univ. Boumerdes, Algeria

INVITED SPEAKERS

Speakers-01: Dr. Moussa HAMADACHE

Title : System Health Condition Inspection, Monitoring, and Prognosis in Transportation-with Applications to Railway Engineering.



Astract : At present, with the 4th Industrial Revolution and its associated rapid change in technology, industries, and societal patterns and processes (increased interconnectivity and smart automation); the efficient and safe operation of any advanced innovative machine should be a must. Transportation is an important and even crucial part of human life, and the monitoring and prognosis of urban infrastructure is safety-critical. This talk intends to highlight some recent and future innovative system health condition inspection, monitoring, and prognosis technologies within transportation (focusing on railway engineering) that are designed and tested toward their specific application(s) within their working conditions and environment. This achieved through test rigs, test benches, demonstrators, and/or field implementations.

Bio

Dr. Moussa Hamadache has an Engineer of State degree in Automation of Industrial Processes (summa cum laude) and a PhD in Electrical & Electronic Engineering with over 14 years of research experience gained in several international excellent research groups including:

- The Dependable Embedded Control System (DECS) Lab, Kyungpook National University, South Korea;
- The System Health & Risk Management (SHRM) lab, Seoul National University, South Korea;
- The MechVib, the Mechanics and Vibrations Research Group, University of Ferrara, Italy;
- The Birmingham Centre for Railway Research and Education (BCRRE), University of Birmingham (UoB), UK; and
- The Department of Automation of Industrial Processes & the Electrical Systems Engineering Department, Faculty of Technology, University of M'hamed-Bougara Boumerdes, Algeria.

His research interests lie in: System health condition monitoring; Fault detection and diagnosis; Prognostics and health management; Predictive and condition-based maintenance, Robotics (robots in manufacturing and robots in domicile); Signal processing and filtration; Control system and automation; Mechatronics systems including applications to railways.

Dr. Moussa Hamadache has several honors and awards including: Awarded the Korean Government Scholarship Program on 2008; Awarded 4 Student Best Paper Awards between 2014 and 2019, Awarded the Best Paper Award by the American Society of Mechanical Engineers (ASME) on 2020. His work has been included in various high quality, world-leading with international excellent publications, and he serves as a reviewer for several journals (such the IEEE Transaction in Industrial Electronics) and conferences. He is an Associate Fellow of the Higher Education Academy (UK).

Speakers-02: Pr. Madjid KIDOUCHE

Title : Decomposition approach for generating Lyapunov function of complex dynamical system.



Bio

Pr. Madjid Kidouche was born in Bordj-Menaiel, Algeria. He received Engineering, Master of Sciences, and Ph.D degrees all in control theory. He joined M'hamed Bougara University of Boumerdes, Algeria in 1990 where he is a full Professor in the department of automation and electrification of industrial process. He is a research group head on "Control of complex dynamical systems" at Applied Automatic Control Laboratory. He has been actively involved in several research projects in the fields of control and power system analysis. He is the author and co-author of numerous research publications in international conferences and recognized journals. His research interests include modeling and control of dynamic nonlinear systems, stability of large scale systems, fuzzy and sliding mode control.

Papers

Title	*Pages*
ID11 : Drone Secure Element Challenges: A Survey Benabderrezak Youcef, Riahla Mohamed Amine, and Hamadouche Samiya	***pp. 1-5***
ID12 : Image segmentation 3D Ali Dilmi, Ahcen Bouzida, and Yassa Nacera	***pp. 6-10***
ID19 : Multi-Deme Evolutionary Algorithm Based Approach to the Identification of Standard Arabic Sounds Mohamed Aissiou	***pp. 11-16***
ID21 : Influence of Interface defects on Ideality Factor and Schottky barrier in Schottky Diodes based on SiC polytypes Fayssal Mekaret, B. Zebentout, A. Rabehi, S. Tizi, and Z. Benamara	***pp. 17-20***
ID22 : Obstacle Avoiding fire fieghting car-like robot Messaouda Benzaoui, ZIDANI Chahinez, and BOUSSAID Inas	***pp. 21-26***
ID25 : Evaluating Machine Learning Models For Intrusion Classification In Industrial Control Systems ANIBA Hicham, BOUTERFAS Malika, and CHERIFI Tarek	***pp. 27-30***
ID26 : A New Combined Two Channel Set-Membership PU NLMS Algorithm Based on Controlling Energy and Correlation for Blind Speech Quality Enhancement Abdelhak Cheffi, Akila Sayoud, Rahima Henni, and Mohamed Djendi	***pp. 31-35***
ID28 : Speech Emotion Recognition in Arabic and Beyond: A Review Hadjadji Imene, Falek Leila, and Hamza Amina	***pp. 36-40***
ID29 : Emotion Recognition in Speech Using CNNs and Mel Spectrograms: Classifying Seven Emotions Hadjadji Imene, Falek Leila, and Hamza Amina	***pp. 41-44***
ID30 : A COMPUTATIONAL FRAMEWORK FOR AUTOMATIC GROUND CONTROL POINT IDENTIFICATION IN SATELLITE IMAGERY Oussama Mezouar and Fatiha Meskine	***pp. 45-49***
ID43 : A Dual Forward Correlated Set-Membership NLMS Algorithm for Noise Reduction Abdelhak Cheffi, Akila Sayoud, Rahima Henni , and Mohamed Djendi	***pp. 50-54***
ID55 : New Symmetric Set Membership Adaptive Filtering for Blind Source Separation Abdelhak Cheffi, Akila Sayoud, Rahima Henni , and Mohamed Djendi	***pp. 55-58***
ID59 : Optimizing the Structural, Morphological, and Optical Properties of (PbS)0.5(SnS)0.5 Composite Thin Films via Spray Pyrolysis for Advanced Optoelectronic Applications MAKHLOUF Oussama, KHADRAOUI Mohammed, and Redouane MILOUA	***pp. 59-61***
ID65 : Machine learning approach for automated gender recognition Amina Hamza and Imene Hadjadji	***pp. 62-65***
ID66 : A Review of Automated Voice Pathology Detection Databases, Acoustic Features, and Classification Algorithms Amina Hamza, Djamel Addou, and Imene Hadjadji	***pp. 66-71***

ID70 : Design and Implementation of a Low-cost Wireless Charger for Mobile Phones Leila SADOUKI, Feriel LAMINI, and Nourhane AYADI	*pp. 72-77*
ID71 : Design and Implementation of a Digital pH Meter with Analog Conditioning Circuit Calibration Leila SADOUKI, Marouane DAOUDI, and Yacine BENGHARABI	*pp. 78-83*
ID72 : Real-Time PCB Defect Detection on Android Devices Using YOLOv9-Tiny: A Mobile Solution for Industrial Quality Control Mohamed Achit, Nacera Yassa, Ahcene Bouzida, Ali Mazari, Okba Fergani, and Billel Naceri	*pp. 84-89*
ID78 : Design and characterisation of dipole photoconductive antenna for Terahertz multiband applications Hamza Zeraoula, Chellali Slimani, and HadjAbdelkader Benzater	*pp. 90-94*
ID82 : Ada-Boost for Delineation of Raining Cloud over Mediterranean Sea BENSAFI Noureddine, LAZRI Mourad, and AMEUR Soltane	*pp. 95-100*
ID84 : A Novel Approach Based on Geometric Mean Algorithm for Optimum Multiple Capacitor and DG Allocation in Radial Energy Distribution Imene Djedidi, Ahmed Salhi, Anes Bouhanik	*pp. 101-106*
ID85 : Selecting an Appropriate Deep Learning Algorithm for Predictive Maintenance Djahida Belayadi, Asma Djaghlouf, Hadjer Laribi, and Ouezna Meliani	*pp. 107-114*
ID96 : Integration of IoT for Effective Leak Detection in WDNs Chaima CHABIRA, Haddi BAKHTI, Miloud BENTOUMI, and Riyadh BENYETTOU	*pp. 115-118*
ID97 : Leak Detection in WDNs Using Cepstrum Analysis of Pressure Signals Chaima CHABIRA, Haddi BAKHTI, Miloud BENTOUMI, and Riyadh BENYETTOU	*pp. 119-122*
ID100 : Performances Evaluation of Electrical Appliances Identification System using COOLL database in brand-independent mode Fateh Ghazali, Abdenour Hacine-Gharbi, and Hamza Roubhi	*pp. 123-128*
ID101 :ECG signal Denoising Techniques based on Wavelet Transforms : Assessment and Comparison Saida BOUDINA, Samia BELKACEM, and Noureddine MESSAOUDI	*pp. 129-133*
ID103 :Soft Sensing Based On Artificial Neural Network Regression And Data Reconstructed by Self-Organizing Maps for Water Quality Monitoring Mohamed Imed KHELIL, and Mohamed LADJAL	*pp. 134-139*
ID109 :Fabric Fault detection based on image processing and SVM classifier Boualem IKHLEF, Mohammed Amine Sahraoui, and Chemseddine RAHMOUNE	*pp. 140-144*
ID112 :Sound traffic light and smart glasses for the Visually impaired people Benahmed Ines Nadia, khelil Ines, Belfedhal Seifeddine, and Drias Souhil	*pp. 145-150*
ID117 : Dynamic Viewpoint Alignment and Sequence-Specifc Optimization in MV-HEVC Amel Belbel, Amara Bekhouch, and Noureddine Doghmane	*pp. 151-155*
ID127 : Automatic Extractive Text Summarization	*pp. 156-161*

Sabrina Karboua	
ID133 : Investigating the Electronic Properties of Doped SrO for Advanced Spintronic Applications Mokaddem khadidja and bensaid djillali	*pp. 162-164*
ID134 : Self-similar solitons in cubic-quintic media Amel Mecelti and Houria Triki	*pp. 165-169*
ID137 : Detection of QRS by Pan-Tompkins Algorithm applied in the middle variations signal Siham MOUSSAOUI, Sidali FELLAG, Hocine CHEBI, and Yasmina Fadila DIDOUCHE	*pp. 170-174*
ID145 : Improving and Numerical Simulation of Highly Stable and Efficient Perovskite Solar Cells FAPbI3-Based Using Fluorinated Spiro-OMeTAD hole transport layers Mohammed Bouladame and Souad Tobbeche	*pp. 175-189*
ID153 : TCAD simulation of AlN/GaN MOS-HEMT using SiO2 as gate insulator Alem Raida, Soltani Ali, Mattalah Maghnia, and Benzekkour Nadjet	*pp.190-192*
ID166 : Improving vehicle detection from complex urban aerial images with yolov8 Yasmine Silhadi and Hocine CHEBI	*pp. 193-196*
ID32 : Drones In Digital Transformation For Industry:Applications, Challenges and Future Trends Tayeb AOUFI and Karima AKSA	*pp. 197-203*
ID41 : Control of Anode Pressure and Electro-Valve Dynamics for Enhanced Stability in Hydrogen Fuel Cells KABACHE Sabah and BOUREGBA Hicham	*pp.204-209*
ID46 : A Novel Control to Enhance Solar Energy with Electrical Grid Integration Menzri Fatima and Naamane Debdouche	*pp. 210-214*
ID57 : Automation and Supervision of Switching Multipl Power Generators within TRC-SONATRACH Headquarters Maizi Abdelatif and Moussa Hamadache	*pp. 215-220*
ID58 : Stick-Slip Vibration Elimination in Rotary Drilling Systems Using a Tuned GA-PID-Based Controller TELLAA Mawloud, AIBECHE Abderrezak, and DOGHMANE Mohamed Zinelabidine	*pp. 221-225*
ID61 : Comparative Analysis of Various Adaptation Mechanisms in a Sliding Mode Observer-Based MRAS for Vector Control of Induction Machine Zoubir Gadouche, Djilali Toumi ,and Mokhtar Mahmoud Mohammedi	*pp. 226-230*
ID62 : Analysis of the Impact of Objective Function Choice on Optimal FOPID Controller Design for Three Tanks Liquid Level Control Karim Benaouicha, Hamza Akroum, Abdelhakim Idir, and Yassine Bensafia	*pp. 231-238*
ID64 : Comparative Analysis of Torsional Vibration Mitigation Strategies for Enhanced Rotary Drilling Efficiency Zebboudj Ahmed Hichem, Akroum Hamza, Doghmane Z.Mohamed, and Madjid Kidouche	*pp. 239-244*
ID75 : Nonsingular terminal sliding mode controller for controlling an microgrid's Rayane Yasmine Mahfoudi, Boumous Zouhir, and Boumous Samira	***pp. 245-248***
ID76 : State space observer design for high frequency stick slip vibrations in drill-	***pp. 249-252***

string of rotary drilling system Bengana Chahrazad, Aibeche Abderezzak, and Mohamed Zinelabidine Doghmane	
ID79 : IDENTIFICATION OF HIGH-ORDER SYNCHRONOUS MACHINE PARAMETERS FROM STAND STILL FREQUENCY RESPONSES TESTING Farid Leguebedj and Djamel Boukhetala	***pp. 253-258***
ID91 : FOC with LDIW-PSO for OEWIM Supplied by SVM Multi-Level Dual Inverter Mohamed Lamine Ladjel and Abdelhakim Dendouga	***pp. 259-264***
ID98 : Neural Network Modeling of Manipulator Robot Chafea Stiti, Mohamed Benrabah, and Abdelhadi Aouaichia	***pp. 265-268***
ID99 : A Second Order Sliding Mode Control For an Overhead Crane System Oussama Boutalbi, Khier Benmahammed, Mabrouk Boubezoula, Elrachid Bendaoud , and Boualem Boukezata	***pp. 269-274***
ID107 :Open-Circuit and Short-Circuit Faults in Three-Phase Three Level Inverter Fatma Khater and Abderrezak Aibeche	***pp. 275-280***
ID120 : Design and Implementation of a Automatic Multi-Function System for Car Ibtissem Adoui, Akram Maouchi, and Abdelmounim Bouadi	***pp. 281-286***
ID121 : Diagnosis of electrical faults in an asynchronous machine using the sliding mode observer AMMAOUI YOUCEF and FARID TAFININE	***pp. 287-292***
ID123 : Practical Validation of a Control Algorithm for Power Factor Correction and Power Quality Improvement in Distribution Systems Messaouda Bedoui, Abderrahmane Berkani, and Karim Negadi	***pp. 293-298***
ID124 : Fractional Adaptative Control Belgacem Mecheri, Aicha Name Djalab, Nora Kcimi, and Abdelkader Nadjem	***pp. 299-301***
ID128 : Comparing Two Control Techniques Implemented on the Permanent Magnet Synchronous Machine (PMSM): A Comparative Study Wassila HATTAB, Abdelhamid BENAKCHA, Seddik TABET, and Amira SLIMANI	***pp. 302-306***
ID129 :Improved sensorless backstepping controller using model reference adaptive system of a Permanent Magnet Synchronous Motors SLIMANI Amira, BOUREK Amor, AMMAR Abdelkarim, DAHNOUN Ilyes, and HATTAB Wassila	***pp. 307-312***
ID131 :Design of a Super Twisting Sliding Mode Controller for a LC filtered Three-Phase Inverter Houssam Eddine Mansouri, Billel Talbi, Idris Messaoudene, and Moussa Abderrahim Mehiris	***pp. 313-316***
ID132 : Examining the Stability of a Coupled Axial-Torsional Drilling System Across Diverse Operating Scenarios Samira Hadibi, Madjid Kidouch, and Mohamed Zinelabidine Doghmane	***pp. 317-319***
ID154 : Trajectory Tracking of Mobile Manipulator Based on Second Order Sliding Mode Control Samah Riache, Madjid Kidouche, and Issam-Younes Hmadache	***pp. 320-324***

ID164 : Combining MLP and Gradient Boosting for Efficient Intrusion Detection Nabila BOUDIA, Hocine CHEBI, and Malika BELKADI	*pp. 325-328*
ID165 : Complete process of developing a plant disease detection system using deep learning Mohammed Djawed Megaiz and Hocine CHEBI	*pp. 329-331*
ID1 : Whale Optimization Algorithm for Maximum Power Point Tracking in Wind Turbines Billel Naceri, Kamel Aoues, Hamou Ait Abbas, Okba Fergani, Ali Mazari, and Younes Belhadjer	*pp. 332-337*
ID2 : Cuckoo Algorithm for Maximum Power Point Tracking in Wind Turbines Billel Naceri, Ali Mazari, Younes Belhadjer, Okba Fergani, Hamou Ait Abbas, and Kamel Aoues	*pp. 338-343*
ID5 : Performance Assessment of Photovoltaic Power Plant in Algerian Ouest High Plains, Case Study: Naàma Assia Habbati Bellia and Fatima Moulay	*pp. 344-349*
ID6 : Extract Parameters of Photovoltaic System Using Metaheuristic Algorithm Elmamoune Halassa and Abdellatif Seghiour	*pp. 350-353*
ID7 : Modeling and Analysis of Shading Effects on Solar Photo-Voltaic Systems Abbes Med Mounir, Zouhir Boumous, Samira Boumous, and Rayane Yasmine Mahfoudi	*pp. 354-357*
ID9 : Wind Energy Conversion System IGBT Open Circuit Fault Compensation Mokhtar Mahmoud Mohammedi, Hocine Guentri, Noureddine Messaoudi, Zoubir Gadouche, Benabied Djamal, and Samah Bouanik	*pp. 35-361*
ID10 : The process involves determining the optimal angle for the inclination and tilt of the solar panels, taking into account the loss in Algiers province Ali Dilmi, Ahcen Bouzida, and Yassa Nacera	*pp. 362-367*
ID16 : First-Principles Investigation of Optoélectronique Behavior in Pérovskites For Photovoltaic and Imaging Technologies Korichi Hamiche, Yamina Sefir, and ali zitouni	*pp. 368-372*
ID17 : Energy-Efficient HVAC Control Using Deep Learning: A Simulated Approach for Optimizing Building Comfort and Consumption Okba Fergani, Ali Mazari, Raihane Mechgoug, Hamou Ait Abbas, Mohamed Achit, and Billel Naceri	*pp. 373-377*
ID18 : Real-Time HIL Implementation of Integrated Wind Turbine and DSIG System Khaled Benzaoui, Ismail Ghadbane, Mohammed Boukhari, Abderrahmen Bouguerra, and Samir Zeghlache	*pp. 378-381*
ID23 : Indirect forecasting of Photovoltaic power Louiza Ait Mouloud, Aissa Kheldoun, and Samira Oussidhoum	*pp. 382-386*
ID157 : Solar tracking system with Arduino Uno Messaouda Benzaoui and Belfeche oussama	*pp. 387-392*
ID24 : Probabilistic Based Deep Learning for Solar Power Forecasting Samira Oussidhoum, Louiza Ait Mouloud, and Aissa Kheldoun	*pp. 393-398*

ID27 : Hardware Implementation of V/f control in induction motor using Dspace microlabox Ismail Ghadbane, BOUZIDI Riad, BOUKHARI Mohammed, and Khaled Belhouchet	***pp. 399-403***
ID31 : A Fuzzy Logic-Enhanced Variable Step Size P&O Algorithm for Efficient MPPT in Photovoltaic Systems DENNAI Mohammed Yassine, ABDELOUAHED Touhami, TEDJINI Hamza	***pp. 404-407***
ID33 : Grid Tied Inverter with Proportional Resonant Controller for Enhanced PV System Performance Salah Anis KRIM, Fateh KRIM, and Hamza AFGHOUL	***pp.408-413***
ID34 : Multilevel Inverter Topologies: A Case Study of 9-Level Cascaded H-Bridge Inverter Salah Anis KRIM, Fateh KRIM, and Hamza AFGHOUL	***pp. 414-419***
ID39 : Finite element modeling of a self-excited induction generator Belynda Fares, Radia Abdelli, Ahcene Bouzida, and Safia Azzougui	***pp. 420-424***
ID40 : Impacts of DC/DC converter faults in the performance of photovoltaic systems FARES Belynda, AIBECHE Abderrezak, and AZZOUGUI Safia	***pp. 425-431***
ID42 : Assessment of Several Mataheuristic Algorithms for Economic Analysis of a Hybrid Renewable Energy Systems Connected to the Electricity Grid Oussama Belhadef, Farid Merahi, and Abdessalam Badoud	***pp. 432-435***
ID50 : Enhancing the Performance of Grid-Connected Photovoltaic Systems Through Storage Integration: A Techno-Economic and Environmental Study Benchenina Yacine, ZEMMIT Abderrahim, and BOUZAKI Mohammed Moustafa	***pp. 436-441***
ID51 : Utilizing Excess Solar Energy for Green Hydrogen Production in Algeria: A Feasibility Study Benchenina Yacine, ZEMMIT Abderrahim, and BOUZAKI Mohammed Moustafa	***pp. 442-446***
ID45 : Pattern Recognition of Discharge Activity on Polluted Insulators Using Fractal Analysis and SVM Imene Ferrah and Fatma Zahra Hadjaidji	***pp.447-450***
ID161 : Enhancing SAPF Bsed on Packed-U-cell inverter using Particle Swarm Optimization Taha Lemdjed Belahcene, Hamza Afghoul, Okba Kraa, and Djallal Eddine Zabia	***pp. 451-455***
ID53 : Application of Solar Irradiance Parameterization in Advanced Electrical Power Systems: Integration of WRF Model for Renewable Energy Forecasting and Grid Management BAHLOULI Abdelhak	***pp. 256-462***
ID73 : Enhanced Defect Detection in Photovoltaic Systems Using YOLOv10: A Deep Learning Approach Mohamed Achit, Nacera Yassa, Ahcene Bouzida, Ali Mazari, Okba Fergani, and Billel Naceri	***pp. 463-468***
ID74 : Environmental and Economic Impact of Photovoltaic Power Plant emplacement in the Electricity Network Abdelatif GADOUM, Naima BENBOUZA, Tarek SAHRAOUI, and Alla Eddine TOUBAL MAAMAR	***pp. 469-474***

ID81 Optimized MPPT for PV Systems With Grey Wolf Optimization Algorithm Under Partial Shading Condition Alla Eddine Boukhedenna, Hamza Afghoul, Djallal Eddine Zabia, Belgacem Med Nassim Bouzidi, Feriel Abdelmalek, and Lyamine Ouchen	***pp.475-479***
ID89 : Performance Evaluation of Fisher Linear Discriminant Models-Based Fault Detection and Diagnosis in PV Power Generation Systems SAIDI Ahmed, ABDELOUAHED Touhami, DRAOUI Abdelghani, and MASMOUDI Ramadhan	***pp. 480-485***
ID94 : Enhancement of Sensorless Direct Torque Controlled Permanent Magnet Synchronous Motors (PMSM) D. Kouchih, Y. Diboune, R. Hachelaf , M. Tadjine , and M.S Boucherit	***pp. 486-491***
ID95 : A Methodology for Solving VAr/Voltage Problems in distribution system connected wind farm based on IG model ABDELOUAHED Touhami, BEGHDADI Azeddine, SAIDI Ahmed, and Bousserhane Ismail Khalil	***pp.492-498***
ID104 :Misalignment effect on efficiency of wireless power transfer BENNIA Fatima and BOUDOUDA Aimad	***pp. 499-502***
ID108 : Synthesis and Application of Graphene Nanoribbon in Electronic Field Benalia Kouini, Hossem Belhamdi, Asma Nour El Houda Sid, and Amina Hachaichi	***pp. 503-509***
ID110 :Benchmarking YOLOv10 Variants for Solar Panel Surface Defect Detection with Low Epoch Training Younes Benazzouz, and Djilalia Guendouz	***pp. 510-514***
ID113 :Modeling and Simulation of Wind Power System: Comparative Study Yaakoub Diboune and Djilali Kouchih	***pp. 515-520***
ID114 : Improving Power Quality in Microgrids by Optimizing Energy Management Using Genetic Algorithm Sabrine Morakeb, Samira Boumous, Zouhir Boumous , and Mohamed Abderrazak Ghellouci	***pp. 521-525***
ID116 : Enhancing Efficiency of Three-Level Inverters by Minimizing Switching Losses Sid-Ahmed Touil, Hemza Gueddar, Nidhal Cherrat, Tarek Bensidhoum, Nasserdine Boudjerda, and Badis Lekouaghet	***pp. 526-529***
ID125 : Artificial Intelligence Advances for Welding Inspection: A Comprehensive Review of X-Ray Digital Radiography Methods Seghier chahinez Ayad Amar, and Seridi Hassina	***pp. 530-534***
ID136 : Enhancing Energy Efficiency in PV-Powered Ventilation Systems: A Cuckoo Search Algorithm Approach Ali Mazari, Okba Fergani, Hamou Ait Abbas, Kouider Laroussi, Billel Naceri, and Younes Belhadjer	***pp. 535-539***
ID140 : Ensemble Tree Model based MPPT for Standalone Photovoltaic System Mohamed Abderrazak Ghellouci, Samira Boumous, Zouhir Boumous, Sabrine Morakeb, and Mohamed Mounir Abbes	***pp. 540-545***
ID141 : Hardware in The Loop Simulation of Sinusoidal Pulse Width Modulation	***pp. 546-551***

Inverter's Controller Based on OPAL-RT & FPGA BOUYAHIAOUI Hadjer, AZZOUGUI Yasmina, and Mohammed Salim BENMERABET	
ID143 : MPPT Design based on partial Swarm optimization and predictive control for photovoltaic system Nora KACIMI, Abdelhakim IDIR, Belgacem MECHERI, Aicha DJALAB, and Said GROUNI	***pp. 552-556***
ID146 : Modelling and Simulation of the Interleaved Boost Converter Okba Mohammed Cherif, Ziyad Younsi, Belferdi Souad, Nadji Bouchra, Sid Ahmed Tadjer, and Bencherif Hichem	***pp. 557-561***
ID149 : Mathematical Model in Simulink/Matlab of a Battery Electric Vehicle Said BELHADJ	***pp. 562-567***
ID151 : Advanced Virtual Impedance Control Scheme for Power-Sharing Improvement of Three-Phase Paralleled VSIs within a Standalone MG Camelia Ait Hammouda, Rafik Bradai, Rachid Boukenoui, Abdelhammid Kherbachi, and Ahmed Bendib	***pp. 568-573***
ID152 : Monitoring systems technologies for photovoltaic systems : Overview Dermouche Reda, BOUDALI Khadidja, BOUABDELLAH Hiba, and TADJER Sid ahmed	***pp. 574-580***
ID155 : Recent Developments in Photovoltaic Energy: Challenges and Opportunities in Achieving Sustainability Aicha Djalab , Belgacem Mechri, and Kacimi Noura	***pp. 581-583***
ID156 : Random Source Coupling to Shielded Cable Hemza Gueddar, Sid-Ahmed Touil, and Nidhal Cherrat	***pp. 584-589***
ID8 : Impact of High Penetration Photo-Voltaic Generation On Distribution System Voltage Regulation and Stability Zouhir Boumous, Samira Boumous, Med Mounir Abbes, Rayane Yasmine Mahfoudi, and Hadia Belhouchet	***pp. 590-596***
ID37 : Modeling and Simulation of an Intelligent Energy Management System for Hybrid Renewable Energy Microgrids Chakib Guebli , Boubakeur Rouabah, Mohamed Abdelbasset Mahboub, and Bilal Benarabi	***pp. 597-600***
ID167 : Study And simulation of PUC5 Inverter For Photovoltaic Application Fateh Khelifi, Aimad Boudouda, and Tadjer Sid ahmed	***pp. 601-606-***
ID168 : Control of a Doubly-Fed Induction Generator Under Fault Condition Djamila Abdeslame and Mokrane DEHMAS	***pp. 607-611***
ID13 : Adaptive Asynchronous Compressed Spectrum Sensing in Cooperative Networks Hadj Abdelkader BENZATER, Nacerredine LASSAMI, Djamal TEGUIG, and Hamza Zeraoula	***pp. 612-617***
ID38 : Towards the new object-oriented metrics calculation system for software fault prediction MEKAHLIA Fatma Zohra, and CHABANE CHAOUCHE Rabah	***pp. 618-623***
ID47 : Performance Evaluation of Uplink and Downlink Cooperative NOMA Communications Faical Khennoufa, Abdellatif Khelil, Safia Beddiaf, and Walid Beddiaf	***pp. 624-628***

ID48 : Cooperative NOMA assisted Multiple Relays with Direct Links Faical Khennoufa, Abdellatif Khelil, Safia Beddiaf, and Walid Beddiaf	***pp. 629-634***
ID54 : Dual Band antenna for 5G millimeter-wave applications HAMMAD Ilhem, MERIAH Sidi Mohammed, and MERAD Lotfi	***pp. 635-638***
ID60 : Stochastic Geometry-based Analysis of Cellular-Connected UAVs Dina Alkama and Lynda Alkama	***pp. 639-647***
ID63 : Comparative Analysis of Machine Learning Algorithms for IoT Network Traffic Classification DJEGHLOUF Asmaa	***pp. 648-652***
ID67 : Performance Analysis of Uplink NOMA under In-phase and Quadrature Imbalance Beddiaf Safia, Abdellatif Khelil, Faical Khennoufa, and Beddiaf Walid	***pp. 653-658***
ID68 : Effect of Hardware Impairments on Cooperative NOMA with Direct Links Safia Beddiaf, Abdellatif Khelil, Faical Khennoufa, and Walid Beddiaf	***pp. 659-663***
ID80 : Design of a Triple-Band High-Gain Fractal MiMo Antenna with Low ECC for Satellite and Radar Systems Abdelbasset Azzouz, Rachid Bouhmidi, and Mohammed Chetioui	***pp. 664-667***
ID92 : Design of Hexagonal Symmetry Meets Circular Polarization Antenna BOUAMAMA Safi, MERIAH SIDI Mohammed, and BELGACEM Wahiba	***pp. 668-671***
ID93 : Advanced Systems in Electrical Engineering SAMET KOURDI Rahma	***pp. 672-678***
ID102 : A deep learning-assisted MCDM Service Selection Strategy in UAV Networks Abdessalam Mohammed Hadjkouider, Abdelhammid Bouazza, and Malika Abid	***pp. 679-682***
ID118 : 3bit Quantized On-device Recommendation System Ramzi Khantouchi, Ibtissem Gasmi, and Abderaouf Bahi	***pp. 683-687***
ID119 : Alex Net-Based Feature Extraction Combined with a KNN Classifier for Facial Expression Recognition Hamza Roubhi, Abdenour Hacine-Gharbi, and Fateh Ghazali	***pp. 688-693***
ID126 Comparison of three image fusion methods: applied to ALSAT-2A images Samia MASSOUT	***pp. 694-698***
ID130 : An Enhanced Artifcial gorilla troops optimizer with Chaotic map for LEDs Placement in Indoor VLC Systems Abdelbaki Benayad, Yassine Meraihi, and Amel Boustil	***pp. 699-703***
ID138 : Facial expression Recognition for mobile robotics Amine Rami BENNACER, Yacine HATEM, Isma AKLI	***pp. 704-706***
ID158 : Application of Support Vector Machine Regression with Radial Basis Function Kernel in Non-Linear Function Approximation Using AI Techniques Hemza Gueddar, Sid-Ahmed Touil, and Nidhal Cherrat	***pp. 707-710***
ID162 : 6G Networks: A Vision for the Future and Comparative Analysis with 5G TEMERICHT IMANE, H. Bellahsene, and A.Taleb-Ahmed	***pp. 711-716***

ID163 : Snake Optimizer for solving the UAV Base stations Placement Problem Rachda El Taani, Karim Baiche, and Yassin Meraihi	*pp. 717-722*
ID171 :Energy Harvesting Study in V-VLC Systems Using Multi-hop and Commercial Headlights Souad Refas, Sidali Faical, Rahmoune, Ihab Megdoud, and Yassine Meraihi	*pp. 723-727*
ID172 :Simulation and Optimization of Quality of Service in 5G Networks: A Rural Deployment Perspective Souad Refas, Anis Guettaf, and Yassine Meraihi	*pp. 728-732*
ID173 :Particle Swarm Optimization for Solving the UAV-mounted Camera Network Placement Problem Sylia Mekhmoukh Taleb, Karim Baiche, and Yassine Meraihi	*pp. 733-733*
ID44 : Lung and Colon Cancer Histopathological Image Classification Using Deep Learning Approaches Oussama Bouguerra, Youcef Brik, and Bilal Attallah	*pp. 734-738*
ID86 : Breast cancer masses segmentation from mammography images using adaptive thresholding and morphological filtering methods Lilia Radjef, Tahar Omari, and Karim Baiche	*pp. 739-744*
ID106 :Comparative Analysis of Classifcation Models for Bone Mineral Density HAMIDAT Ibrahim, HARRAR Khaled, and AOUINANE Bilal	*pp. 745-750*
ID111 : Chest CT Scan Super Resolution using Deep Residual Attention Fusion Network Rania Saoudi	*pp. 751-756*
ID160 : Phonocardiogram Segmentation Based on HMM Modelling combined with LWE: Application for Heart Valve Disorder Classification Rima Touahria, AbdenourHacine-Gharbi, Philippe Ravier, Karim Abed-Meraim, and Fateh Ghazali	*pp. 757-762*
ID169 : Implementation of VGG for Pneumonia Detection and Brain Tumor Classification Using Raspberry Pi Cherifa ABDELAOUI, Samia BELKACEM, Noureddine MESSAOUDI, Redouane DRAI, and Mounia BACHI	*pp. 763-766*

Drone Secure Element Challenges: A Survey

1st Benabderrezak Youcef
Department of Computer Science,
Faculty of Sciences, LIMOSE Laboratory,
University M'Hamed Bougara of Boumerdes
Boumerdes, Algeria
y.brnabderrezak@univ-boumerdes.dz

2nd Riahla Mohamed Amine
Department of Electrical Systems,
Faculty of Sciences, LIMOSE Laboratory,
University M'Hamed Bougara of Boumerdes
Boumerdes, Algeria
ma.riahla@univ-boumerdes.dz

3rd Hamadouche Samiya
Department of Computer Science,
Faculty of Sciences, LIMOSE Laboratory,
University M'Hamed Bougara of Boumerdes
Boumerdes, Algeria
hamadouche.samiya@univ-boumerdes.dz

***Abstract*—This paper explores the security challenges faced by drones, focusing on using secure elements. It discusses secure element architecture, applications in drone security, drone types and components, and security threats like malware. It highlights the benefits of secure element integration and suggests countermeasures against attacks and secure communication protocols.**

***Index Terms*—UAV, drone, secure element, drone security, side-channel attacks, firmware attacks, malwares**

I. Introduction

The rapid expansion of Unmanned Aerial Vehicles (UAVs) across various sectors, from surveillance to logistics, has underscored pressing security vulnerabilities. This paper begins by examining secure elements, crucial for shielding UAVs from cyber threats, detailing their types and architecture to underscore their importance in UAV security. It also provides an overview of UAVs, including their applications, designs, and key components, while addressing the security challenges they face, particularly from malware and attacks targeting secure elements. This groundwork paves the way for an in-depth discussion on the advantages of integrating secure elements in UAVs and a detailed analysis of the architecture of drone secure elements.

II. Secure Element

A secure element (SE) is a tamper-resistant hardware component designed to securely host and perform cryptographic operations for sensitive applications, such as payment systems, personal data protection, secure identification, and mobile telecommunications.It thwarts unauthorized access and tampering, crucial for secure authentication, digital signatures, contactless payments, cryptocurrency wallets, and mobile transactions. Secure Element (SE) can be implemented through various methods: Removable devices like Universal Integrated Circuit Card (UICC) or MicroSD cards, Embedded SE (eSE) within devices, and Cloud-based SE (CBSE). Secure elements consist of several key components, including a Cryptographic engine, Tamper-proof memory, True random number generator (TRNG), Monotonic counters, Communication interfaces, General-purpose memory, and a Device number.[1,2,3,4,5,6,7,12]

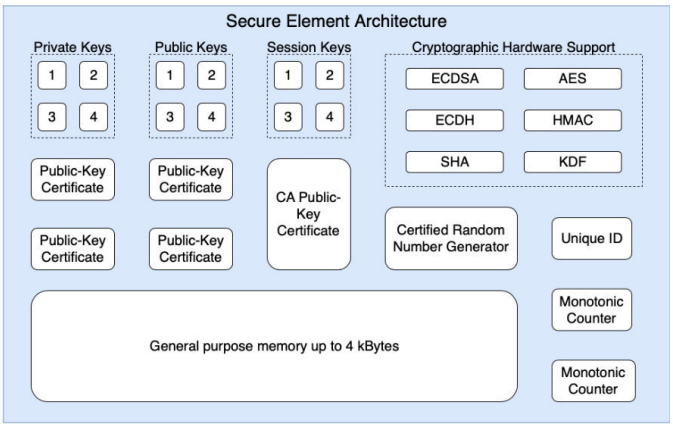


Fig. 1. Secure Element Architecture

III. UAVs

UAVs, or Unmanned Aerial Vehicles, are autonomous or remotely controlled aircraft that operate without a human pilot onboard. UAVs play a pivotal role in diverse industries. They excel in aerial photography and videography, utilizing high-resolution cameras for applications like cinematography, real estate, surveying, and advertising. UAVs are made up of several important parts that work together to enable their flight and operation; including: Airframe, Power Source, Power Distribution Board (PDB), Avionics, Payload, Propulsion System, Control Station, Sensors, and Data Link.[11] (Fig. 2) Based on the type of aerial platform used, there are 4 major types of UAVs : Multi Rotor Drone, Fixed Wing Drone, Single Rotor Drone andFixed Wing Hybrid VTOL.[13][14] (Fig. 3)



Fig. 2. Main components of UAV

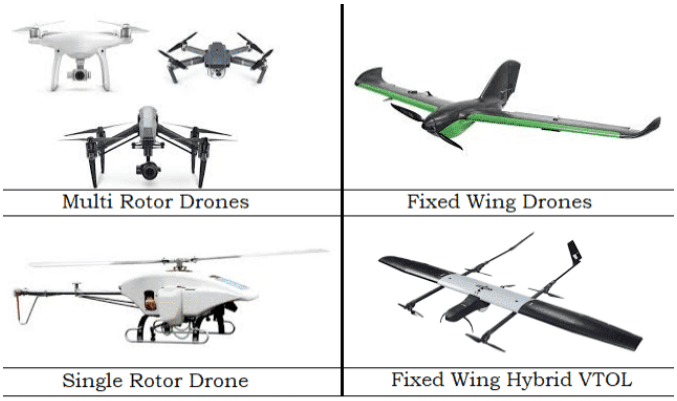


Fig. 3. UAVs types based on the aerial platform used

A. UAV Security Challenges

UAVs encounter multifaceted security challenges that demand attention to ensure their safe operation. Key concerns encompass unauthorized access and control, where malicious entities exploit vulnerabilities in communication and control systems, potentially leading to unauthorized takeovers and misuse. Data security is paramount to protect sensitive information collected and transmitted by UAVs, requiring robust encryption and secure storage. Communication interference risks, such as jamming and interception, threaten the integrity of UAV operations. Cyberattacks targeting onboard and control systems necessitate secure software practices and authentication mechanisms. Physical security, airspace intrusion, and privacy concerns also demand comprehensive measures, including anti-tamper mechanisms, airspace monitoring, and adherence to privacy regulations, to ensure the responsible and secure use of UAVs.[41,42,43]

IV. Malware

Malware, short for malicious software, is any code intended to disrupt systems, whether by unauthorized access, data theft, or damage. It encompasses viruses, worms, Trojans, ransomware, spyware, and adware, each posing distinct threats. To evade detection, malware uses tactics like obfuscation and encryption. Encryption conceals the malware until decrypted, while oligomorphic viruses enhance concealment with diverse iterations. Metamorphic and polymorphic malware adapt their code, challenging antivirus software. Obfuscation techniques, such as junk code insertion and instruction permutation, hinder analysis. Malware analysis involves static scrutiny, dynamic observation, or hybrid approaches. Detection methods like signature-based, heuristic-based, behavior-based, and sandboxing aim to protect systems. These techniques, vital for effective detection, illustrate the evolving landscape of malware evasion and analysis, underscoring the ongoing battle between malicious actors and cybersecurity professionals to safeguard digital environments.[15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29, 40]

V. Challenges

A. Secure elements and Malwares

Malware that targets secure elements, such as smart cards, embedded devices, or cryptographic modules, can be categorized into two main categories: side channel attacks and firmware attacks.

1) ***Side Channel Attacks****:* Side channel attacks exploit physical characteristics like power consumption, electromagnetic radiation, timing, or sound in a device to uncover confidential information or bypass security measures. These attacks reveal details about cryptographic systems, exposing secret keys or operations. For example, monitoring the power usage of a smart card can unveil encryption keys or PIN codes. Various types of side channel attacks target secure elements, including power analysis (SPA and DPA), timing attacks, electromagnetic attacks (SEMA and DEMA), fault injection attacks, and probing and read-out attacks. Protecting against these threats requires implementing countermeasures such as masking, hiding, blinding, randomization, error detection, and tamper resistance in secure elements. However, these measures may not be universally effective and may introduce performance or cost overheads. Thus, designing and evaluating secure elements against side channel attacks remains an ongoing and challenging research focus.[53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64]
in Fig. 5. a general process of how side-channel attacks work.

2) ***Firmware Attacks****:* Firmware attacks, targeting the critical software managing a device's foundational functions, pose severe cybersecurity risks. By illicitly altering or replacing firmware, attackers can compromise device security and functionality, or create backdoors for further exploits. This strategy exploits firmware vulnerabilities to gain unauthorized access, steal data, or disrupt operations, highlighting its potency in undermining device integrity. [44,45,46,47,48,49,50,51,52]
Examples of Firmware Attacks:

- **Firmware Malware:** Malicious code is injected into the firmware to compromise the device's operation. This could involve replacing the legitimate firmware with a malicious version.
- **Firmware Backdoors:** Unauthorized access points (backdoors) are inserted into the firmware, allowing attackers to gain control of the device.
- **Firmware Spoofing:** Attackers may manipulate firmware to present false information about the device's identity or status.

Here's a general outline of the process involved in executing firmware attacks :

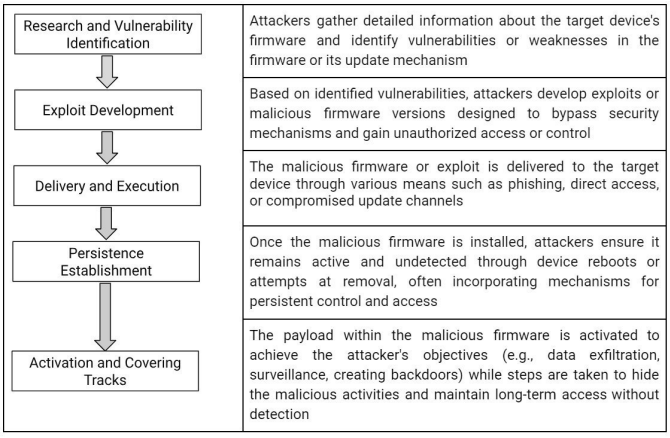


Fig. 4. general process of how firmware attacks work

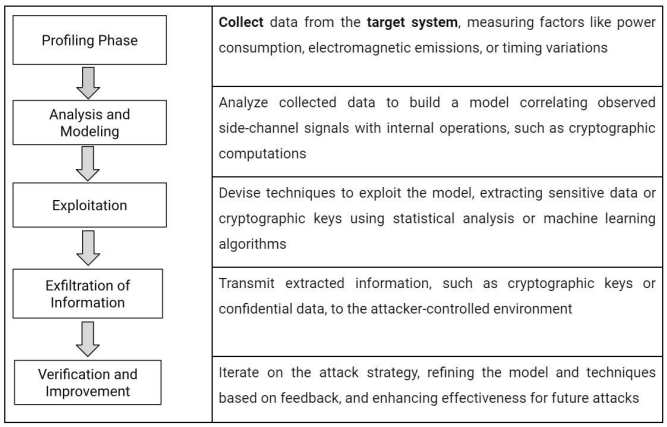


Fig. 5. general process of how side-channel attacks work

B. Secure element into UAV

Incorporating secure elements into UAVs significantly enhances their security, safeguarding sensitive data, ensuring secure and authenticated communications, and protecting against unauthorized firmware modifications and physical tampering. These elements play a vital role in meeting stringent security standards, facilitating robust protection mechanisms that are essential for the safe and reliable operation of UAVs

1) Internal structure of drone secure element: The internal structure of the drone secure element consists of three main layers: the hardware layer, the kernel layer, and the application layer.[12]

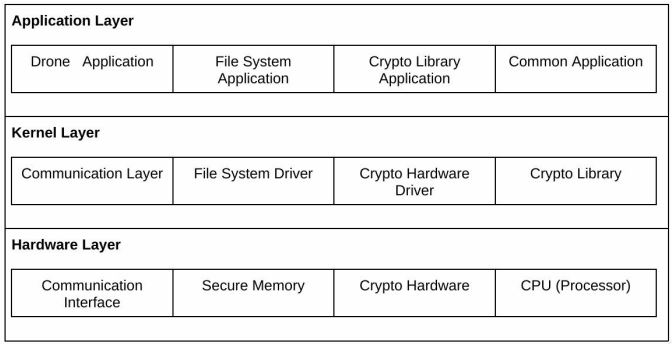


Fig. 6. Internal structure of drone secure element

- **Hardware Layer :** The processor, a central processing unit (CPU), executes instructions and manages the secure element's overall functionality. The communication interface facilitates external communication, while crypto hardware ensures efficient and secure cryptographic operations. Secure memory, resistant to physical and side-channel attacks, safeguards encryption keys and sensitive data, ensuring confidentiality and integrity.
- **Kernel Layer:** Communication and file system drivers enable external communication and manage file operations within the secure element. Crypto hardware drivers interface with dedicated hardware, executing cryptographic operations securely. The KCMVP-certified crypto library undergoes Key Management and Key Validation Program (KCMVP) certification, ensuring compliance with industry standards.
- **Application Layer :** The drone application, specific software on the secure element, handles drone operations. File system application components interact with the file system driver for data management. Cryptographic library applications utilize the certified library for secure cryptographic operations. Additionally, common applications may offer general-purpose functionality or support services for other secure element applications.

C. Malwares Threats to Drone Secure Element

Side-channel attacks pose potential threats to the secure elements in drones, exploiting physical attributes like power consumption, electromagnetic radiation, timing, or sound to extract sensitive information or bypass security measures. While specific side-channel attacks on drone secure elements aren't explicitly mentioned, potential risks include power analysis, timing, electromagnetic, fault injection, and probing attacks. These attacks could compromise cryptographic keys or data stored in secure elements . Awareness of these risks is crucial for drone manufacturers and users to implement necessary safeguards and ensure the security of drone oper-

ations.[35] On the other hand, Firmware attacks pose significant threats to drone secure elements, potentially resulting in unauthorized access, data theft, manipulation, or disruption of operations. Although specific firmware attacks targeting drone secure elements aren't explicitly mentioned, potential risks include vulnerability exploitation, firmware replacement, fault injection, malware injection, and unauthorized access. Vulnerability analysis reveals critical flaws that attackers could exploit. Firmware replacement and malware injection could occur through physical access or remote vulnerabilities. Electromagnetic fault injection (EMF) can hack drones even with robust security measures. To mitigate these risks, drone manufacturers and users must adopt countermeasures such as regular updates, encryption, and secure communication channels.[35,36,39]

VI. Research Avenues

As we grapple with the evolving landscape of UAV security, several avenues of research emerge, beckoning the attention of scholars and practitioners alike. The imperative need to develop robust countermeasures against side-channel attacks, which pose a significant threat to the integrity of secure elements, is paramount. Additionally, firmware attacks and malware incursions demand focused scrutiny, with an emphasis on devising preemptive strategies to mitigate these threats effectively. Exploring innovative secure communication protocols and authentication mechanisms stands as another crucial area of investigation, ensuring the establishment of secure channels for UAV operations. The task of developing secure firmware update mechanisms remains at the forefront, acknowledging the dynamic nature of security challenges in the UAV domain.

VII. Conclusion

In conclusion, this paper emphasizes the pivotal role of secure elements in fortifying unmanned aerial vehicles (UAVs) against evolving threats. The exploration of secure element architecture, coupled with insights into UAV types and vulnerabilities, advocates for their proactive integration. The internal structure of a drone secure element serves as a catalyst for transformative advancements. The imperative to counter side-channel attacks, address firmware vulnerabilities, and innovate in secure communication protocols and authentication mechanisms underscores the pressing need for ongoing research.

References

[1] Ozdenizci, B., Ok, K., & Coskun, V. (2016). A Tokenization-Based Communication Architecture for HCE-Enabled NFC Services. Mobile Information Systems, 2016.

[2] Kaspersky. (n.d.). Secure Element. Kaspersky Encyclopedia Glossary. Retrieved March 9, 2023, from: https://encyclopedia.kaspersky.com/glossary/secure-element/

[3] Park, J., Kim, K., Kim, M. (2008, December). Aegis: A UICC-based security framework for mobile network operators. In 2008 Second International Conference on Future Generation Communication and Networking (pp. 264-269). IEEE.

[4] "MicroSD definition." (n.d.). Computer Hope. Accessed March 15, 2023, from https://www.computerhope.com/jargon/m/microsd.htm

[5] Vincent, A., Marc, P. (2009). Post-distribution provisioning and personalization of a payment application on a UICC-based Secure Element. IN 2009 International Conference on Availability, Reliability and Security

[6] Schläpfer, T., Rüst, A. (n.d.). Security on IoT Devices with Secure Elements. Zurich University of Applied Science (ZHAW), Institute of Embedded Systems (InES), Winterthur, Switzerland.

[7] ZAKARET, Carine, PELADARINOS, Nikolaos, CHEIMARAS, Vasileios, et al. Blockchain and Secure Element, a Hybrid Approach for Secure Energy Smart Meter Gateways. Sensors, 2022, vol. 22, no 24, p. 9664.

[8] SIHAG, Vikas, CHOUDHARY, Gaurav, CHOUDHARY, Pankaj, et al. Cyber4Drone: A Systematic Review of Cyber Security and Forensics in Next-Generation Drones. Drones, 2023, vol. 7, no 7, p. 430.

[9] Nonami, Kenzo, et al. Autonomous flying robots: unmanned aerial vehicles and micro aerial vehicles. Springer Science Business Media, 2010.p.2

[10] C. Watts, V. G. Ambrosia, and E. A. Hinkley, "Unmanned Aircraft Systems in Remote Sensing and Scientific Research: Classification and Considerations of Use," Remote Sensing, vol. 4, no. 6, Art. no. 6, Jun. 2012, doi: 10.3390/rs4061671.

[11] KARAR, Mohamed Esmail, ALOTAIBI, Faris, RASHEED, Abdullah AL, et al. A pilot study of smart agricultural irrigation using unmanned aerial vehicles and IoT-based cloud system. arXiv preprint arXiv:2101.01851, 2021.

[12] KIM, Keonwoo; KANG, Yousung. Drone security module for UAV data encryption. In: 2020 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2020. p. 1672-1674.

[13] SAMANTH, Snehal, KV, Prema, et BALACHANDRA, Mamatha. Security in internet of drones: a comprehensive review. Cogent Engineering, 2022, vol. 9, no 1, p. 2029080.

[14] https://inxee.com/blog/types-of-drones/

[15] M. Christodorescu, S. Jha, S. Seshia, D. Song, and R. Bryant. Semantics-aware malware detection. In Proceedings of the 2005 IEEE Symposium on Security and Privacy, pages 32–46, 2005.

[16] Idika, Nwokedi; MATHUR, Aditya P. A survey of malware detection techniques. Purdue University, 2007, 48.2: 32-46.

[17] VINOD, P., et al. Survey on malware detection methods. In: Proceedings of the 3rd Hackers' Workshop on computer and internet security (IITKHACK'09). 2009. p. 74-79.

[18] ALSMADI, Tibra; ALQUDAH, Nour. A Survey on malware detection techniques. In: 2021 International Conference on Information Technology (ICIT). IEEE, 2021. p. 371-376.

[19] TAHIR, Rabia. A study on malware and malware detection techniques. International Journal of Education and Management Engineering, 2018, 8.2: 20.

[20] Aycock, J. (2006). Computer viruses and malware. Springer US. p. 14. ISBN: 978-0-387-46301-1.

[21] QAMAR, Attia; KARIM, Ahmad; CHANG, Victor. Mobile malware attacks: Review, taxonomy future directions. Future Generation Computer Systems, 2019, 97: 887-909.

[22] RAD, Babak Bashari; MASROM, Maslin; IBRAHIM, Suhaimi. Camouflage in malware: from encryption to metamorphism. International Journal of Computer Science and Network Security, 2012, 12.8: 74-83.

[23] SINGLA, Sanjam, et al. Detecting and classifying morphed malwares: A survey. International Journal of Computer Applications, 2015, 122.10.

[24] ALMASHHADANI, Ahmad O., et al. A multi-classifier network-based crypto ransomware detection system: A case study of locky ransomware. IEEE access, 2019, 7: 47053-47067.

[25] BAZRAFSHAN, Zahra, et al. "A survey on heuristic malware detection techniques." In: *The 5th Conference on Information and Knowledge Technology*. IEEE, 2013, pp. 113-120.

[26] GREAMO, Chris; GHOSH, Anup. "Sandboxing and virtualization: Modern tools for combating malware." *IEEE Security & Privacy*, 2011, 9.2: 79-82.

[27] CUSACK, Greg; MICHEL, Oliver; KELLER, Eric. "Machine learning-based detection of ransomware using SDN." In: *Proceedings of the 2018 ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization*. 2018, pp. 1-6.

[28] ARSLAN, Bilgehan; GUNDUZ, Sedef; SAGIROGLU, Seref. "A review on mobile threats and machine learning based detection approaches." In: *2016 4th International Symposium on Digital Forensic and Security (ISDFS)*. IEEE, 2016, pp. 7-13.

[29] ARSLAN, Bilgehan; GUNDUZ, Sedef; SAGIROGLU, Seref. "A review on mobile threats and machine learning based detection approaches." In: *2016 4th International Symposium on Digital Forensic and Security (ISDFS)*. IEEE, 2016, pp. 7-13.

[30] http://damien.sauveron.fr/fileadmin/damiensauveron/publications/improving-security-autonomous-CameraReady14.pdf

[31] https://developer.parrot.com/docs/airsdk/general/security.html

[32] AKRAM, Raja Naeem, et al. "A secure and trusted channel protocol for uavs fleets." In: *Information Security Theory and Practice: 11th IFIP WG 11.2 International Conference, WISTP 2017, Heraklion, Crete, Greece, September 28–29, 2017, Proceedings 11*. Springer International Publishing, 2018, pp. 3-24.

[33] CEVIZ, Ozlem; SADIOGLU, Pinar; SEN, Sevil. "A Survey of Security in UAVs and FANETs: Issues, Threats, Analysis of Attacks, and Solutions." arXiv preprint arXiv:2306.14281, 2023.

[34] AKRAM, Raja Naeem, et al. "Secure autonomous uavs fleets by using new specific embedded secure elements." In: *2016 IEEE Trustcom/BigDataSE/ISPA*. IEEE, 2016, pp. 606-614.

[35] SCHILLER, Nico, et al. "Drone Security and the Mysterious Case of DJI's DroneID." In: *Network and Distributed System Security Symposium (NDSS)*. 2023.

[36] SIHAG, Vikas, et al. "Cyber4Drone: A Systematic Review of Cyber Security and Forensics in Next-Generation Drones." *Drones*, 2023, vol. 7, no. 7, p. 430.

[37] SALAMH, Fahad E.; KARABIYIK, Umit; ROGERS, Marcus. "A constructive direst security threat modeling for drone as a service." *Journal of Digital Forensics, Security and Law*, 2021, vol. 16, no. 1, p. 2.

[38] Online resource available at https://www.unmannedairspace.info/counter-uas-systems-and-policies/research-by-ioactive-reveals-potential-drone-vulnerabilities-to-non-invasive-attacks/

[39] YAACOUB, Jean-Paul, et al. "Security analysis of drones systems: Attacks, limitations, and recommendations." *Internet of Things*, 2020, vol. 11, 100218.

[40] NGO, Quoc-Dung, et al. "A survey of IoT malware and detection methods based on static features." *ICT Express*, 2020, vol. 6, no. 4, pp. 280-286.

[41] YANG, Wencheng, et al. "A review on security issues and solutions of the Internet of Drones." *IEEE Open Journal of the Computer Society*, 2022.

[42] DHAKAL, Raju; KANDEL, Laxima Niure. "A Survey of Physical Layer-Aided UAV Security." In: *2023 Integrated Communication, Navigation and Surveillance Conference (ICNS)*. IEEE, 2023, p. 1-8.

[43] MEKDAD, Yassine, et al. "A survey on security and privacy issues of UAVs." *Computer Networks*, 2023, vol. 224, 109626.

[44] SASI, Tinshu, LASHKARI, Arash Habibi, LU, Rongxing, et al. "A Comprehensive Survey on IoT Attacks: Taxonomy, Detection Mechanisms and Challenges." *Journal of Information and Intelligence*, 2023.

[45] TSAUR, Woei-Jiunn, CHANG, Jen-Chun, et CHEN, Chin-Ling. "A highly secure IoT firmware update mechanism using blockchain." *Sensors*, 2022, vol. 22, no. 2, p. 530.

[46] ZHOU, Xu, WANG, Pengfei, ZHOU, Lei, et al. "A Survey of the Security Analysis of Embedded Devices." *Sensors*, 2023, vol. 23, no. 22, p. 9221.

[47] NOMAN, Haitham Ameen et ABU-SHARKH, Osama MF. "Code Injection Attacks in Wireless-Based Internet of Things (IoT): A Comprehensive Review and Practical Implementations." *Sensors*, 2023, vol. 23, no. 13, p. 6067.

[48] LITVINOV, Egor, LLUMIGUANO, Henry, SANTOFIMIA, Maria J., et al. "Code Integrity and Confidentiality: An Active Data Approach for Active and Healthy Ageing." *Sensors*, 2023, vol. 23, no. 10, p. 4794.

[49] ZHAO, Yang et KUERBAN, Alifu. "MDABP: A Novel Approach to Detect Cross-Architecture IoT Malware Based on PaaS." *Sensors*, 2023, vol. 23, no. 6, p. 3060.

[50] DIAZ, Alvaro et SANCHEZ, Pablo. "Simulation of attacks for security in wireless sensor network." *Sensors*, 2016, vol. 16, no. 11, p. 1932.

[51] NOMAN, Haitham Ameen, ABU-SHARKH, Osama MF, et NOMAN, Sinan Ameen. "Log Poisoning Attacks in Internet of Things (IoT)." 2023.

[52] MUNICIO, Esteban, MARQUEZ-BARJA, Johann, LATRÉ, Steven, et al. "Whisper: Programmable and flexible control on industrial IoT networks." *Sensors*, 2018, vol. 18, no. 11, p. 4048.

[53] KAUR, Manjit, RAJ, Manish, et LEE, Heung-No. "Cross channel scripting and code injection attacks on web and cloud-based applications: a comprehensive review." *Sensors*, 2022, vol. 22, no. 5, p. 1959.

[54] SELVAM, Ravikumar et TYAGI, Akhilesh. "An Evaluation of Power Side-Channel Resistance for RNS Secure Logic." *Sensors*, 2022, vol. 22, no. 6, p. 2242.

[55] ALAHMADI, Adel N., REHMAN, Saeed Ur, ALHAZMI, Husain S., et al. "Cyber-Security Threats and Side-Channel Attacks for Digital Agriculture." *Sensors*, 2022, vol. 22, no. 9, p. 3520.

[56] ZHANG, Qingqing, ZHANG, Hongxing, CUI, Xiaotong, et al. "Side channel analysis of speck based on transfer learning." *Sensors*, 2022, vol. 22, no. 13, p. 4671.

[57] NERINI, Matteo, FAVARELLI, Elia, et CHIANI, Marco. "Augmented PIN authentication through behavioral biometrics." *Sensors*, 2022, vol. 22, no. 13, p. 4857.

[58] GUPTA, Manik, KUMAR, Rakesh, SHEKHAR, Shashi, et al. "Game theory-based authentication framework to secure internet of vehicles with blockchain." *Sensors*, 2022, vol. 22, no. 14, p. 5119.

[59] SOCHA, Petr, MIŠKOVSKÝ, Vojtěch, et NOVOTNÝ, Martin. "A Comprehensive Survey on the Non-Invasive Passive Side-Channel Analysis." *Sensors*, 2022, vol. 22, no. 21, p. 8096.

[60] NASSIRI ABRISHAMCHI, Mohammad Ali, ZAINAL, Anazida, GHALEB, Fuad A., et al. "Smart home privacy protection methods against a passive wireless Snooping side-channel attack." *Sensors*, 2022, vol. 22, no. 21, p. 8564.

[61] PARK, Jangyong, YOO, Jaehoon, YU, Jaehyun, et al. "A Survey on Air-Gap Attacks: Fundamentals, Transport Means, Attack Scenarios and Challenges." *Sensors*, 2023, vol. 23, no. 6, p. 3215.

[62] MAHMOUD, Dina G., LENDERS, Vincent, et STOJILOVIĆ, Mirjana. "Electrical-level attacks on CPUs, FPGAs, and GPUs: Survey and implications in the heterogeneous era." *ACM Computing Surveys (CSUR)*, 2022, vol. 55, no. 3, p. 1-40.

[63] GUPTA, Himanshu, MONDAL, Subhash, MAJUMDAR, Rana, et al. "Impact of side channel attack in information security." In: *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*. IEEE, 2019, p. 291-295.

[64] MEUNIER, Quentin L., PONS, Etienne, et HEYDEMANN, Karine. "LeakageVerif: Efficient and Scalable Formal Verification of Leakage in Symbolic Expressions." *IEEE Transactions on Software Engineering*, 2023.

Image segmentation 3D

1st Ali Dilmi
Department of Electrical Engineering

Laboratoire des matériaux et développement durable (LMDD)
University of Akli Mohand Oulhadj
University of Akli Mohand Oulhadj
Bouira, Algeria
ali.dilmi@univ-bouira.dz

2nd Ahcen Bouzida
Department of Electrical Engineering

Laboratoire des matériaux et développement durable (LMDD)
University of Akli Mohand Oulhadj
University of Akli Mohand Oulhadj
Bouira, Algeria
a.bouzida@univ-bouira.dz

3rd Yassa Nacera
Department of Electrical Engineering
Laboratoire des matériaux et développement durable (LMDD)
University of Akli Mohand Oulhadj
University of Akli Mohand Oulhadj
Bouira, Algeria
n.yassa@univ-bouira.dz

***Abstract*— Image segmentation generates significant objects for the analysis of digital images. Edge detection is a crucial method for image segmentation. The contour delineates the boundary between two homogeneous zones. The quality of the image is defined by the discontinuities present within it. The objective is to identify the outlines of an item while preserving the original structural characteristics of the image. This thesis. We employed the K-Means method to analyses the 3D MRI brain picture to enhance image quality and diagnose tumors.**

Keywords— Edge detection, Segmentation, MRI brain image, K-means

I. INTRODUCTION

Segmentation algorithms are components of scene analysis designed to divide a scene into significant and uniform areas. In recent years, these techniques have undergone significant advancement, particularly in the segmentation of Magnetic Resonance Images.

This work aims to introduce a novel algorithm for the segmentation of Brain MR Images. In recent years, MRI segmentation has seen significant advancement as a tool for medical diagnosis, particularly in the segmentation of brain MRI, where these approaches facilitate the detection of tumors, necrosis, and the examination of brain-size related disorders, such as Alzheimer's.

Segmentation techniques serve not only as a pre-processing step to assist physicians in disease detection but can also be employed directly to localize tumors and other ailments. The K-means clustering technique has been utilized to identify tumors, oedemas, and necrosis by segmenting and labelling each slice of a 3D Brain MRI through an unsupervised K-means clustering algorithm, subsequently merging them into a cohesive volume.

Such techniques typically accept a greyscale image comprised of pixels or voxels (in the case of 3D) as input, and the segmentation algorithm produces an output image in which the pixels or voxels are replaced by labels that denote the class to which the original image's pixels or voxels belong. In certain instances, a definitive decision rule that allocates each pixel to a singular entity within the image cannot be used, as some voxels or pixels may result from the amalgamation of many tissues. The conventional K-means algorithm typically segments brain MRI images into white matter (WM), grey matter (GM), cerebrospinal fluid (CSF), and background. However, this algorithm solely focusses on the grey value of the brain image, neglecting pixel interrelationships, which results in diminished segmentation accuracy, particularly in low signal-to-noise ratio (SNR) data. This study utilizes the average value of a small neighborhood around each pixel and the pixel's grey value to create a new sample point, aiming to mitigate the influence of noise on clustering accuracy, given that nearby pixels in brain MRI images are likely to belong to the same class. This paper reviews the predominant strategies utilized for brain MRI segmentation. We delineate the distinctions between them and examine their functionalities, benefits, and constraints. To elucidate the intricacies of the brain MRI segmentation issue and confront its problems, we initially present the overarching notion of image processing. This encompasses the definition of 2D images and the description of an image processing system. In the second chapter, we first introduce the fundamental concepts of image processing methods, followed by an explanation of various edge detection techniques and region-based segmentation. We emphasize different approaches to contour detection. Finally, in this paper, we elucidate various MRI preprocessing steps, including image registration, bias field correction, and the removal of non-brain tissue then we explain our approach of clustering methods called K-means and we compare our approach with FCM finally we describe our application of contour slices MRI detection.

II. IMPLEMENTATION

A. Related work

The current methodology relies on thresholding and region growth techniques. The thresholding method disregarded spatial factors. Spatial factors are typically crucial for the detection of malignant tumors.

In thresholding-based segmentation, the image is perceived as possessing only two values: black or white. The bitmap image has greyscale values ranging from 0 to 255. At times, it also disregards the tumor cells. Region growing-based segmentation requires increased user input for seed

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

selection. The seed refers to the core of the cancer cells, which may exacerbate the issue of homogeneity. Moreover, it will not yield satisfactory results in our feature extraction for all photos. We are eschewing thresholding and region growth methods as they are unsuitable for feature extraction techniques.

The proposed system has three primary modules: pre-processing, segmentation, and feature extraction. Preprocessing is accomplished by median filtering. Segmentation is executed using K-means clustering techniques. Feature extraction is an approximation reasoning technique utilized to identify the shape and location of tumors in MRI images by edge detection methods. Numerous methods have been created for segmentation in the current methodology. However, they are not suitable for all sorts of MRI scans. [1]

B. Magnetic Resonance Imaging

A magnetic resonance imaging (MRI) scanner use strong magnets to polarize and excite hydrogen nuclei (single protons) in human tissue, generating a detectable signal that is spatially recorded, yielding images of the body. The MRI equipment generates radio frequency (RF) pulses that selectively target hydrogen. The mechanism transmits the pulse to the designated location of the body requiring examination. The RF pulse causes protons in that region to absorb energy, enabling them to rotate in another direction. This refers to the resonance in MRI. The RF pulse induces the protons to rotate at the Larmor frequency in a designated direction. The frequency is determined by the specific tissue being scanned and the intensity of the primary magnetic field. MRI employs three electromagnetic fields: a strong static magnetic field that polarizes hydrogen nuclei; a weaker, time-varying gradient field utilized for spatial encoding; and a weak radio frequency field that manipulates hydrogen nuclei to generate measurable signals, which are captured via a radiofrequency antenna. [2]

C. The challenges

The brain is the foremost component of the central nervous system. A brain tumor is an intracranial solid neoplasm. Neoplasms arise from aberrant and unregulated cellular proliferation in the brain. An axial view of the brain picture (2D) from an MRI scan was utilized due to the lower danger associated with MRI compared to a CT brain scan. A patient undergoes various diagnostic procedures to ascertain the etiology of the reported symptoms. Procedures such as doing a biopsy and utilizing imaging techniques, including MRI or CT scans of the brain, will be executed. In a biopsy, pathologists obtain a sampling of the brain tissue to assess for the presence of a tumor. A pathologist examines tissue cells under a microscope to identify abnormalities. While a biopsy reveals the presence and pathology of a tumor, surgeons must ascertain the tumor's extent and precise location within the brain. This information can be obtained through an MRI scan, which does not utilize harmful radiation, unlike a CT scan. The conventional approach in hospitals involves manually segmenting the medical image, relying on the physician's ability to perceive the image accurately to extract the necessary region. This process is complicated by subtle variations and similarities between the original and affected biological structures in the image. The scarcity of radiologists and the substantial number of MRIs requiring analysis render these evaluations labor-intensive and costly.[2]

D. Clustering Method

This is an iterative method employed to segment an image into clusters. In the clustering methodology, the number of clusters K serves as an input parameter. The primary challenge remains unchanged. The procedure for clustering is as follows.

1. Select K cluster centers, either randomly or according to a heuristic method.
2. Allocate each pixel in the image to the cluster that minimizes the distance to the cluster centroid.
3. Recalculate the cluster centroids by averaging all pixels within the cluster.
4. Iterate steps 2 and 3 until convergence is achieved (e.g., no pixel's transition across clusters).

Clusters may be selected manually, randomly, or according to certain criteria. The distance between a pixel and the cluster center is determined by the squared or absolute difference between the pixel and the cluster center. [2]

E. Image segmentation by clustering method

Clustering is regarded as the paramount unsupervised learning problem, as it involves identifying a structure within a set of unlabeled data. Clustering is defined as "the process of categorizing objects into groups based on their similarities." A cluster is a collection of things that are "similar" to one another and "dissimilar" to objects in other clusters.

F. K-Means Algorithm

K-means is a fundamental unsupervised learning algorithm that addresses the established clustering problem. The process employs a straightforward method to categories a certain data set into a predetermined number of clusters established a priori. This algorithm seeks to minimize an objective function, specifically a squared error function.

The algorithm consists of the subsequent steps:

1. Determine a value for K, the quantity of clusters.
2. Randomly initialize the K cluster centroids, if required.
3. Determine the class affiliations of the N items by allocating them to the nearest cluster centroid. [2]
4. Recalculate the K cluster centers, presuming the previously determined memberships are accurate.
5. Continue steps 3 and 4 until no items among the N have altered their membership in the most recent iteration. [3]

G. Fuzzy C-Means Clustering Algorithm

The Fuzzy C-Means (FCM) clustering algorithm was initially proposed by Dunn and then expanded by Bezdek. The procedure is an iterative clustering method that generates an ideal c-partition by minimizing the weighted within-group sum of squared error objective function J FCM.

$$J_{FCM} = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^q \, d^2 \, (x_k, v_i) \tag{1}$$

Let X = {x1, x2, ..., xn} ⊆ Rp represent the data set in the p-dimensional vector space, where n denotes the number of data items, c signifies the number of clusters with the constraint 2 ⩽ c < n, uik indicates the degree of membership of xk in the ith cluster, q is a weighting exponent applied to each fuzzy membership, vi represents the prototype of the center of cluster i, and d2(xk, vi) denotes the distance measure between object xk and cluster center vi. [4]

Basis	K-Means	Fuzzy C means	Reason
Efficiency	Fairer	Slower	K-Means just needs to do a distance calculation, whereas Fuzzy C Means needs to do a full inverse-Distance weighting
Objective function	$J = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\|x_i^{(j)} - c_j\right\|^2$	$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m}$	The objective functions are virtually identical, the only difference being the introduction of a vector which expresses the percentage of belonging of a given point to each of the clusters
Performance	Traditional and limited	Can be used in variety of clusters and can handle uncertainty	FCM may converge faster than hard K-Means, somewhat offsetting the bigger computational requirement of FCM
Applications	In image retrieval algorithms	Segmentation of magnetic resonance imaging (MRI) analysis of network traffic-Fourier – transform infrared spectroscopy	

H. Comparative analysis of K-means and Fuzzy c-means algorithm [5].

TABLEI. Comparative analysis of K-means and Fuzzy c-means algorithm

III. MATLAB FOR MRI APPLICATIONS

A. DICOM file Strucutre:

DICOM images are denoted by the '.dcm' extension. A DICOM file comprises a 'Header' and a 'Dataset'.
The header includes details regarding the encased dataset. It comprises a File Preamble, a DICOM prefix, and the File Meta Elements.

The header includes details regarding the encased dataset. It comprises a File Preamble, a DICOM prefix, and the File Meta Elements.

```
Untitled.m
1 -  dataFolder = 'C:\Users\Abderrahim\Desktop\Head_Axial_DICOM\';
2 -  imageName = 'CTHead1.dcm';
3    %%dcmMetaData = dicominfo(dataFolder imageName );
4 -  dcmMetaData = dicominfo([dataFolder imageName]);
5 -  mriImage1 = dicomread([dataFolder imageName]);
6    % Converting from uint16 to double
7 -  mriImage1 = double(mriImage1);
8    %%figure;
9    %%%%%%%%%%imagesc(mriImage1);
10 -  colormap('gray')
11   %%Turn off axis
12 -  axis off
13 -  for iImg = 1:22
14 -  mriImage(:,:,iImg) = dicomread([dataFolder sprintf('CTHead%d.dcm', iImg)]);
15 -  end
16   % Converting MRI data from uint16 to double
17 -  mriImage = double(mriImage);
18 -  montageImg = reshape(mriImage,[512 512 1 22]);
19 -  montageCMap = gray(128);
20 -  montage(montageImg, montageCMap);
21   %%%figure;
22   %%figure;
23   %%for iImg = 1:93
24   %%imagesc(mriImage(:,:,iImg));
25   %%axis off; colormap gray
26   %%hold on;
27   % pause(n) slows down image display
28   %%pause(0.3)
29   %%end
30
4 usages of "dataFolder" found
```

Fig.1. source code (part1)

B. Accessing DICOM Metadata:

The 'dicominfo' function retrieves the metadata from a DICOM file.

```
1 -  dataFolder = 'C:\Users\Abderrahim\Desktop\Head_Axial_DICOM\';
2 -  imageName = 'CTHead1.dcm';
```

Fig.2. The 'dicominfo' function retrieves the metadata from a DICOM file.

C. Loading DICOM image:

DICOM image data is read into MATLAB using 'dicomread' function.

```
4 -  dcmMetaData = dicominfo([dataFolder imageName]);
5 -  mriImage1 = dicomread([dataFolder imageName]);
6    % Converting from uint16 to double
7 -  mriImage1 = double(mriImage1);
8    %%figure;
```

Fig.3. DICOM image data is read into MATLAB using 'dicomread' function.

Read a DICOM image by using dicomread and converting from uint16 to double

```
13 - for iImg = 1:22
14 - mriImage(:,:,iImg) = dicomread([dataFolder sprintf('CTHead%d.dcm', iImg)]);
15 - end
16   % Converting MRI data from uint16 to double
17 - mriImage = double(mriImage);
```

Fig.4. Read a DICOM image by using dicomread And Converting from uint16 to double.

Generally, a single DICOM file contains a single MRI slice data. To read in MRI data for all slices, each individual DICOM files are loaded separately. 'For' loop is used to read in all data files.

D. Montage

A montage is formed by assembling visual fragments into a cohesive image.

The MATLAB function 'montage' accepts a MxNx1xK image dataset and presents numerous image frames in a montage format. 'Reshape ()' is employed to resize the MRI image collection for the purpose of constructing a montage.

```
18 - montageImg = reshape(mriImage,[512 512 1 22]);
19 - montageCMap = gray(128);
20 - montage(montageImg, montageCMap);
```

Fig.6. image frames as a montage. 'Reshape ()' is used to resize MRI image dataset for creating montage.

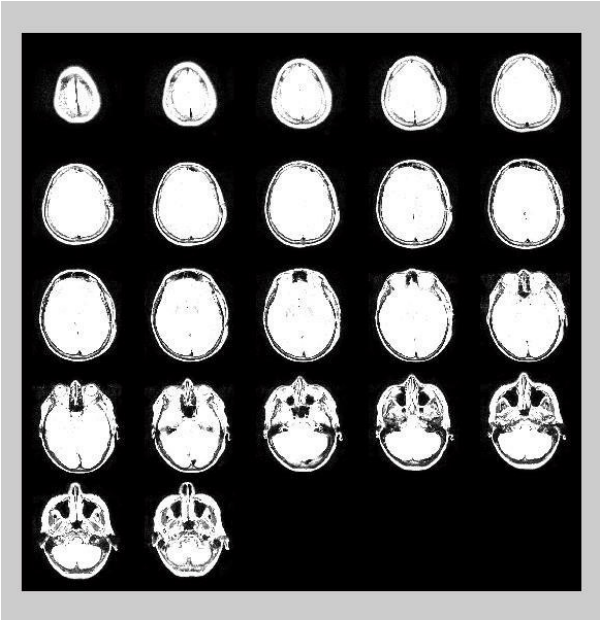


Fig.5. brain slices Contour Slices

Each slice in the image collection constitutes a 2D object, and when aggregated, they form a 3D volumetric data object. Utilizing 'contourslice' enables the creation of a three-dimensional object composed of slice contours, which can then be observed from any orientation.

```
68 - figure
69 - cm = gray(50);
70 - cHandle = contourslice(mriImage,[],[],[1 5 15],8);
71 - colormap(cm)
72 - set(cHandle,'FaceColor','interp')
73
```

Fig.6. create a 3D object consisting of slice contours By using 'contour slice'

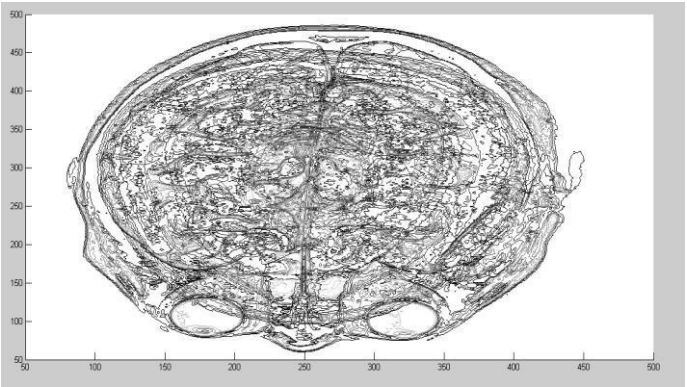


Fig.7. Brain slices

IV. Conclusion

This work aims to offer a 3D volume segmentation technique applicable to MR brain data. The primary characteristics of our algorithm are: K-means voxel clustering those accounts for both general and local features

of the 3D picture. The segmentation of brain structures not only facilitates a more precise characterization of their functions but also enables the isolation of markers for specific disorders to assist physicians in diagnosis.

The aim of this study is to implement K-means algorithms for the segmentation of brain structures utilizing magnetic resonance imaging (MRI). We have contrasted our methodology with the fuzzy c-means algorithm, which is inherently similar to the one delineated in this study, however diverges in several significant aspects. The fuzzy c-means method operates on two-dimensional images. We are exploring the application of the proposed approach for segmentation in additional organs and various imaging modalities. Our research demonstrated that favorable outcomes are anticipated in CT and MRI modalities; however, other modalities, such as X-ray and ultrasound imaging, may require enhanced discriminative qualities. Utilizing more robust discriminative dictionary learning methods could significantly enhance the method's distinguishing properties. The primary objective is to develop a technique that can segment any region exhibiting distinct features (such as texture, intensity, frequency information, etc.) across several imaging modalities.

REFERENCES

[1] [SivaSankari.S, Sindhu.M, angeetha.R, ShenbagaRajan]," Feature Extraction of Brain Tumor Using MRI ",vol3,march 2014.

[2] [Kalpana Shrivastava, Neelesh Gupta, Neetu Sharma],"Medical Image Segmentation using Modified K Means Clustering"," International Journal of Computer Applications", Volume 103 – No 16, October 2014.

[3] [Jianwei Liu, Lei Guo], " An Improved K-means Algorithm for Brain MRI Image Segmentation "," School of Automation, Northwestern Polytechnical University, Xi'an, Shanxi, 710072, China ", 2015.

[4] [Yong Yang, Shuying Huang]," IMAGE SEGMENTATION BY FUZZY C-MEANS CLUSTERING ALGORITHM WITH A NOVEL PENALTY TERM"," Nanchang 330013, P.R. China", vol26, 2007.

[5] [Ankita Singh, Dr Prerna Mahajan]," Comparison Of K- Means And Fuzzy C- Means Algorithms"," International Journal of Engineering Research & Technology (IJERT)"," Vol. 2 Issue 5, May – 2013.

.

Multi-Deme Evolutionary Algorithm Based Approach to the Identification of Standard Arabic Sounds

Mohamed Aissiou
Electronics department
National Higher School of Advanced Technologies
Algiers 16001, BP 474, Algeria
mohamed.aissiou@ensta.edu.dz

Abstract--The goal of this paper is the application of the Multi Deme Parallel Evolutionary Algorithm (MDPEA) such as Evolutionary Genetic Algorithms (EGAs) to the Automatic Speech Identification domain at the acoustic sequences identification level. The Automatic identification has been formulated as functions optimization problem Thus, we have looked for recognizing Standard Arabic (SA) sounds of continuous, naturally spoken, speech by using the MDPAA witch have several advantages in resolving complicated optimization problems. In SA, there are forty distinctive sounds. We have analyzed a corpus that contains several sentences composed of the whole SA phonemes types in the initial, medium and final positions, recorded by several Algerian male speakers. We have used the decision rule Manhattan distance as the fitness functions for our MDPEA evaluations whose topology chosen is the Multi Deme isolated island one. The Corpus phonemes were extracted and identified successfully with an overall accuracy equals to 92.60%. In addition, the computational cost was greatly reduced, and so the performances of the MDPEA were improved.

Key words-- Parallel Evolutionary Algorithms, Isolated Multi-deme, Arabic Sounds, Automatic phonemes Identification..

I. INTRODUCTION

The essential task of speech recognition is to capture a speech signal from an input device such as microphone and to identify the label of this signal. However, this task is not simple because of the time-varying properties of speech signals. Thus, in the course of developing a speech recognition system, the crux of matter would seem to be to perform a robust and accurate matching in the recognition process. This is by no means an easy task, as speech utterance in general is fuzzy, varies in time and is sometimes unpredictable. Once the process of extracting feature vector has completed, the recognition systems compare this set of feature vectors with each reference pattern in the database by using some kind of pattern classification technique. The one with the largest value of similarity would be considered as the most likely candidate of the unknown utterance.

The Evolutionary Algorithms (AEs) are search methods for good solutions in a large population of candidate solutions. They have several advantages. They work with both continuous and discrete parameters, execute simultaneous searches over several regions of the search space and work with a population instead of a unique point. In addition, they optimize a large number of parameters; have successfully found global minimum even on very complex and complicated objective functions and their computer implementations are portable and modular. In addition, they are tolerant to incomplete and noise data [1].

However, computing cost is the main problem of EAs, especially in our situation of large size of population. In order to apply the evolutionary search into large-scale problems, we have used an isolated multiple-deme parallel EGA topology. It simply consists in evolving all the tasks of a basic genetic algorithm on each deme or subpopulations, in parallel manner. For each phoneme class, we have a subpopulation. The demes individuals or chromosomes are defined as the segmentation acoustic vectors. So, we have forty demes. For each one, we apply a basic GA in parallel with the others, no time will be spent on communication and thus, the algorithm is run much faster and has found more accurate results than the sequential one.

II. STANDARD ARABIC SOUNDS

The Standard Arabic language is composed of 28 phonetically distinct consonant phonemes and three short vowels, [a, u, i], which contrast phonemically with their long counterparts, [a:, u :, i :]. Throughout the text, phonemic length is indicated by writing the vowels symbol twice. In addition, there are six (06) correspondent variants of the short and the long vowels, in emphatic context.

Generally, consonants have less energy than vowels. The characteristics which form a vowel are relatively more prominent and stable than those of the consonants. The consonants vary individually, making it easier to deal with them in group. All the Arabic vowels are oral and fully voiced. But, they can be nasalized in nasal consonants context (assimilation). The difference between short and long ones is approximately double duration or more [2].

The 28 consonants of SA are classified physiologically as follow :

- 13 fricatives ; [θ],[ħ], [X], [ð], [s], [ʃ], [$_s$],[f], [z], [ð],[ʕ], [ɣ], [h]. Fricatives are consonant sounds where the vocal tract is excited by a turbulent airflow, produced when the airflow passes a constriction in the vocal tract.
- 08 stops : /ʔ/ /q/,/ k/, /t/, /ṭ/, /d/,/ḍ/, /b /. Stop consonants, are produced in two phases. In the first phase, air pressure is built up behind a complete constriction at some point in the vocal tract. In the second phase, there is a sudden release of this air which produces the plosive sound.
- 02 nasals : [n], [m]. These are all voiced and involve moving air through the nasal cavities by blocking it with the lips, gums, and so on. Nasals are lower in energy than most vowels due to the closure of the oral cavity and the limited ability of the nasal cavity to radiate sound.
- 04 semi Vowels : the two glide [j], [w], the trill [r] and the liquid : [l]. They have vowel-like acoustic characteristics.
- 01 affricate: [dz]. It is a dynamic consonant sound that results from the combination of two sounds: the transition from a plosive to a fricative.

The relative duration of the consonants depends upon whether they occur initially, medially or finally. It also depends on whether they are aspirated or unaspirated, voiced or unvoiced and single or geminated.

The SA is characterized by three phonetic phenomena which are the presence of the emphatic consonants, the geminate ones and by the presence of glottal, pharyngeal, velar and uvular ones called back consonants. The SA possesses eight back phonemes. The phonemes in Arabic that are not found in English include the following: /ʔ/ /q/,/ k/, /t/, /ṭ/, /d/,/ḍ/, /b / [2].

III. SPEECH SOUNDS ANALYSIS

Several speech analysis methods can be used to extract spectral envelope characteristics of the speech signal such as filter bank analysis, **L**inear **P**redictive **C**oding (**LPC**), and **M**el-**F**requency **C**epstral (**MFC**) analysis. Among these techniques, we have used both the LPC and the MFC analysis because the LPC and the MFC Coefficients are the best known and most commonly used features for ASR. Just the first coefficients of the predictive coding and the cepstral sequence are interesting for smoothing the spectrum and minimising the pitch influence [3].

A. Parametric vectors of phonemes

It turns out that the vowel and semi-vowel segments have a great steady-state part. Consequently, a 10th order linear prediction analysis performs a sufficiently high prediction gain. Broadly, the autocorrelation method of linear prediction ensures the stability of the estimated all-pole filter and hence, is used here for analysis (with 20 ms Hamming window). Each 10 ms frame or segment n°m noted S_m is represented by a short parametric vector noted Vv_{10m} composed of ten prediction coefficients a_{im} . The description of this vector is as follows :

$$Vv_{10m} = (a_{1m},...,a_{im},...,a_{10m}) \quad (1)$$

Temporal changes, in speech spectra, play an important role in perception. This information is captured in the form of velocity and acceleration coefficients (collectively referred to as differential or dynamic features). It is often the dynamic characteristics of the features that provide most information about phonetic properties of speech sounds (related to, for example, formant transitions or the closures and releases of stop consonants). No time evolution information is included in MFCCs. But it is often included in the feature set by cepstral derivatives. The first order derivatives of MFCCs are called Delta coefficients, and their second order derivatives are called Delta-Delta coefficients.

It is common to append an energy coefficient to the cepstrum feature vector. Differences in energy among phonemes show that it is a good feature to distinguish between them. **S**hort **T**ime **E**nergy (**STE**) is simply the mean square power in the frame.

Consequently, the fricative and nasal consonants segments n°m are represented by a features vector noted Vcf_{39m} which contains the twelve first MFCCs, the STE and their first and second order derivatives. The description of this vector is as follows :

$$Vcf_{39m} = (cf_{1m}, cf_{2m},..., cf_{41m}) \quad (2)$$

The **B**urst **F**requency (**BF**) was found to be one of the most important perceptual clues to human classification of stops. The BF is a parameter aimed at finding the frequency where power is concentrated in a stop. Essentially, at each time position of the stop burst, there exists a position of

maximum intensity. Each of these positions has a corresponding frequency. The BF is the minimum value of that set. The **V**oicing **O**nset Time (**VOT**), defined as the duration from the release to the start of voicing (burst duration) was used to detect voicing. Unvoiced stops were found to have longer VOTs [4].

To search stop and affricate consonants, each parameter vector noted Vcs_{41m} of the acoustic segment n° m is composed of the first twelve MFCCs, the STE and their first and second time derivatives. Also, we add to this set of parameters the VOT and the BF ones. The description of this vector is as follows :

$$Vcs_{41m} = (cs_{1m}, cs_{2m}, ..., cs_{41m}) \quad (3)$$

B. Reference acoustic data

The choice of the type of the discriminative acoustic parameters of phonemes to be recognized is very important. These last ones, as acoustic vectors form, constitute the set of reference data for the AG. Each SA phoneme of class k is represented by a mean parameters vector noted μ_{km} :

$$\mu_{km} = (\mu_{k1}, \mu_{k2}, ..., \mu_{km}) \quad (4)$$

k is the number of SA phonemes. The SA alphabet is composed of 28 consonants, 06 vowels and their 06 vocalic variants in emphatic context.

These reference acoustic vectors are obtained during the learning phase. Each of the forty phonemes is represented by an acoustic reference vector of a specific order. The mean vectors of the forty phonemes classes are computed from the data in the training set by using the following relations:

$$\mu_{km} = \frac{1}{N_k} \sum_{z=1}^{N_k} v_{kz}, \qquad 1 \le k \le 40 \quad (5)$$

Where N_k is the number of preclassified vectors in the class k and v_{kz} the vector z of the class k.

IV. EVOLUTIONARY ALGORITHM

An EGA is a global optimization procedure that uses an analogy of the genetic evolution of biological organisms. It is a heuristic search procedure that modifies function values of individuals coded as binary (or real or symbol) strings, through the application of predefined reproduction operators in a stochastic manner. In our case, the EA consist on looking for, among the various acoustic segments of a vocal sequence, structures which correspond to those of the representative reference vectors of every class. To associate entities to be recognized in classes, it is enough to find an analytical correspondence between the two groups of the acoustic vectors respectively vocal sequences and entities to be recognized of reference. The vectors of the initial population represent the search space for potential solutions. The EGA will have for task to make on these initial vectors some modifications, reorganizations, to produce the best vector as the potential solution [5].

A. Genetic encoding of speech sounds

In EGAs, each potential solution is represented in the form of a bit string which is dubbed a chromosome or an individual. A pool of strings forms a population. We use a fixed length chromosome for the EGA, where each chromosome contains, as genes, the coefficients of several segments noted S_m of the vocalic continuum of the chosen corpus. The bit string length depends on the required numerical precision. The precision of the solutions is bounded by the precision of the data type used in the implementation of the genetic algorithm. We have used a double precision data type of 64 bits following the specification IEEE Standard for Binary Floating-Point Arithmetic. This data type has a precision of 15 - 17 digits.

The individuals are elaborated relatively to the phoneme to be classified. For example, if we want to classify the vowel [u], we proceed to the segmentation of the vocalic continuum into a set of acoustic vectors according to the corresponding individual that will be definite as the chromosome.

This acoustic vector noted $V_{vo_k\ m}$ is composed of the coefficients of a certain number of successive concatenated acoustic segments and it's size is equal to the order noted vo_K of the phoneme reference vector and it is expressed as follows :

a. Stop and affricate consonants :

$$V_{vo_k\ vm} = (cs_{1m}, ..., cs_{41m}, ..., cs_{1(m+1)}, ..., cs_{41(m+1)}, ..., c_{vo_k}) \quad (6)$$

With : $vo_k = r \times 41$;

And : r is the number of acoustic segments concatenated to form the individual vector.

b. Fricative and nasal consonants:

$$V_{vo_k\ m} = (cf_{1m}, ..., cf_{39m}, ..., cf_{1(m+1)}, ..., cf_{39(m+1)}, ..., cf_{vo_k}) \quad (7)$$

With $vo_k = r \times 39$;

c. Vowels and semivowels :

$$A_{vo_k\ m} = \left(a_{1m},...,a_{im},...,a_{10m},...a_{1(m+1)},...,a_{vo_k}\right) \quad (8)$$

With : $vo_k = r \times 10$;

It results that the initial population noted pop will be composed of all possible acoustic vectors $V_{m\ vo_K}$ when we vary the values of m and vo_K. The different values of m are as follow: $m = 1, \ . \ . \ . \ , M$.

With : $M = T/10$ and T is the vocalic continuum duration. Each individual is made up of vo_K genes.

B. EGA Objective Function

To evolve good solutions and to implement natural selection, we need a measure for distinguishing good solutions from bad solutions. The measure could be an objective function that is a mathematical model or a computer simulation, or it can be a subjective function where humans choose better solutions over worse ones. In essence, the fitness measure must determine a candidate solution's relative fitness, which will subsequently be used by the EGA to guide the evolution of good solutions. Once the problem is encoded in a chromosomal manner and a fitness measure for discriminating good solutions from bad ones has been chosen, we can start to *evolve* solutions to the search problem [6].

The **F**itness **f**unction (**Ff)** can be represented rigorously by the Manhattan distance between the vectors of the acoustic segments and the vectors of reference of every class of stops. It can be expressed as follows [7].

<u>*a- vowels and semivowels :*</u>

$$Ff_k = \sum_{m=1}^{n} \frac{\left|A_{mvok} - \mu_{mk}\right|}{\sigma} \quad (9)$$

<u>*b- consonants :*</u>

$$Ff_k = \sum_{m=1}^{n} \frac{\left|V_{mvok} - \mu_{mk}\right|}{\sigma} \quad (10)$$

Each deme which corresponds to one phoneme constitutes the initial population of the basic GA.

To estimate potential solutions of our problem, we minimize the Ff to decide of the type of membership classes of the acoustic segment or the sets of these segments. So, we maximize the inverse function of Ff, and looking for the extrema which can give two types of results: the global maximum means the existence of the phoneme we are looking for. In that case, we continue the search of the same type of sound for the rest of the vocal continuum. Otherwise, the end of the vocal continuum means that there is no phoneme of class k in the vocal continuum, so we repeat the research for the other types of Ssounds.

V. MULTI-DEME PEA

Multiple-deme parallel PEAs are also called "distributed" EGAs, because the communication to computation ratio is low, and they are often implemented on distributed memory computers. They are also known as "island model" EGAs, because they resemble a model that is used to describe natural populations isolated by the distance between them (as in islands). Both in the model and in the EGAs individuals may migrate occasionally to any population. The population is divided into a few subpopulations or demes, and each of these relatively large demes evolves separately on different processors. Exchange between subpopulations is possible via a migration operator. On each deme or island the population is free to converge toward different optima. Technically, there are three important features in the coarse grained PGA: the topology that defines connections between subpopulations, migration rate that controls how many individuals migrate, migration intervals that affect how often the migration occurs [8].

This bounding case of our MDPEA considers that the demes evolve in complete isolation. Without communication, the migration rate is zero, and this is clearly a lower bound. Also, no connections between the demes represent a lower bound in the connectivity of the topology. Since the demes are completely isolated, the parallel speedup is simply the ratio of the time used to evaluate individuals in the serial and parallel cases [9].

This is the topology we have chosen to our MDPEA. For each phoneme class, we have a subpopulation or deme. The deme individuals or chromosomes are defined as the segmentation acoustic vectors. We devise our population into three main groups of subpopulations according to the nature of their individuals.

The first subpopulations group is noted V_{ST}. It contains as individuals, the acoustic vectors $A_{vo_k\ m}$. This group is divided into sixteen demes according to the number of vowels and semivowels.

The second subpopulations group is noted S_A. It contains as individuals, the acoustic vectors $V_{vo_k\ m}$. This

group is divided into nine demes according to the number of stops and the affricate consonants.

The third group is noted F_N and contains as individuals, the acoustic vectors $V_{vo_k\ m}$. This group is divided into fifteen demes according to the number of fricative and nasal consonants

For each deme, basic GA evolves in parallel and without communication with the other GAs.

VI. RESULTS ANS DISCUSSIONS

Our approach is based on Multiple-deme parallel PEAs. The population is divided into a few subpopulations or demes, and each of these relatively large demes evolves separately on different processors in complete isolation. Without communication, the migration rate is zero This is the topology we have chosen to our PEA. For each phoneme class, we have a subpopulation or deme. For each deme, basic EGA evolves in parallel and without communication with the other EGAs. We have chosen uniform crossover operator because it combines any material of the parents and the ordering of genes is irrelevant. Its probability noted p_c is equal to 1.0. The probability uniform mutation noted p_m was fixed to zero. This means that too poor attribute values were modified in each recombination. These values are chosen empirically to improve each deme EGA by performing multiple executions and then choosing the best solution [10].

For each EGA, we have used tournament selection and an elitist generational population model without any overlap of the populations. Fixed length chromosomes were used relatively to the order of reference vector of each phoneme sound.

The experiment was conducted using a medium-sized corpus composed of 50 naturally spoken sentences of Standard Arabic continuous speech from Arabic normalized database. This corpus is spoken by several male Algerian speakers in low noisy environment. It contains hundreds of consonants (stops, fricatives, semivowels) which occurred in the three possible positions: initially, medially and finally. Speakers were asked to read a list of meaningful and grammatically correct sentences. The sentences were selected so that the number of consonants with the two voicing properties is fairly well balanced.

The results analysis shows that the MDPEA application to the speech classification domain is very interesting, since we have obtained good results. This evaluation was conducted in terms of two criteria: classification accuracy and computation time. The Corpus phonemes were extracted and identified successfully with an overall accuracy of 92.60% in different contexts. Approximately, there is similar identification accuracy between the same types of consonants. It means that we have used a good front-end signal processing used in our system to characterize, acoustically, the consonants.

The results show that an identification accuracy of vowels and semivowels which is equal to 94.50% is higher than the accuracy of the other phonemes. It can be explained by the steady character of their acoustic characteristics. Also, they show that the classification accuracy of stops which is equal to 89.3%, is the worse one because of their dynamic acoustic character.

The unidentified phonemes are essentially due to the segmentation difficulties and to the coarticulation phenomenon problems. In natural speech, there are no marked boundaries between acoustic data segments. Word and phoneme boundaries are non-existent. Even with expert labelling of the acoustic data, it is very difficult to establish a hard boundary between the phonemes and words that form an utterance.

VII. CONCLUSIONS

This paper shows a speech segments automatic identification method based on the Parallel Genetic Algorithms in order to recognize the Standard Arabic phonemes. We have used a corpus composed of 50 naturally spoken sentences of continuous speech from Arabic corpus recorded by several Algerian male speakers in low noisy environment.

We have formulated the segments automatic identification as a function optimization problem and have used an isolated multiple-deme parallel EGA. For each phoneme class, there is an isolated subpopulation or deme. For each one, basic EGA evolves in parallel with the decision rule Manhattan distance as the EGA evaluation or fitness function. Also, we have used three structure modification operators: uniform crossover, uniform mutation and tournament selection. The first difficult task is how to set up the EGAs parameters. Unfortunately, there is no unified guidance for this. Empirically, we set the crossover probability 1.0 and the mutation probability zero. In order to maximize the reproducibility and minimize the ambiguity in the EGA implementation, the deme population size was chosen to be dependant on the size chosen corpus. The binary tournament selection was used to increase the selection pressure, which allowed us to measure whether each crossover was able to keep the population diversity [20].

Our promising results show that our MDPEA is capable to precisely identify almost all instances. Despite its simplicity, it may give competitive performance compared to many other methods. By parallelising the algorithm, we have increased population size, reduced the computational cost, and so have improved the performance of the EGAs. We have compared these results with those of other promising machine learning approaches and we have found that our identification MDPEA is approximately as performing as the other known approaches.

VIII. REFERENCES

[1] M Kumar, DM Husain and N Upreti, "Genetic algorithm: Review and application," Available at SSRN: https://ssrn.com/abstract=3529843 or http://dx.doi.org/10.2139/ssrn.3529843, 2010.

[2] S. h. Al Ani, "Arabic Phonology an acoustical and physiological investigation," Edition Mouton. 1970.

[3] R. Vergin, & all, Generalized Mel frequency cepstral coefficients for large-vocabulary speaker-independent continuous-speech recognition," IEEE Transactions on Speech and Audio Processing, pp: 525–532, September. 1999.

[4] A.M. A. Ali, & all, "Acoustic-phonetic Features for the Automatic Classification of Stop Consonants," IEEE Transaction on Speech and Audio Processing, in press, 1999.

[5] M. Aissiou & M. Guerti, "Classification Génétique des Voyelles de l'Arabe Standard International," Conference on Control, Modelling and Diagnostic, Annaba, Algeria, pp : 75, 2006.

[6] O. Kramer. " Genetic Algorithms. SpringerIn: Genetic Algorithm Essentials. Studies in Computational Intelligence," vol. 679. Springer, Cham. https://doi.org/10.1007/978-3-319-52156-5_2, 2017

[7] D.E.Goldberg, "Design of Innovation: Lessons From and For Competent Genetic Algorithms," Kluwer, Boston, MA. 2002.

[8] E. Cantu_Paz, "A survey of parallel genetic algorithms. Calculateurs Paralléles. Reseaux et Systems Repartis," vol.10(2), pp :141-171. 1998.

[9] F.Rothlauf and F Rothlauf. "Representations for Genetic and Evolutionary Algorithms. In: Representations for Genetic and Evolutionary Algorithms," Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-32444-5-2.2006.

[10] F. Herrera, & all, "Taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study," International Journal of Intelligent Systems, vol.18, pp: 09–338. 2003.

Influence of Interface defects on Ideality Factor and Schottky barrier in Schottky Diodes based on SiC polytypes

Fayssal Mekaret[a], B. Zebentout[a], A. Rabehi[a,b,*],S. Tizi[a], Z. Benamara[a]
[a] Applied Microelectronics Laboratory (AMEL), Electronics Department, Faculty of Technology, DjillaliLiabes University of Sidi Bel Abbes, BP 89, 22000 Sidi Bel Abbes, Algeria
[b]University of Ziane Achor, Djelfa, 17000, Algeria
* Corresponding author: mekaretf@gmail.com.

Abstract

This paper presents an analytical simulation performed using MATLAB to investigate the characteristics of Schottky diodes based on three different SiC polytypes: 3C-SiC, 4H-SiC, and 6H-SiC.

The objective of this study is to examine the impact of two key factors—the thickness of the interfacial layer (*dox*) and the density of interface states (*Nss*) on the ideality factor and the Schottky barrier height.

The findings demonstrate a significant influence of these parameters on the Schottky barrier height, which, in turn, substantially affects the diode's current characteristics. By shedding light on the interplay between these factors, this study contributes to a deeper understanding of the performance and optimization of Schottky diodes for practical applications.

Keywords: SiC polytype, Schottky Barrier, interface states, layer thickness.

1. Introduction

Schottky diodes based on silicon carbide (SiC) polytypes, such as 3C-SiC, 4H-SiC, and 6H-SiC, have gained significant attention in high-power and high-frequency applications due to their exceptional electrical and thermal properties. However, the performance of these devices is often influenced by non-ideal factors, such as interface states and the thickness of the interfacial oxide layer**.** [1]

These factors can modify the Schottky barrier height and significantly affect the ideality factor, a critical parameter that reflects the quality of the metal-semiconductor junction. Understanding the impact of interface states and oxide layer thickness is essential for optimizing the rectifying behavior and enhancing the efficiency of SiC-based Schottky diodes.

Investigating these phenomena is crucial for the development of robust models that can predict performance under real-world conditions.

This study focuses on analyzing these effects across different SiC polytypes using advanced analytical and simulation methods to provide insights into device design and performance improvements.

By delving into the nuances of Nss (interface state density) and dox (oxide layer thickness) and their ramifications on device performance, this research contributes to the advancement of SiC-based electronic systems tailored for diverse applications.

The outcomes of this study not only enhance our understanding of Schottky diode behavior but also lay the groundwork for developing optimized designs with superior functionality and performance in diverse technological fields.

2. Simulation details

In terms of technology, Schottky diodes are manufactured with a moderately doped of n-SiC epitaxial layer and therefore the chosen doping type is Nd = 5×10^{15}cm^{-3}. In practice, the M/S interface often exhibits defects [2]**,** impurities, or roughness that can influence the barrier height and charge transport mechanisms as a result of several practical factors.

The real Schottky barrier Φ_{Breal} is influenced by the presence of an oxide layer (dox) between the metal and the semiconductor [3], The oxide layer can also contribute to the formation of interface states or traps, which can pin the Fermi level and influence the effective Schottky barrier height.

To investigate the effect of Nss (interface state density) and dox (oxide layer thickness) on the barrier height (SBH) and the ideality factor (n) of a Schottky diode based on the three SiC polytypes, a simulation is conducted using MATLAB tools at room temperature.

This simulation visualizes how the ideality factor's value changes with the values taken from Nss and dox.

$$\Phi_{B\,reel} = C_2\left(\Phi_M - \chi_{sc}\right) + (1 - C_2)\left(\frac{Eg}{q} - \Phi_0\right) - \Delta\Phi_B \quad (1)$$

$$C_2 = \frac{\varepsilon_{ox}}{\varepsilon_{ox} + Nss.d_{ox}.q^2} \quad (2)$$

$$n = 1 + \frac{dox}{\varepsilon_{ox}}\left[\frac{\varepsilon_{sc}}{W} + qN_{SS}\right] \quad (3)$$

where ε_{SC} and ε_{ox} are respectively the semiconductor and the oxyde layer permittivity, W is the width of the depletion region.

Due to the low intrinsic concentration of SiC materials, the generation-recombination current is negligible. Similarly, the diffusion current is also negligible because of

the low charge carrier mobility across all SiC polytypes. The low doping level chosen in the simulation further ensures that the tunneling current is insignificant. Consequently, the charge transport model considered in this study is the thermionic current [4].

$$I = I_S \left[\exp\left(\frac{qV - IR_S}{nkT}\right) - 1\right] + \frac{V - R_S \times I}{R_P} \qquad (4)$$

In this context, the series resistance R_s encompasses ohmic losses arising from the bulk resistance of the semiconductor material and the metallic contacts, R_p is the shunt resistance can indicate significant leakage that deteriorates the device performance [5].

The three SiC polytypes do not have the same electrical and physical parameters [6], which means that for optimal rectification quality, the choice of metal for forming the junction must be different for each SiC polytype.

In this study, copper is the metal chosen for 3C-SiC, aluminum for 4H-SiC, and titanium for 6H-SiC.

3. Results and discussions

To understand the impact of interface states on the real Schottky barrier height, a MATLAB script was developed for the three polytypes to plot the real SBH for a range of N_{SS} values from 10^{11} to 5×10^{13} $cm^{-2}eV^{-1}$ (Fig.1) while the thickness of interface layer is 20 Å and for range of dox values from 20 to 50Å with Nss=5×10^{-12} cm^{-2} eV^{-1}.

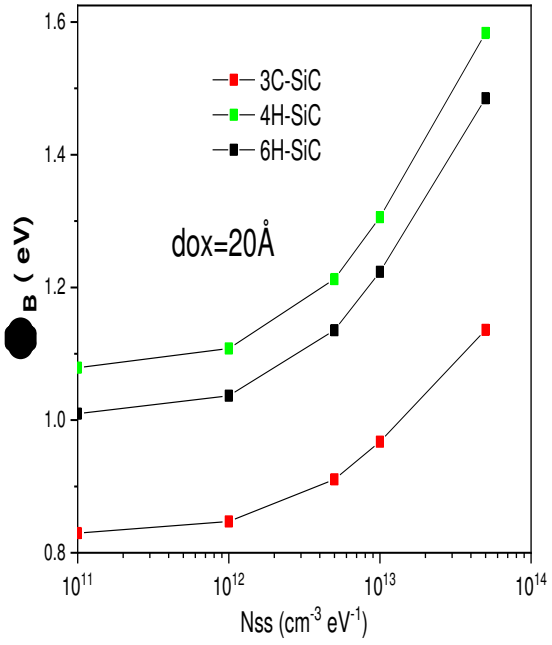


Fig 1: Evolution of the real Schottky barrier Height (SBH) with N_{SS} values.

The observed increase in the Schottky barrier height (SBH) with an increase in the number of surface states (N_{SS}) for the three polytypes can be attributed to the modulation of the metal-semiconductor interface properties.

As N_{SS} increases, there is a higher density of localized states at the interface, which can pin the Fermi level more strongly. This Fermi level pinning effect results in a shift of the effective barrier height. Essentially, the presence of these surface states leads to a redistribution of charge, influencing the overall energy band alignment and thus increasing the Schottky barrier height for all polytypes.

This trend highlights the critical role of surface states in defining the electronic properties of metal-semiconductor interfaces, which is consistent across different polytypes.

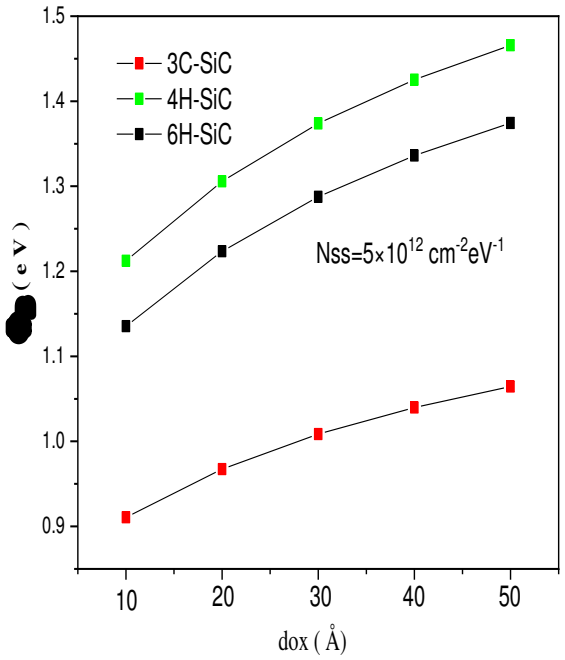


Fig 2: Evolution of the real Schottky barrier Height (SBH) with d_{OX} values.

Based on Fig 2 The observed increase in the Schottky barrier height with the thickness of the interfacial layer (dox) can be attributed to the influence of the insulating layer on the metal-semiconductor interface properties [3]. As the interfacial layer thickens, it creates an additional potential barrier that modifies the energy band alignment at the interface. This layer reduces the direct interaction between the metal and the semiconductor, suppressing the effect of Fermi-level pinning and allowing the barrier height to be more strongly influenced by the intrinsic properties of the interfacial layer and the semiconductor [7].
Furthermore, the thicker oxide layer can reduce tunneling effects by increasing the effective width of the potential barrier, leading to a more pronounced Schottky barrier. This

behavior highlights the critical role of the interfacial layer in engineering the electronic properties of the metal-semiconductor junction [8], particularly in controlling the barrier height for optimized device performance.

The increase in the ideality factor with oxide layer thickness can be attributed to the enhanced interface states and the increased barrier inhomogeneities introduced by a thicker oxide layer. These inhomogeneities and interface states facilitate non-ideal current transport mechanisms, such as tunneling and recombination, further deviating from the ideal diode behavior.

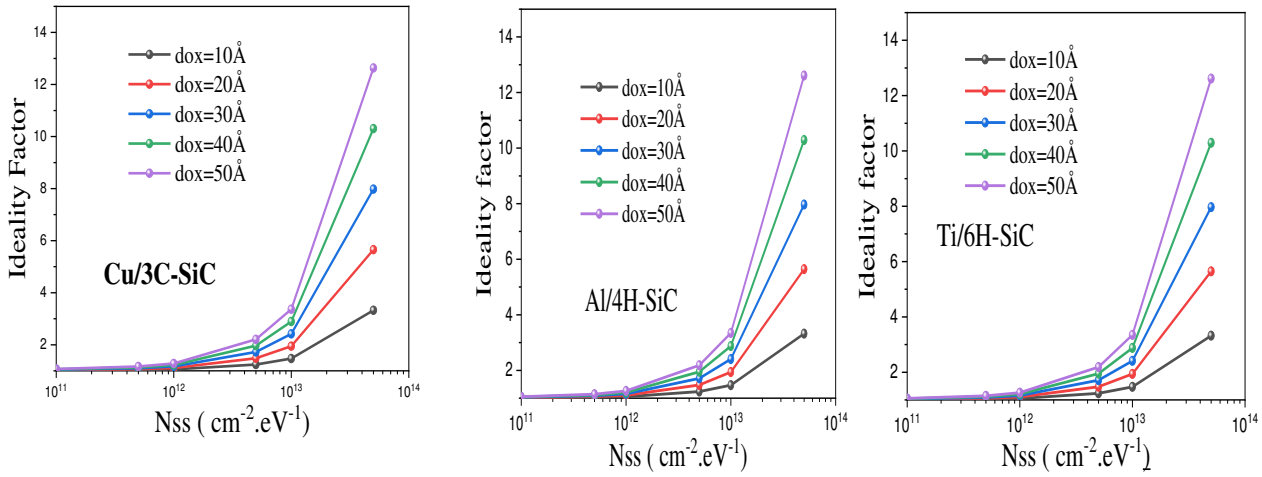


Fig 3: Variation of ideality factor as a function of interface states density N_{SS} for different values of interface layer dox.

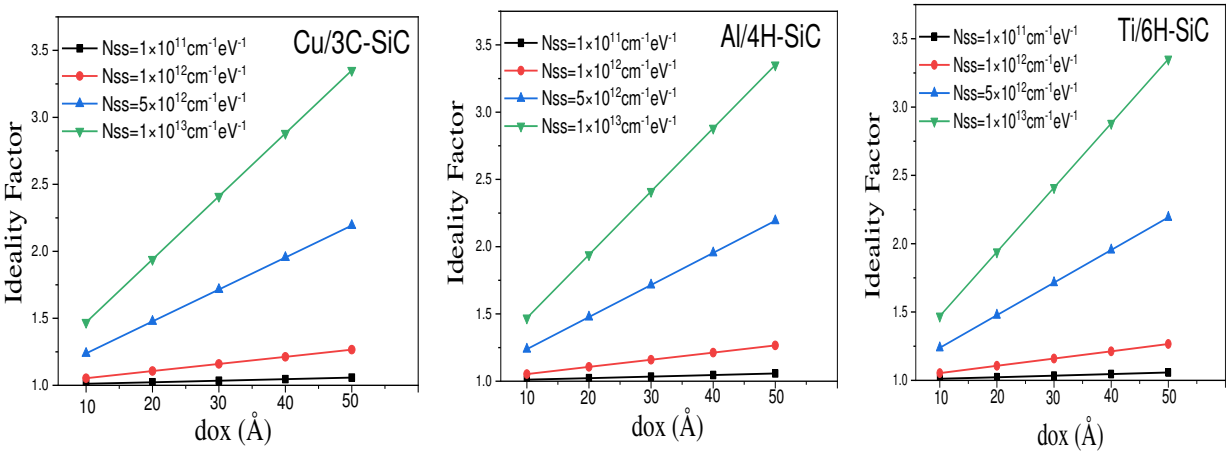


Fig 4: Variation of ideality factor as a function of interface layer dox for different values of interface states density N_{SS}

Based on Fig 3 and 4, it is evident that the thickness of the oxide layer (dox) has a significant influence on the ideality factor of the diode. Specifically, for oxide thicknesses below 10 Å, the ideality factor remains relatively low and does not exceed a value of 3. However, as the oxide thickness increases to 50 Å, the ideality factor becomes significantly larger,

exceeding 10. This trend is consistent across all three polytypes.

Conclusion

In this study, we have thoroughly analyzed the impact of surface states on the Schottky barrier height and the ideality factor for different polytypes. Our findings reveal that an increase in the number of surface states (Nss) leads to a significant increase in the Schottky barrier height across all polytypes, which can be attributed to the enhanced Fermi level pinning effect. Furthermore, the ideality factor was shown to increase with rising NSS, indicating a pronounced deviation from ideal diode behavior. This non-ideal behavior results from enhanced recombination and tunneling effects at the metal-semiconductor interface, which are exacerbated by the presence of surface states.
These results underscore the critical role that surface states play in influencing the electronic properties of metal-semiconductor interfaces, impacting both energy band alignment and current transport mechanisms. Our work highlights the importance of careful interface engineering to optimize the performance of Schottky diodes, especially in applications where minimizing non-ideal behavior is crucial. Future research should focus on developing methods to control or mitigate the effects of surface states to achieve more reliable and efficient semiconductor devices

References

[1] T. Kimoto , J. Cooper . Fundamentals of silicon Carbride Technology : Growth, Characterisation, Devices, and Applications. 1st éd., John Wiley & Sons 2014.

[2] M.E. Levinsthtein , S.L.Rumyantsev andM.S. Shur, Properties of advanced semiconductor materials GaN, AlN, InN, BN, SiC, SiGe. 2001.

[3] S.M. Sze, Y. Li, K.K. Ng, Physics of Semiconductor Device, John wiley& sons, 2021.

[4] A. Rabehi, Study of the characteristics current-voltage and capacitancevoltage in nitride GaAs Schottky diode. Eur. Phys. J. Appl. Phys, vol. 72, 2015,https://doi.org/10.1051/epjap/2015150

[5] A. Bekaddour, A. Rabehi, S. Tizi, B. Zebentout, B. Akkal, Z. Benamara, .Effect of the contact area on the electrical characteristics of the Ti/6H–SiC (n) Schottky diode. Micro and Nanostructures, vol. 173, 2022.

[6] Helal H, Benamara Z, Arbia MB, Rabehi A, Chaouche AC, Maaref H. Electrical behavior of n-GaAs based Schottky diode for different contacts: Temperature dependence of current-voltage. Int J Numer Model .2021.

[7] Nagaraj M. K1, Munikrishna Reddy, Interface states and electrical properties of Cu/Ni Schottky contacts on n-InP substrate, International Journal of Advanced Scientific Technologies in Engineering and Management Sciences Volume.2,Issue.7,July.2016

[8] Dean k, Bruno C, on the role of unterface charge in non ideal metal-semiconductors contact, applied surface science ,250,63-69(2005)

Obstacle Avoiding firefieghting car-like robot

Messaouda Benzaoui
University of Mhamed Bougara
Automatic Applied and Signal
Processing Laboratory
Boumerd· s Algeria
m.benzaoui@univ-boumerdes.dz

ZIDANI Chahinez
Institute of Electrical and
Electronic Engineering
University of Mhamed Bougara
Boumerd· s Algeria

BOUSSAID Inas
Institute of Electrical and
Electronic Engineering
University of Mhamed Bougara
Boumerd· s Algeria
inasboussaid30@gmail.com

Abstract´ the idea of this paper is to improve emergency response capabilities, and ensure the safety of both firefighters and the communities they serve. it is about a smart car system consists of two major parts: The first one is Firefighting system which depends on a flame sensor that detects the presence of flames, allowing the vehicle to swiftly respond to fire incidents by activating fire suppression systems. The second is Obstacle avoiding system. It consists of a main sensor that is ultra-sonic, it detects obstacles by emitting sound waves and measuring their reflections, enabling the vehicle to navigate safely by avoiding collisions and adjusting its path as needed. The whole system is managed with an Arduino UNO microcontroller, relay module, water pump, and power battery. This system is implemented and tested. Results shows that it works properly as expected.

Keywords´ Firefighting robot, Obstacle avoidance, flame sensor, Arduino UNO microcontroller.

I. INTRODUCTION

Firefighting is a dangerous job that requires split-second decisions, quick reflexes. Firefighters often face smoke, flames, and debris that can block their path to victims or hinder their escape. Since robots are introduced in various industries in the recent years. The use of robots are actively carried out to minimize firefighters˘ injuries and deaths as well as increasing productivity, safety, efficiency and quality of the task given. In response, to find ways and to put an end for these disasters and their consequences. The use of `a fire robot_ may help to save lives by navigating fires and obstacles that would be difficult or dangerous for humans.

Firefighting robot plays a crucial role in daily life by enhancing the efficiency and effectiveness of firefighting operations. The vehicles are equipped with advanced technology and automated systems that enable them to swiftly respond to fire emergencies with minimal human intervention. By autonomously detecting, extinguishing, and containing fires, that can reduce response times[1-10]. In recent years, different types of vehicles for firefighting was burned: rover tank [2], forest fires [11], fire bird robot [12], ..etc.

Additionally, automatic fire-fighting cars can access hazardous environments that may be unsafe for human firefighters, allowing them to tackle fires in challenging conditions more effectively [13-15].

Building this automatic fire fighting car-like involves electronics and coding. But the result is quite impressive! With further improvements, this technology could be used to build autonomous firefighting robots that can assist firefighters in dangerous situations.

The brain for robot is the microcontroller, which exists in several forms. Controllers or microcomputers can be the Arduino microcontroller, Peripheral Interface Controller (PIC), and raspberry Pi microcomputer. The choice depends the speed of processing of the data, the design cost or the number of sensors connected to the robot[2].

This paper is organized as follows: Section 2 delivers the hardware and the software description of the designed robot used. Section 3 presents the implementation of the robot and the working principle of the proposed robot. Also, a presentation for the obtained results robot. Finally, Section 4 shows the concludes of this paper.

II. HARDWARE & SOFTWARE DESIGN

This work involves building obstacle avoiding and firefighting robot car, using Arduino Uno board as microcontroller. while a second board to DC motor, an ultrasonic sensor to prevent collisions with walls other obstacles and finally flame sensor to detect the fire .

A. Hardware

This session presents the hardware development. It includes the several types of sensors, DC motor, microcontroller, Transmitter and Water pump.

1) Arduino Uno board which is based on ATmega328p microcontroller.

The term Arduino contains:

- The Arduino IDE (PC software)
- The Arduino board (hardware)
- The Arduino firmware (microcontroller software)

The microcontroller on the Arduino Uno R3 board is the ATmega328P that is a low power and low cost 8-bit microcontroller, which serves as the main processing unit. It runs the user's Arduino sketches and interacts with various

XXX-X-XXXX-XXXX-X/XX/$XX.00 ø 20XX IEEE

peripherals and sensors connected to the board. The ATmega328P has flash memory for storing the user's program (32KB), RAM for runtime data storage(2KB), and EEPROM for non-volatile data storage(1KB). It also manages serial communication with the computer through the USB-to-serial converter chip.

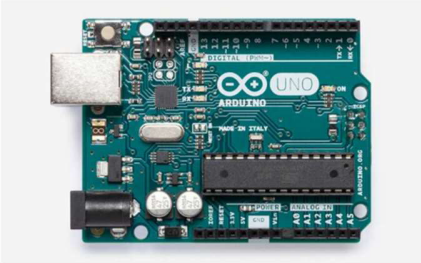


Fig1. Arduino Uno R3 board

2) DC motor

A dc motor is a fundamental electromechanical device in mechatronic systems that plays an important role in maintaining accuracy in task execution in Automotive Industry, Robotics, Household Appliances ŭ ect. This DC or direct current motor works on the principal, when a current carrying conductor is placed in a magnetic field; it experiences a torque and tends to move.



Fig2. DC motor

3) Motor driver

The motor driver plays a critical role in robotics and mechatronics projects. It functions as an amplifier, converting low current signals into high currents necessary for driving motors. Acting as an interface between the microcontroller and the DC motor, it resolves issues where the microcontroller output falls short. For instance, if a motor requires 12 volts and draws 1 ampere, but the Arduino Uno can only supply 5 volts and 20 milliamperes, this output is insufficient and could potentially damage the microcontroller. Hence, the motor driver becomes indispensable in such scenarios.

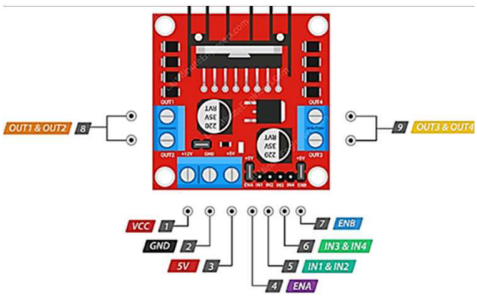


Fig3. L298N motor driver

As shown in fig.3, the L298N motor driver is high voltage high current transistor based on dual H-bridge driver which means this driver can operate two motors separately. all control inputs are gives in pin header form at bottom side which contains four direction pins (IN1, IN2, IN3,IN4)along with two enable pins (ENA,ENB) , in the following driver output channels For the motors are broken out to the edge of the module with two screw terminals (OUT1,OUT2 & OUT3,OUT4).

The L298N motor driver powered through three pins screw terminals it consists of pins motor power followed by ground and 5 volts logic power supply.

4) Servo motor

Servo motor is part of closed-loop system which is a self-contained electrical device, that rotate parts of a machine with high efficiency and with great precision. it provides a precise control on the movement of its drive shaft; it might be the best choice. Internally a Servo motor is divided into two main sections: mechanical section, which contains gear that produce a low-speed high torque output electronic section: Servo Motors uses potentiometer as feedback sensor that sends signal to PCB about the current position of Servo motor.

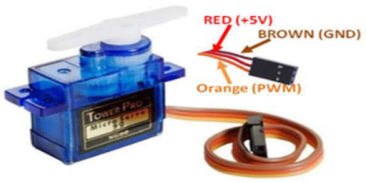


Fig4. L298N motor driver

5) Ultrasonic sensor

An ultrasonic sensor is a device that uses high frequency sound waves to detect and locate objects. Ultrasonic sensors work by emitting a high frequency sound wave that bounces off nearby objects and returns to the sensor. The sensor then measures the time it takes for the wave to return, giving the distance to that object. It operates at frequencies higher than 20 kHz, which is beyond the range of human hearing.

Because firefighting robots should be able to sense things with enough limits to let robots to react and travel appropriately. The existing sensors ultrasonic sensors suit more. These sensors have several advantages: they are low cost, have a wide detection range from a few centimeters up to 10 meters, and can detect both solid and soft objects [1]. The basic components of an ultrasonic sensor are an emitter, a receiver, and a control circuit. The emitter produces high frequency sound waves that propagate outward in a beam. When these waves hit an object, some of the waves reflect towards the receiver. The receiver then detects these reflected waves and converts them into electrical signals. The control circuit analyzes these signals to determine the distance to the object.

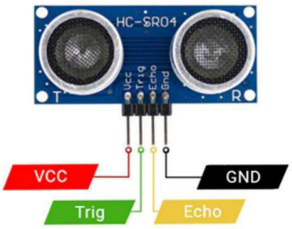


Fig5. Ultrasonic sensor

6) Flame sensor

It is a device used to detect and respond to the presence of a flame or fire. The most common type of flame sensor is the infrared flame detector. Infrared flame detectors work by detecting infrared radiation emitted by flames. Infrared flame detectors typically consist of an infrared sensor that is sensitive to the specific wavelengths of infrared light emitted by flames. When a flame enters the sensor's field of view, the infrared radiation emitted by the flame is detected. This triggers an alarm condition, which can activate a fire suppression system or other safety response. In firefighting robots, fire sensors can be considered as robot eyes to discover sources of fire. It can be utilized to identify fire based on wavelength of the light [1]. Flame sensor has two signal: a digital Output gives two kind of information that it˘s has flame or non-flame, and Analog Output pins will detect exact wavelength of different light.

Inside the sensor is a special filter that only allows infrared wavelengths associated with flames to reach the sensor element. This helps eliminate false alarms from other infrared sources like sunlight. The sensor element is typically made of a material that generates an electrical current when exposed to infrared radiation. The larger the flame, the infrared radiation is detected and the greater the electrical current produced. Flame sensors provide a fast and reliable way to detect flames and activate safety systems.

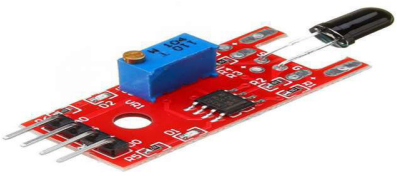


Fig6. Flam sensor

7) Water pump

A 5V water pump is a small water pump that runs on 5 volts of DC power. They are commonly used in aquariums, fountains, and other small water circulation systems. 5V pumps are typically very small, with pumps heads ranging from just a few centimeters up to around 15 centimeters. Since they run on 5 volts, 5V water pumps require a low-power DC power source like a USB adapter or battery pack. They are designed to pump relatively small volumes of water, usually measured in liters per hour rather than gallons per minute like larger pumps.



Fig7. Water pump

8) Relay module

A relay is an electrically operated switch that can be turned on or off, letting the current go through or not, and can be controlled with low voltages, like the 5V provided by the Arduino pins. It is used here, to control the water pump (turn it on /off).



Fig8. Relay module

B. Softwore

Arduino IDE (Integrated Development Environment) is an open-source software platform used to write, compile, and upload code to Arduino microcontroller boards developed by Arduino.cc. It consists of many libraries and a set of examples, it supports C and C++ programming language .it used to get started with electronics programming and robotics, and build interactive prototypes.

Programs written using Arduino Software (IDE) are called sketches. These sketches are written in the text editor and are saved with the file extension.

Arduino programs can be divided in three main parts structure, values (variables and constants), and functions.

Let take in consideration the structure which consists of two main functions:

Setup () function:

The setup () function is called when a sketch starts. It is used to initialize the variables, pin modes, start using libraries, etc. The setup function will only run once, after each power up or reset of the Arduino board.

Loop () function:

The loop function, integral to Arduino programming, operates continuously after the setup function. Its primary role is to execute the main program logic repeatedly until the Arduino

board is powered off or reset. Within this function, tasks such as reading inputs, processing data, and controlling outputs are performed based on the defined program instructions.

III. IMPLEMENTATION

The purpose here, is to present the proposed system˘s hardware design and implementation. This provides a full explanation of the system˘s components, their functions, and how they are linked together to make the system. Also, cover the process of assembling the hardware and configuring the system, as well as show the results of any tests performed to ensure the system is working properly.

The paper uses a flame sensor when the fire is detected, the car goes towards the desired place using Ultrasonic sensor placed on the front of the car to detect any obstacles in its way.

When an impediment is detected, the car avoids it by changing direction using servo motor and continues moving towards the fire source.

A. Design the Hardware Architecture

The robot car˘s hardware architecture incorporates a robust system for obstacle avoidance and fire-fighting capabilities. For obstacle avoidance, ultrasonic sensors are strategically placed at the front of the car to detect obstacles within the vicinity. This sensor provides real-time distance measurements to the microcontroller, which processes the data and adjusts the car˘s path by controlling the motors through motor drivers. This ensures the car can navigate around obstacles smoothly and efficiently. And this part is represented in the Following circuit Fig9:

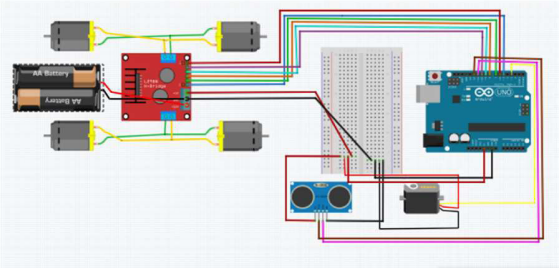


Fig9. Obstacle avoidance circuit

For firefighting part, flame sensors are installed to detect heat sources that signify the presence of fire. Upon detection, the microcontroller activates the fire extinguishing mechanism, which includes a water pump. The combination of these systems allows the robot car to autonomously navigate its environment while effectively addressing fire hazards. We combine the firefighting to the previous circuit as shown below:

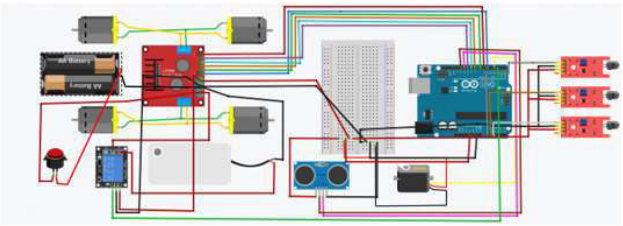


Fig10. Interfacing of the overall circuit

We develop the Arduino sketch (code) to control the car's behavior. This involves programming the Arduino to:

- ¿ Read data from the flame sensor module to detect the presence of fire.

Read data from the ultrasonic sensor module to detect obstacles.

- ¿ Make decisions based on sensor data to navigate toward the fire source and avoid obstacles.
- ¿ Control the motors to move the car and adjust its direction.
- ¿ Activate the water pump to extinguish the fire.

The overall code with some comments is exhibited on figure: fig11.

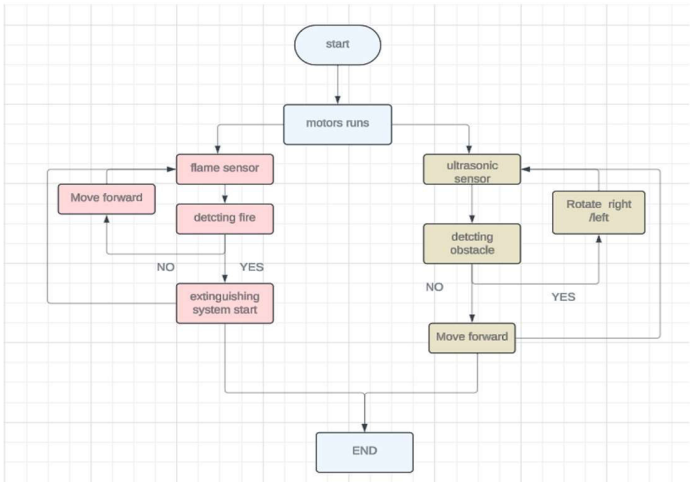


Fig11. General flowchart

B. Results and functional testing

The implementation yielded promising results, with the obstacle-avoiding and fire-fighting smart car demonstrating robust functionality during testing. The robot's obstacle avoidance system performed admirably, utilizing the ultrasonic sensor to accurately detect nearby objects and obstacles. By emitting sound waves and analyzing their reflections, the car autonomously adjusted its trajectory to navigate through complex environments, showcasing its ability to maneuver safely and effectively. Additionally, the fire detection system proved to be highly reliable, promptly identifying flames using the flame sensor and triggering the activation of the fire extinguishing mechanism. This swift response to fire incidents highlights the system's effectiveness in mitigating potential hazards and ensuring the safety of both the robot and its surroundings.

Throughout the development process, we encountered various obstacles that tested our problem-solving abilities and technical expertise. One significant challenge arose when the flame sensor initially struggled to reliably detect flames, leading to delays in the fire detection process. This issue required extensive troubleshooting and adjustment of sensor parameters to enhance its sensitivity and accuracy. Additionally, we faced challenges with the motor driver module, encountering issues with motor control and coordination. Fine-tuning the motor driver settings and ensuring proper communication between the

microcontroller and motors proved to be a complex task. Furthermore, integration challenges between hardware components occasionally disrupted system functionality, requiring thorough analysis and adjustments to optimize performance. Despite these obstacles, collaborative problem-solving efforts and persistent experimentation enabled us to

```
void loop() {
    // Measure distance using the ultrasonic sensor
    int distance = mySensor.ping_cm();
    // Read the flame sensor values
    int leftSensorValue = analogRead(leftFlameSensorPin);
    int centerSensorValue = analogRead(centerFlameSensorPin);
    int rightSensorValue = analogRead(rightFlameSensorPin);
    // Check if there is a fire within the center flame threshold
    if (centerSensorValue < flameThreshold) {
    // Stop the car and activate the water pump
         rotateMotor(0, 0);
        digitalWrite(waterPumpRelayPin, HIGH);
        delay(3000); // Keep the pump on for 3 seconds (adjust as needed)
        digitalWrite(waterPumpRelayPin, LOW);}
    // Check if there is a fire within the left flame threshold
    else if(rightSensorValue < flameThreshold) {
        // Stop the car and activate the water pump
        rotateMotor(0, 0);
        digitalWrite(waterPumpRelayPin, HIGH);
        delay(3000); // Keep the pump on for 3 seconds (adjust as needed)
        digitalWrite(waterPumpRelayPin, LOW);

    // Read right side distance using the ultrasonic sensor
    int distanceRight = mySensor.ping_cm();
    // Bring servo to center
    myServo.write(90);
    delay(500);
     if (distanceLeft == 0) {
    // Turn the car right
      rotateMotor(ADJUST_SPEED, -ADJUST_SPEED);
       delay(200);
    } else if (distanceRight == 0) {
    // Turn the car left
        rotateMotor(-ADJUST_SPEED, ADJUST_SPEED);
        delay(200);
    } else if (distanceLeft >= distanceRight) {
    // Turn the car right
        rotateMotor(ADJUST_SPEED, -ADJUST_SPEED);
        delay(200);
    } else {
    // Turn the car left
    rotateMotor(-ADJUST_SPEED, ADJUST_SPEED);
    delay(200);
    }
  else {
    // Move the car forward
    rotateMotor(REGULAR_SPEED, REGULAR_SPEED);}}

  //Motor Control Function:
  //rotateMotor(): Controls the speed and direction of each motor.
  //Uses digital writes to control the direction of the motors and
  void rotateMotor(int rightMotorSpeed, int leftMotorSpeed) {
      if (rightMotorSpeed < 0) {
          digitalWrite(rightMotorPin1, LOW);
          digitalWrite(rightMotorPin2, HIGH);
      } else if (rightMotorSpeed >= 0) {
          digitalWrite(rightMotorPin1, HIGH);
  }       digitalWrite(rightMotorPin2, LOW);
      }

      if (leftMotorSpeed < 0) {
          digitalWrite(leftMotorPin1, LOW);
          digitalWrite(leftMotorPin2, HIGH);
      } else if (leftMotorSpeed >= 0) {
          digitalWrite(leftMotorPin1, HIGH);
          digitalWrite(leftMotorPin2, LOW);
      }

      analogWrite(enableRightMotor, abs(rightMotorSpeed));
      analogWrite(enableLeftMotor, abs(leftMotorSpeed));
```

Fig12. Overall code of robot

overcome these challenges and refine our system to achieve the desired functionality. Each challenge presented valuable learning opportunities, allowing us to deepen our understanding of hardware-software integration and refine our engineering skills.

Integrating a camera, Bluetooth communication, GPS, and an alarm system into the circuit can significantly enhance its capabilities and make it more effective in detecting and responding to fire incidents. Here's how you can incorporate these features into your system:

¿ Camera for Fire Action Recording:

Integrate a camera module capable of capturing high-quality video footage.

Implement algorithms for fire detection using image processing techniques.

Configure the camera to record video footage when a fire is detected, providing visual documentation of the incident.

¿ Bluetooth Communication for Stakeholder Notification:

Add a Bluetooth module to our system for wireless communication.

Develop a notification system that sends messages to stakeholders (e.g., building occupants, emergency services) when a fire is detected.

Include relevant information such as the location of the fire, timestamp, and instructions for evacuation or response.

¿ GPS for Fire Location Reporting:

Integrate a GPS module to determine the robot's location accurately.

Develop algorithms to detect and localize the source of the fire using sensor data and environmental cues.

Send the precise location of the fire to stakeholders or emergency services via Bluetooth or other communication channels.

¿ Alarm System for Immediate Alerting:

Implement an alarm system that triggers a loud sound or visual alert when a fire is detected.

Design the alarm system to be highly noticeable and attention-grabbing to prompt immediate action from nearby individuals.

Ensure the alarm system is integrated seamlessly with the fire detection and notification mechanisms.

By integrating these features, we can create a comprehensive fire detection and response system that not only detects fires but also communicates with stakeholders, reports the fire's location, and alerts individuals to take immediate action. Ensure thorough testing and validation of the integrated system to verify its functionality and reliability in real-world scenarios.

IV. CONCLUSION

Firefighting is a dangerous job that requires split-second decisions, quick reflexes. Firefighters often face smoke, flames, and debris that can block their path to victims or hinder their escape. In response, we decided to start this work to find ways to put an end to these disasters and their consequences by a firebot that may help save lives by navigating fires and obstacles that would be difficult or dangerous for humans.

This work plays a crucial role in daily life by enhancing the efficiency and effectiveness of firefighting operations. These vehicles are equipped with advanced technology and automated systems that enable them to swiftly respond to fire emergencies with minimal human intervention. By autonomously detecting, extinguishing, and containing fires, that can reduce response times. Additionally, automatic fire-fighting cars can access hazardous environments that may be unsafe for human firefighters, allowing them to tackle fires in challenging conditions more effectively.

The car is equipped with sensors like flame sensors, and ultrasonic sensors. The flame sensor detects the presence of fire. Once a fire is detected, the car navigates towards the source using an Ultrasonic sensor placed on the front of the car to detect any obstacles in its path. When an obstacle is detected, the car avoids it by changing direction and continues moving towards the fire source.

It is also equipped with a water pump and nozzle that is activated once the car reaches the fire source. The nozzle sprays water to put out the fire.

Building this automatic fire fighting car involves electronics and coding. But the result is quite impressive! With further improvements, this technology could be used to build autonomous firefighting robots that can assist firefighters in dangerous situations.

REFERENCES

[1] Amano, H. (2002, August). Present status and problems of fire fighting robots. In Proceedings of the 41st SICE Annual Conference. SICE 2002. (Vol. 2, pp. 880-885). IEEE.

[2] Aliff, Mohd, et al. "Development of fire fighting robot (QROB)." International Journal of Advanced Computer Science and Applications 10.1 (2019).

[3] MURAD, Ahmed, BAYAT, OÄŸuz, et MARHOON, Hamzah M. Implementation of rover tank firefighting robot for closed areas based on arduino microcontroller. 2021.

[4] KADAM, Kirti, BIDKAR, Ayush, PIMPALE, Vaishnavi, et al. Fire fighting robot. International Journal of Engineering and Computer Science, 2018, vol. 7, no 1, p. 23383-23485.

[5] RAJU, Joyal, MOHAMMED, S. Sheik, PAUL, Johaan Varkey, et al. Development and implementation of arduino microcontroller based dual mode fire extinguishing robot. In : 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS). IEEE, 2017. p. 1-4.

[6] RAKIB, Tawfiqur et SARKAR, MA Rashid. Design and fabrication of an autonomous fire fighting robot with multisensor fire detection using PID controller. In : 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV). IEEE, 2016. p. 909-914.

[7] TAHA, Ihsan A. et MARHOON, Hamzah M. Implementation of controlled robot for fire detection and extinguish to closed areas based on Arduino. TELKOMNIKA (Telecommunication Computing Electronics and Control), 2018, vol. 16, no 2, p. 654-664.

[8] SUCUOGLU, H. Saygin, BOGREKCI, Ismail, et DEMIRCIOGLU, Pinar. Development of mobile robot with sensor fusion fire detection unit. IFAC-PapersOnLine, 2018, vol. 51, no 30, p. 430-435.

[9] ZHANG, Shuo, YAO, Jiantao, WANG, Ruochao, et al. Design of intelligent fire-fighting robot based on multi-sensor fusion and experimental study on fire scene patrol. Robotics and Autonomous Systems, 2022, vol. 154, p. 104122.

[10] ZHANG, JunJie, YE, ZiYang, et LI, KaiFeng. Multi-sensor information fusion detection system for fire robot through back propagation neural network. Plos one, 2020, vol. 15, no 7, p. e0236482.

[11] R. N. Haksar and M. Schwager, Distributed Deep Reinforcement Learning for Fighting Forest Fires with a Network of Aerial Robots, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 1067-1074.

[12] Rao DC, Kabat MR, Das PK, et al. Hybrid IWD-DE: A Novel Approach to Model Cooperative Navigation Planning for Multi-robot in Unknown Dynamic Environment. Journal of Bionic Engineering, 2019, 16 (2), pp. 235¯252.

[13] ANH, Phung Quang, DUC CHUNG, Tran, TUAN, Tran, et al. Design and Development of an Obstacle Avoidance Mobile-controlled Robot. In : 2019 IEEE student conference on research and development (SCOReD). IEEE, 2019. p. 90-94.

[14] ALLI, Kolapo Sulaimon, ONIBONOJE, Moses Oluwafemi, OLUWOLE, Akinola S., et al. Development of an Arduino-based obstacle avoidance robotic system for an unmanned vehicle. ARPN Journal of Engineering and Applied Sciences, 2018, vol. 13, no 3, p. 1-7.

[15] BABALLE, Muhammad Ahmad, BELLO, Mukhtar Ibrahim, AYAGI, Shehu Hassan, et al. Obstacle Avoidance Robot using an ultrasonic Sensor with Arduino Uno. Journal homepage: https://gjrpublication. com/gjrecs, 2023, vol. 3, no 05.

Evaluating Machine Learning Models For Intrusion Classification In Industrial Control Systems

1st ANIBA Hicham
Laboratory of Innovative Technology (LTI)
National School of Advanced Technologies
Algiers, Algeria
h_aniba@ensta.edu.dz

2nd BOUTERFAS Malika
Laboratory of Innovative Technology (LTI)
National School of Advanced Technologies
Algiers, Algeria
malika.bouterfas@ensta.edu.dz

3rd CHERIFI Tarek
Laboratory of Innovative Technology (LTI)
National School of Advanced Technologies
Algiers, Algeria
tarek.cherifi@ensta.edu.dz

***Abstract*—As cyber threats to industrial controls systems (ICT) evolve , effective intrusion detection and classification mechanisms are needed for protecting critical infrastructure. This study presents a comparative analysis of deep learning (DL) and traditional machine learning (ML) models for intrusion classification using a public Modbus TCP/IP communication data from industrial control systems. We evaluate the performance of XGBoost, Random Forest, Decision Tree, Multilayer Perceptron (MLP), and Recurrent Neural Networks (RNN) models based on key metrics, including precision, recall, and F1-score. the results shows that traditional ML models outperformed deep learning models in classification accuracy and efficiency.**

***Index Terms*—intrusion Detection System (IDS) , Power Grid Cybersecurity, Modbus TCP/IP, Machine Learning.**

I. Introduction

The rapid adoption of the Industry 4.0 has led to the interconnection of industrial machines and devices through industrial field networks and internet [1].Industrial Control Systems (ICS) integrate Information Technology (IT) and Operational Technology (OT) and play a vital role in monitoring and controlling critical infrastructure such as power grids offering benefits in terms of efficiency and automation[2, 3]. The increasing number of cyberattacks targeting ICS, including prominent incidents like the Stuxnet attack on Iran's nuclear facilities [4], and the ransomware attack on the American oil pipeline [5], highlights the significant economic and operational consequences of such threats. Therefore implementing robust cybersecurity defense mechanisms are essential for ICS environments [6]. One of the key strategies for protecting these systems is the use of Intrusion Detection Systems (IDS). IDS continuously monitor network traffic and system activities to detect anomalies that may indicate malicious behavior [7, 3] Recently, due to the increasing complexity of cyber threats the use of advanced detection methods involving machine learning (ML) and deep learning (DL) techniques has gained traction in research [8].

Prior work has studied various machine learning techniques for ICS intrusion detection.Authors of the dataset [6] compared between the performance of support vector machine (SVM) , Isolation Forest (IF) , Neural Network (NN) and One-Class SVM. Ning et al [9] proposed the use of the generative adversarial network (GAN) for synthetic data generation and the deep neural network (DNN) model for intrusion detection.Authors in [3] introduced a DNN with multi-attention blocks for capturing core features, and with residual blocks to avoid gradient vanishing problem. More recent work in [10] implemented a CNN with and without data augmentation through GAN. although the mentioned studies achieves a good detection performance , they introduce the use of complex, high-computation approaches that require extensive time and. data manipulation. This study aims to compare the effectiveness of stand-alone traditional ML models and DL models for intrusion classification in ICS systems using the original Modbus TCP/IP communications dataset [6].

The performance of extreme gradient boosting (XGBoost), Random Forest (RF), Decision Tree (DT), Multi layer perception (MLP), and Recurrent Neural Networks(RNN) models is evaluated using metrics such as precision, recall, and F1-score, examining the effectiveness of each model for this intrusion classification problem.

II. Methodology

A. Electra Modbus Dataset

The electra Modbus dataset [6] used in this study was generated during normal and abnormal operation using a testbed that models the behavior of an electric traction substation used in a real high-speed railway area. the testbed shown in figure 1is composed of 5 PLCS (1 master PLC and 4 slave PLCs) , a SCADA system , a firewall, a switch, some ICS devices, and a device that can launch man-in-the-middle (MitM) attacks. The SCADA system consists of a Nanobox (A1) and an HMI (A4) and acts as a master of both Modbus slaves A2 and A3.

Table I summaries the number of samples contained in the dataset after removing duplicates. The resulted Modbus dataset contains 11 features which are described in Table II

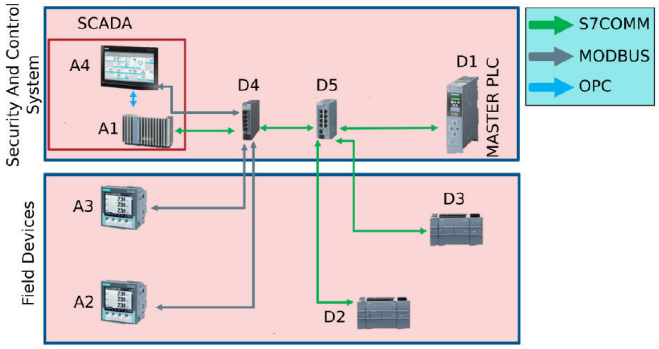


Fig. 1. Industrial testbed used to collect network traffic

TABLE I
ELECTRA SUBSET SAMPLES DISTRIBUTION

Subset	Normal Samples	Anomalous Samples	Total Samples
Modbus	24409	17299	41708

The dataset contains 41708 samples labeled as normal traffic and 7 types of attacks launched at the testbed , table III present the different ratios of each attack classes with in the data.

1) Preprocessing Steps: Before applying the models we ensure the quality and integrity of the dataset, a certain preprocessing steps are conducted.Firstly duplicate samples were identified and removed to avoid bias in model training and evaluation.Then label encoding was employed to deal with categorical features.After that all numerical features, including those resulting from label encoding, were scaled using Min-Max scaling 1.Finally we convert attack categories into a numeric values through label encoding or training the models.

$$X_{\text{scaled}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{1}$$

where:

- X is the original feature value,
- X_{min} is the minimum value of the feature in the dataset,
- X_{max} is the maximum value of the feature in the dataset,
- X_{scaled} is the normalized feature value, scaled between 0 and 1.

To ensure proper evaluation of the model's performance, the dataset was split into training and testing subsets as shown in IV.We used an 80-20 split for training and testing,The training data was used to train the used models, while the test set remain unseen for evaluation.

B. Model Selection

In this study, we evaluate traditional machine learning (ML) models and deep learning (DL) models for the task of intrusion classification in Industrial Control system.We experiment with several well-established algorithms from both the ML and DL domains and assess their performance on the dataset derived from Modbus TCP/IP communication.

Traditional ML models such as XGBoost [11], Random Forest [12], and Decision Tree [13] are selected due to their proven effectiveness in handling structured data and, interpretable results. In contrast, Multilayer Perception (MLP) [14] and Recurrent Neural Networks (RNNs) [15] represent deep learning approaches that are capable of capturing complex, non-linear relationships and temporal characteristics, respectively

By comparing the performance of these models, we aim to identify the most suitable approach for intrusion classification .

1) Model Training and Evaluation: Each model will be trained using the same training set (as described in IV). The models will be trained using standard supervised learning techniques, with default parameters for each model . For deep learning models, we employ a simple architecture to ensure both effective learning and computational efficiency. The Multilayer Perceptron (MLP) consists of two hidden layers with 128 and 64 neurons respectivelly .The Recurrent Neural Network consists of an LSTM layer with 64 units followed by a dense layer with 128 neurons.

In order to asses the performance of the models, several key metrics will be used.

- **Precision:** Precision measures the proportion of true positive predictions out of all the instances classified as positive (i.e., attacks).

$$\text{Precision} = \frac{TP}{TP + FP}$$

Where:

  - TP = True Positives (correctly identified attacks)
  - FP = False Positives (incorrectly identified as attacks)

TABLE II
FEATURE DESCRIPTION OF DATASET

Feature	Description	Data Type
time	Timestamp	String
smac	Source MAC address	String
dmac	Destination MAC address	String
sip	Source IP address	String
dip	Destination IP address	String
request	Indicates whether the packet is a request (packet from master to slave)	Boolean
fc	Function code	Integer
error	Indicates whether there has been an error in the reading/writing operation	Boolean
madd	Memory address to perform read/write operation	Integer
data	Data exchanged during read/write operations	Integer
label	Label for attacks and normal samples	String

TABLE III
ATTACK CLASSES OF THE ELECTRA MODBUS DATASET

Classes	Percentage of Samples
Normal	94.8%
Function code recognition attack	0.19%
Response modification attack	0.1%
Force error in response attack	0.007%
Read attack	4.83%
Write attack	0.06%
Replay attack	0.006%

TABLE IV
TRAIN-TEST SPLIT

max width=

Dataset	Total Samples	Attack Samples	Normal Samples
Training Set	33366	13850	19516
Test Set	8342	3449	4893

- **Recall:** Recall, or sensitivity, evaluates the model's ability to correctly identify all actual positive cases (i.e., attacks).

$$\text{Recall} = \frac{TP}{TP + FN}$$

  Where:
  - FN = False Negatives (attacks incorrectly classified as normal)
- **F1-Score:** The F1-Score is the harmonic mean of precision and recall. It provides a balanced measure of the model's ability to correctly identify both attack and normal classes,

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Accuracy:** Accuracy provides the overall percentage of correct predictions (both normal and attack) made by the model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

  Where:
  - TN = True Negatives (correctly identified normal cases)
- **Confusion Matrix:** A confusion matrix is used to visualize the performance of the classification model by showing the actual versus predicted classifications.

III. RESULTS AND DISCUSSION

TableV summarizes the numerical performance of each model across the various evaluation metrics. The results indicate that XGBoost achieved the highest performance among all models, with an F1 score of 99.94%, closely followed by Random Forest and Decision Tree with F1 score of 99.91% and 99.90% respectively. highlighting the strength of ML methods in handling the Modbus TCP/IP dataset. Deep learning models, both MLP and RNN, showed lower performance, with anf F1 score of 99.70% for MLP model and 98.22% for RNN .This suggests that ML models may better exploit the feature set of this dataset, likely due to the lack of strong temporal patterns in the data and deep learning models typically require larger training datasets to capture complex patterns and achieve optimal performance which wasn't the case for the used dataset.

Although detection performance is critical, ensuring computational efficiency is also important in industrial control systems applications, the analysis of training and inference times for different models VI indicates that Decision Tree , Random Forest and MLP model are faster to train and inference, with Decision Tree being the most computationally efficient. However, XGBoost achieves a good balance between training time and inference performance, making it suitable for real-time applications.the RNN's higher computational cost may limit its usage with in the intrusion detection context.

TABLE V
MODEL PERFORMANCE METRICS

Model	Accuracy	Precision	Recall	F1-Score
XGBoost	0.999401	0.999410	0.999401	0.999381
Random Forest	0.999161	0.999162	0.999161	0.999156
Decision Tree	0.999041	0.999042	0.999041	0.999016
Neural Network	0.997003	0.997244	0.997003	0.997052
RNN	0.986694	0.987370	0.986694	0.982254

TABLE VI
COMPARISON OF TRAINING AND INFERENCE TIMES FOR DIFFERENT MODELS

Model	Training Time (seconds)	Inference Time per Sample (millisec-onds)	Total Inference Time (seconds)
XGBoost	**11.83**	**0.0001**	**0.04**
Random Forest	**3.52**	**0.0145**	**0.12**
Decision Tree	**0.06**	**0.0011**	**0.01**
Neural Network (MLP)	**1.68**	**0.0001**	**0.08**
RNN	**59.80**	**0.0002**	**1.48**

We further investigate the confusion matrix of the best-performing model, XGBoost, to gain deeper insights into the classification behavior. The resulting confusion matrix is shown in Figure 2, where the rows represent the actual labels and the columns represent the predicted labels.The model successfully predicts the "Normal" class with no false positives, as evidenced by 4882 correct predictions. It also performs well with the "Read Attack" class, correctly identifying 534 instances, and the "Recognition Attack" class, with 50 correct predictions. Misclassifications are minimal , such a "Replay Attack" instance incorrectly labeled as "Read Attack," and a "Write Attack" instance misclassified as "Recognition Attack.

IV. CONCLUSION

This study compared the performance of machine learning (ML) models and deep learning (DL) models for intrusion classification in Indusrial control systems using a Modbus TCP/IP dataset.XGBoost demonstrated the highest accuracy and performance, followed by Random Forest and Decision Tree models. Although deep learning models, particularly MLP, showed promise, their performance was slightly lower due to the limited size of the training data.Future work should focus on expanding training datasets to improve deep learning models and exploring hybrid ML-DL approaches.

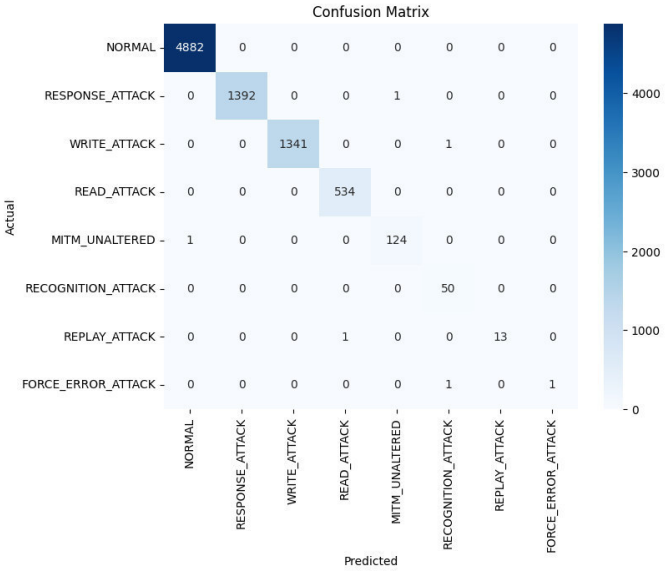


Fig. 2. Confusion Matrix for XGBoost Model

REFERENCES

[1] Martin Wollschlaeger, Thilo Sauter, and Juergen Jasperneite. "The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0". In: *IEEE industrial electronics magazine* 11.1 (2017), pp. 17–27.

[2] Mauro Conti, Denis Donadel, and Federico Turrin. "A survey on industrial control system testbeds and datasets for security research". In: *IEEE Communications Surveys & Tutorials* 23.4 (2021), pp. 2248–2294.

[3] Jehn-Ruey Jiang and Yan-Ting Lin. "Deep Learning Anomaly Classification Using Multi-Attention Residual Blocks for Industrial Control Systems". In: *Sensors* 22.23 (2022). ISSN: 1424-8220. DOI: 10.3390/s22239084. URL: https://www.mdpi.com/1424-8220/22/23/9084.

[4] Stamatis Karnouskos. "Stuxnet worm impact on industrial cyber-physical system security". In: *IECON 2011-37th Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2011, pp. 4490–4494.

[5] Chunjie Zhou et al. "A unified architectural approach for cyberattack-resilient industrial control systems". In: *Proceedings of the IEEE* 109.4 (2020), pp. 517–541.

[6] Ángel Luis Perales Gómez et al. "On the generation of anomaly detection datasets in industrial control systems". In: *IEEE Access* 7 (2019), pp. 177460–177473.

[7] Jehn-Ruey Jiang and Yan-Ting Chen. "Industrial Control System Anomaly Detection and Classification Based on Network Traffic". In: *IEEE Access* 10 (2022), pp. 41874–41888. DOI: 10.1109/ACCESS.2022.3167814.

[8] Mustafa Altaha and Sugwon Hong. "Anomaly detection for SCADA system security based on unsupervised learning and function codes analysis in the DNP3 protocol". In: *Electronics* 11.14 (2022), p. 2184.

[9] Baifeng Ning et al. "Power IoT attack samples generation and detection using generative adversarial networks". In: *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*. IEEE. 2020, pp. 3721–3724.

[10] Mercedes Oriol Rocabert. "Attack detection in a railway scenario using ML techniques". B.S. thesis. Universitat Politècnica de Catalunya, 2023.

[11] Sukhpreet Singh Dhaliwal, Abdullah-Al Nahid, and Robert Abbas. "Effective intrusion detection system using XGBoost". In: *Information* 9.7 (2018), p. 149.

[12] Rifkie Primartha and Bayu Adhi Tama. "Anomaly detection using random forest: A performance revisited". In: *2017 International conference on data and software engineering (ICoDSE)*. IEEE. 2017, pp. 1–6.

[13] Siva S Sivatha Sindhu, Suryakumar Geetha, and Arputharaj Kannan. "Decision tree based light weight intrusion detection using a wrapper approach". In: *Expert Systems with applications* 39.1 (2012), pp. 129–141.

[14] Lennart Van Efferen and Amr MT Ali-Eldin. "A multilayer perceptron approach for flow-based anomaly detection". In: *2017 international symposium on networks, computers and communications (ISNCC)*. IEEE. 2017, pp. 1–6.

[15] Benjamin J Radford et al. "Network traffic anomaly detection using recurrent neural networks". In: *arXiv preprint arXiv:1803.10769* (2018).

A New Combined Two Channel Set-Membership PU NLMS Algorithm Based On Controlling Energy and Correlation for Blind Speech Quality Enhancement

Abdelhak Cheffi[1], Akila Sayoud[2] , Rahima Henni [2] and Mohamed Djendi[1]

[1]Signal Processing and Image (LATSI) Laboratory, University of Blida 1, Blida 09000, Algeria. Emails: cheffi.abdelhak@yahoo.com, m_djendi@yahoo.fr
[2]LCPTS Laboratory, Faculty of Electrical Engineering, USTHB University, Algiers, Algeria. Emails: Akila.sayoud@gmail.com, henni93rahima@gmail.com

***Abstract*—this work addresses the problem of speech enhancement by adaptive and blind source separation techniques. We propose a new two-channel forward partial update algorithm for acoustic noise reduction and speech quality enhancement application. This new algorithm is got by the use of a new coefficients update to control adaptive NLMS cross filters by using of the cross-correlation of the output error signal and the noisy observation. We focus on the comparison between our proposed algorithm with it conventional version TC-SM-PU-NLMS, in terms of speed convergence and stability. In order to examine the good performance of this proposed algorithm, we have evaluated it through three objective criteria: System Mismatch (SM), Segmental Mean Square Error (SegMSE) and the Segmental SNR (SegSNR).**

Keywords*—*speech enhancement, adaptive algorithm, convergence speed, variable-step-size, correlation, MSE.

1. INTRODUCTION

Usually in the systems telecommunication as hand-free mobile, teleconferencing systems and hearing aids, we use a variation of adaptive filters [1-2]. In these adaptive filters algorithms that we are often used on these applications, we the least mean square (LMS), the normalized least mean square (NLMS), the affine projection (AP) and the recursive least squares (RLS) algorithms. By using these adaptive filters algorithms in the real time, we have to respect one the most important proprieties of the implementation that is the computational complexity. The computational complexity of these adaptive algorithms must be taken in consideration. We find a various approaches that have been proposed to solve this problem of the computational complexity. We note two proposed techniques of the most important approaches that reducing the problem of the computational complexity of many adaptive algorithms. The first one, we note the partial update (PU) approach of the adaptive filter coefficients [3-6]. The second one is the set membership filtering [7-11]. We can also note a combination of these two last approaches (i.e. note the partial update approach and the set membership filtering) that called as the set membership partial update filtering (SM-PU) [8]. In one of the most popular techniques for acoustic noise reduction (ANR), we find the blind source separation technique (BSS). This last technique is used to find the original source signals by using only information's of mixed signals observed in the two input channel. Recently several two channel algorithms are proposed to solve the problem the blind source separation. We note two-sensors forward BSS structure [13]. In this work, we propose a new algorithm of the two channel set membership partial update normalized least mean square (TC-SM-PU-NLMS) algorithm for the acoustic noise reduction and speech quality enhancement applications. The rest of this article is organized as follows: In Section 2 is for the description of the classical one channel set membership partial update filtering especially the OC-SM-PU-NLMS. The section 3 describes the two channel mixing model. The mathematical derivation of the proposed two channel set membership partial update NLMS basing on controlling energy and correlation algorithm is given in Section 4. The analysis of simulations results are reported on Section 5, and finally a conclusion is given in Section 6.

2. CLASSICAL PARTIAL UPDATE FILTERING ALGORITHMS

In this section, we describe the important mathematical derivation of the one channel set membership partial update normalized least mean squares (OC-SM-PU-NLMS) algorithm that is presented in [3].

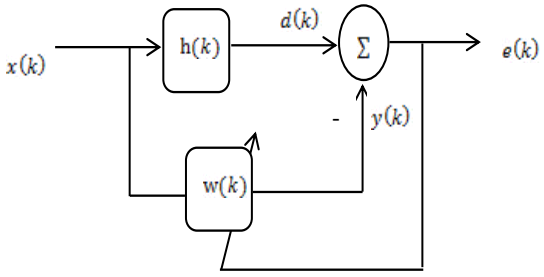


Fig. 1. One Channel PU-NLMS [3].

The OC-SM-PU-NLMS is combined especially of the two mains ideas: the first one is the partial updating approach presented in [3-6] and the second one we note the set-membership filtering approach that is given in [7-11].

The partial updating approach is based principally on derive an algorithm that updates only L coefficients from the all of $(N+1)$ filter coefficients. Firstly, we partition the input source signal vector $\boldsymbol{x}(k)$, and

respectively the adaptive filter vector $\boldsymbol{w}(k)$ into B blocks equal coefficients, where each blocks has L coefficients. As follow:

$$\boldsymbol{x}(k) = [x(k)\ x(k-1) \ldots x(k-N+1)]^T = [\boldsymbol{x}(k)_1^T \boldsymbol{x}(k)_2^T \ \ldots \ \boldsymbol{x}(k)_B^T]^T \quad (1)$$

$$\boldsymbol{w}(k) = [w(k)\ w(k-1) \ \ldots \ w(k-N+1)]^T = [\boldsymbol{w}(k)_1^T \boldsymbol{w}(k)_2^T \ \ldots \ \boldsymbol{w}(k)_B^T]^T \quad (2)$$

According to Fig. 1, the estimated output error signal $e(k)$ is given by the following equation:

$$e(k) = d(k) - \boldsymbol{w}(k)^T \boldsymbol{x}(k) \quad (3)$$

Where the signal $d(k) = \boldsymbol{x}(k)^T \boldsymbol{h}(k)$ represent the desired signal and $\boldsymbol{h}(k)$ is a real impulse response. The first aim in PU-NLMS algorithm is to find an adaptive algorithm that updates only Nb blocks from the total B available blocks. Let the chosen Nb blocks to be specified by the following index set:

$$\boldsymbol{I}_{Nb}(k) = \{i_1(k) i_2(k) \ \ldots \ i_{Nb}(k)\} \quad (4)$$

The main idea in set-membership filtering approach is to find a coefficient filter vector that minimizes the Euclidean distance $\|\mathbf{w}(\mathrm{k}+1) - \mathbf{w}(\mathrm{k})\|^2$ subject to the constraint $\mathbf{w}(\mathrm{k}) \in \mathrm{H_k}$ but with the additional constraint of update only Nb coefficients.

$$\boldsymbol{w}(k) \in \boldsymbol{H}_k \quad \Leftrightarrow \quad \{\boldsymbol{w}(k) \in \mathbb{R}; |e(k)| > \gamma\} \quad (5)$$

Where γ is a scalar parameter that defines the threshold error [3]. The relation update is given by the following relations:

$$\boldsymbol{w}(k) = \begin{cases} w(k-1) + \mu(k) \frac{e(k)\boldsymbol{x}(k)\boldsymbol{A}_{I_{Nb}(k)}}{\left\|\boldsymbol{A}_{I_{Nb}(k)}\boldsymbol{x}(k)\right\|^2 + \delta} & if \ |e(k)| > \gamma \\ w(k-1) & else \end{cases} \quad (6)$$

and

$$\mu(k) = 1 - \frac{\gamma}{|e(k)|} \quad (7)$$

where |.| symbolizes the magnitude operator and $\mu(k)$ is a variable step size.

3. CONVOLUTED MIXING MODEL

In this section, we present the two-channel convoluted mixing model that is used in our simulations. This model allows to mix two uncorrelated sources that are respectively the speech signal $s(n)$ in the first input and the punctual noise $b(n)$ in the second input as shown in Fig. 2[12-13].

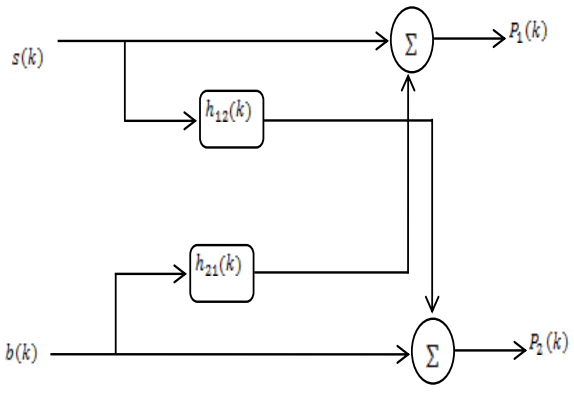


Figure 2.The mixing model [12-13].

where $h_{12}(k)$ and $h_{21}(k)$ represent the mutual impulse responses. According to this figure, the observed signals $p_1(k)$ and $p_2(k)$of this model can be given by the following relations:

$$p_1(k) = s(k) + b(k) * h_{21}(k) \quad (8)$$

$$p_2(k) = b(k) + s(k) * h_{12}(k) \quad (9)$$

Where * symbolize the convolution operator.

4. THE MATHEMATICAL DERIVATION OF THE NEW PROPOSED TC-SM-PU-NLMS ALGORITHM

The mathematical derivation of the proposed two channel set membership partial update normalized least mean square based on controlling energy and correlation algorithm is given in this section. In the first, the mixing signals $p_1(k)$ and $p_2(k)$ obtained from section 3 will be injected on the inputs of the forward blind source separation (FBSS) structure illustrated in Fig.3. The observed outputs signal $u_1(k)$and $u_2(k)$ are given as follows:

$$u_1(k) = p_1(k) - p_2(k) * w_{21}(k) \quad (10)$$

$$u_2(k) = p_2(k) - p_1(k) * w_{12}(k) \quad (11)$$

the insetion of the equations (8) and (9) in (10) and (11) respectively, and by using the optimality assumption for both adaptive filters, $i.e. w_{12}^{opt}(k) = h_{12}(k)$ and $w_{21}^{opt}(k) = h_{21}(k)$, we obtain the following equations [14]

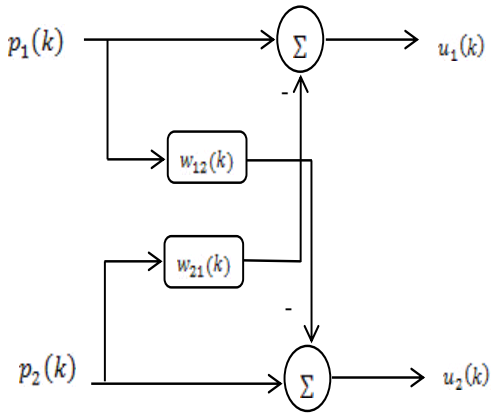


Figure 3.The forward structure [14].

$$u_1(k) = s(k) * [\delta(k) - h_{12}(k) * w_{12}(k)] \quad (12)$$

$$u_2(k) = b(k) * [\delta(k) - h_{21}(k) * w_{21}(k)] \quad (13)$$

It is clearly that the output signals $u_1(k)$ and $u_2(k)$ converge respectively to the original source signals $s(k)$ and $b(k)$. But we have noticed a small distortion. On the use of the post-filters at the output, we have avoid these distortions [14].

In the proposed algorithm, we have combined two mains idea. In the first one is the implementation of a new coefficients $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$. These two coefficients are used to control normalization energy for LMS. This normalization is combined by the energy of noisy

signal $\left\|\boldsymbol{A}_{I_{Nb}(k)}\boldsymbol{p}_1(k)\right\|^2$ respectively $\left\|\boldsymbol{A}_{I_{Nb}(k)}\boldsymbol{p}_2(k)\right\|^2$ that we control by the coefficient $\boldsymbol{\alpha}_1$ and the filtering error energy $\left\|\boldsymbol{A}_{I_{Nb}(k)}\boldsymbol{u}_2(k)\right\|^2$ respectively $\left\|\boldsymbol{A}_{I_{Nb}(k)}\boldsymbol{u}_1(k)\right\|^2$ that we control by $\boldsymbol{\alpha}_2$. The second main idea is to use two variable step-sizes $\mu_1(k)$ and $\mu_2(k)$. These variable step-sizes are controlled by the using respectively a cross-correlation between the noisy signal $\boldsymbol{p}_1(k)$ and the output error signal $u_2(k)$, and the cross-correlation between the noisy signal $\boldsymbol{p}_2(k)$ and the output error signal $u_1(k)$. The following equations report the two variable step-sizes:

$$\mu_1(k) = \begin{cases} 1 - \frac{\gamma}{\|u_2(k)\boldsymbol{p}_1^T(k)\|} & if \quad u_2(k) > \gamma_1 \\ 0 & else \end{cases} \quad (14)$$

$$\mu_2(k) = \begin{cases} 1 - \frac{\gamma}{\|u_1(k)\boldsymbol{p}_2^T(k)\|} & if \quad u_1(k) > \gamma_2 \\ 0 & else \end{cases} \quad (15)$$

By this modification, the new updating relations are presented by the following relations:

$$\boldsymbol{w}_{12}(k) = \begin{cases} \boldsymbol{w}_{12}(k-1) + \mu_1(k)\beta_1(k) & if \ |u_2(k)| > \gamma_1 \\ \boldsymbol{w}_{12}(k-1) & else \end{cases} \quad (16)$$

$$\boldsymbol{w}_{21}(k) = \begin{cases} \boldsymbol{w}_{21}(k-1) + \mu_2(k)\beta_2(k) & if \ |u_1(k)| > \gamma_2 \\ \boldsymbol{w}_{21}(k-1) & else \end{cases} \quad (17)$$

Where

$$\beta_1(k) = \frac{u_2(k)\boldsymbol{p}_1(k)\boldsymbol{A}_{I_{Nb}(k)}(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)}{\boldsymbol{\alpha}_1\left\|\boldsymbol{A}_{I_{Nb}(k)}\boldsymbol{p}_1(k)\right\|^2+\boldsymbol{\alpha}_2\left\|\boldsymbol{A}_{I_{Nb}(k)}\boldsymbol{u}_2(k)\right\|^2} \quad (18)$$

$$\beta_2(k) = \frac{u_1(k)\boldsymbol{p}_2(k)\boldsymbol{A}_{I_{Nb}(k)}(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)}{\boldsymbol{\alpha}_1\left\|\boldsymbol{A}_{I_{Nb}(k)}\boldsymbol{p}_2(k)\right\|^2+\boldsymbol{\alpha}_2\left\|\boldsymbol{A}_{I_{Nb}(k)}\boldsymbol{u}_1(k)\right\|^2} \quad (19)$$

Where $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are controlling coefficients given by the following relation:

$$\alpha_1 = 1 - \alpha_2 \quad (20)$$

The coefficient α_1 equal to one if $|u_2(k)| > 5.\gamma_1$ or $|u_1(k)| > 5.\gamma_2$.

5. THE ANALYSIS SIMULATION RESULTS

The analysis simulation results of the new proposed algorithm are reported in this section. We have carried on the comparison of the proposed Two Channel Set-Membership PU NLMS algorithm based on controlling energy and correlation and the classical TC-SM-PU-NLMS [3] algorithm in terms of stability and fast convergence. Two types of noise are used in our simulations the first one is the White Gaussian noise WGN, and the second one is United States of America Standard Institute (USASI) noise. The obtained simulations results are performed at a sampling frequency of $Fs = 8\,kHz$.

The length of the real impulse responses h_{12} and h_{21} is 128. In Fig4, the examples of the two impulse responses h_{12} and h_{21}, the original speech signal and its manual voice activity detector (MVAD), and an example of the mixture signal $\boldsymbol{p}_1(k)$ that we have used in these simulations are presented.

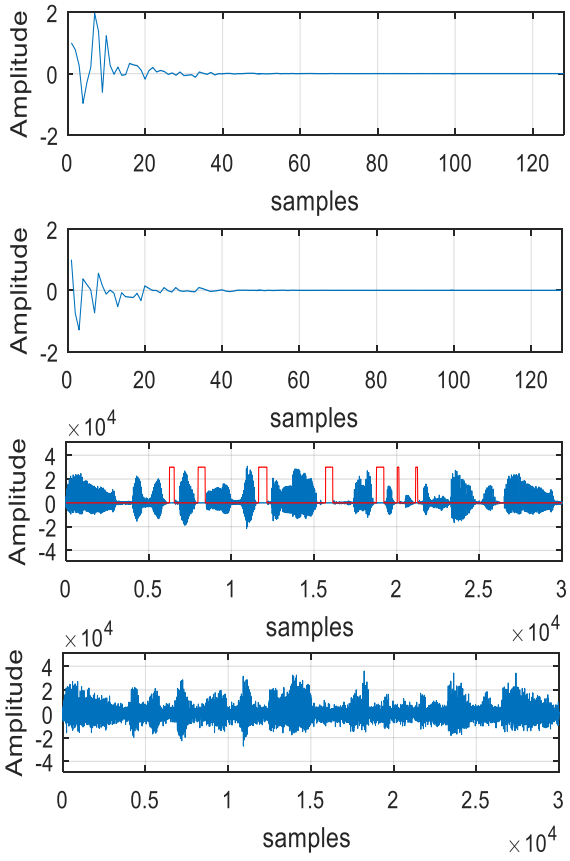


Figure 4. Impulse responses $h_{12}(k)$, $h_{21}(k)$ speech signal and its MVAD, and noisy signal [from top to bottom].

5.1. System mismatch (SM) evaluation

In this subsection, the evaluation of the system mismatch criterion (SM) for the both algorithms (.i.e. proposed algorithm and the classical TC-SM-PU-NLMS) is presented. We focus primarily only on the output $u_1(k)$ and the adaptive filter $w_{21}(k)$. The SM is given as follows:

$$SM_{dB} = 20Log_{10}\left(\frac{\|\boldsymbol{h}_{21}(k)-\boldsymbol{w}_{21}(k)\|}{\|\boldsymbol{h}_{21}(k)\|}\right) \quad (21)$$

Where $h_{21}(k)$ and $w_{21}(k)$ are respectively the real and adaptive filters vectors and ‖.‖ represents the norm-2. In this simulation we have fixed the parameters as follow:

Initial step-sizes: $\mu_1 = \mu_2 = 0.1$.

Adaptive filter length $\boldsymbol{w}_{12}, \boldsymbol{w}_{21}$: $N = 128$.

Number of blocks $B = 128$.

Selected blocks number $Nb = 120$.

Fixed threshold error: $\gamma_1 = \gamma_2 = 15$ and 90.

Input $SNR1 = 3$ dB, Input $SNR2 = 3$ dB.

The obtained results are reported on Fig. 5. Basing on the evaluation of the system mismatch (SM), we can clearly observe the good behavior of the proposed algorithm in comparison with the classical SM-PU-NLMS ones. Also, we can say that the proposed algorithm converge to the optimal solution faster than the SM-PU-NLMS one.

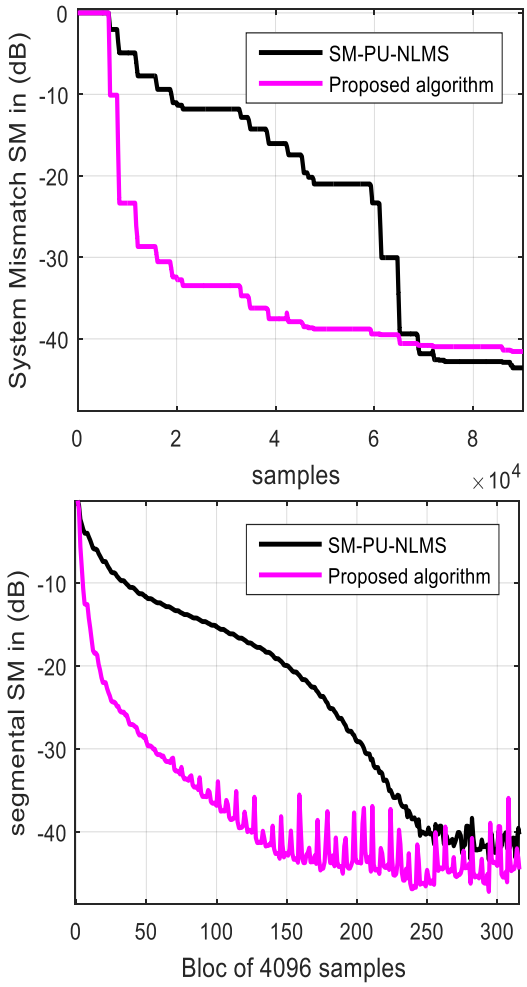


Figure 5. System mismatch evaluation of TC-SM-PU-NLMS and Proposed algorithm. Input signals: speech signal and WGN and speech signal in the top. Input signals: speech signal and USASI.

5.2. Segmental Mean Square Error SegMSE evaluation

To confirm the good performance of our proposed algorithm in term of the convergence speed comparing with the classical SM-PU-NLMS. This subsection describe the segmental mean square error SegMSE evaluation for the bothalgorithms (.i.e. proposed algorithm and the classical TC-SM-PU-NLMS). The SegMSE criterion is given by the following equation:

$$SegMSE_{dB} = 20Log_{10}(\sum_{k=0}^{M-1} u_1(k)) \qquad (22)$$

Where $M$is the time averaging frame length of output$u_1(k)$.We note that the SegMSE criterion is evaluated only in speech presence periods at the processing output. In this simulation we have fixed the parameters as follow:

Initial step-sizes: $\mu_1 = \mu_2 = 0.1$
Adaptive filter length $\boldsymbol{w}_{12}, \boldsymbol{w}_{21}$: $N = 128$
Number of blocks $B = 128$
Selected blocks number $Nb = 120$
Fixed threshold error: $\gamma_1 = \gamma_2 = 15$ and 90
Input $SNR1 = 3$ dB, Input $SNR2 = 3$ dB

The obtained results are reported on Fig.6.
From these simulation results presented in fig.6, we seewell the faster convergence speed performance of the proposed algorithmin comparison with the convontional SM-PU-NLMS. This good property of fast convergence speed is due to the modifications that are included on our proposed algorithm.

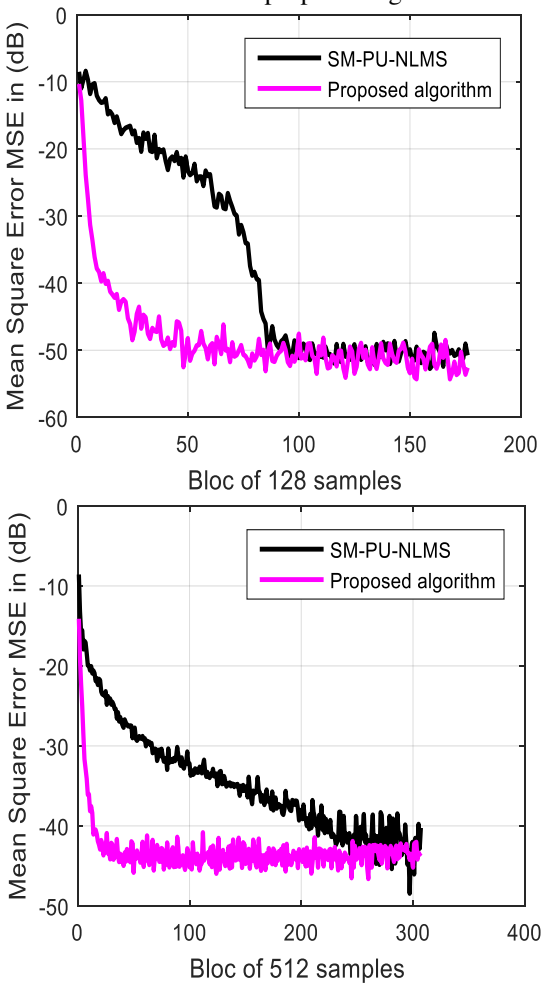


Figure 6. Mean square error evaluation of TC-SM-PU-NLMS and Proposed algorithm. Input signals: speech signal and WGN and speech signal in the top. Input signals: speech signal and USASI.

5.3. Output Segmental SNR evaluation

In order to improve the good performance of proposed algorithm in term of speech quality comparing with the conventional SM-PU-NLMS algorithm, we have evaluate the segmental signal to noise ratio criterion (SegSNR) for the both algorithms (.i.e. proposed algorithm and the classical TC-SM-PU-NLMS).

In this simulation we have fixed the parameters as follow:

Initial step-sizes: $\mu_1 = \mu_2 = 0.1$
Adaptive filter length $\boldsymbol{w}_{12}, \boldsymbol{w}_{21}$: $N = 128$
Number of blocks $B = 128$
Selected blocks number $Nb = 120$
Fixed threshold error: $\gamma_1 = \gamma_2 = 15$ and 90
Input $SNR1 = 3$ dB, Input $SNR2 = 3$ dB

Figure 7 illustrate the obtained results. The SegSNR criterion is given by the following relation:

$$SNR_{dB} = \frac{1}{Q}\sum_{i=0}^{Q-1} 10Log_{10}\left(\frac{\sum_{k=0}^{M-1}|s_i(k)|^2}{\sum_{k=0}^{M-1}|s_i(k)-u_{1,i}(k)|^2}\right) \quad (23)$$

Where $s(k)$ and $u_1(k)$ are the original speech signal and the enhanced one, respectively. The parameter M is the mean averaging value of the output SNR. The parameter Q is the number of only-speech periods.

According to the simulation results presented in fig.7, we show clearly the good performance of the proposed algorithm in comparison with the original SM-PU-NLMS algorithm.

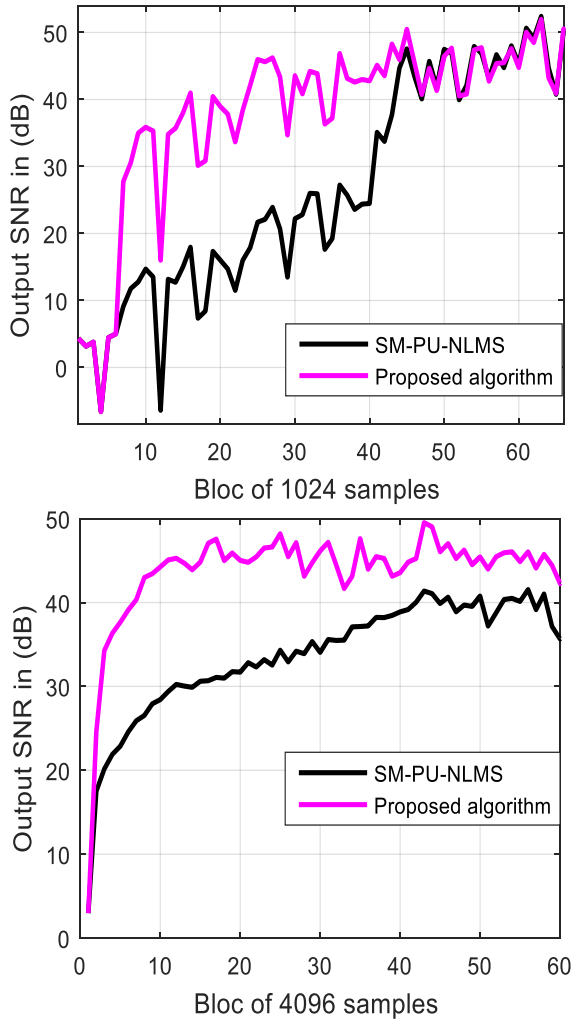


Figure 7. Output signal to noise evolutionevaluation of TC-SM-PU-NLMS and Proposed algorithm. Input signals: speech signal and WGN and speech signal in the top. Input signals: speech signal and USASI.

4. CONCLUSION

In this work, we have proposed a new version of the two-channel forward set membership partial update algorithm for blind source separation and speech quality enhancement. We have focus on a comparison between the classical SM-PU- NLMS algorithm and the proposed algorithm to demonstrate the performance of convergence speed and stability. In order toexhibita comparison between the both algorithms, we have carried a three objective criteria that are: the system mismatch criterion (SM), the segmental mean square error criterion (SegMSE) and of the output segmental signal to noise ratio criterion (SegSNR). From the obtained simulation results, we can show clearly the good performance of the proposed algorithm in term the convergence speed and stability.All the obtained results confirmed the superiority of the proposed algorithm in speech enhancement and acoustic noise reduction.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 4th edition, 2002.

[2] P. S. R. Diniz, "Data-selective adaptive filtering", *Adaptive Filtering*, Springer, 2008.

[3] S. Werner, M. L. R. Campos, and P. S. R. Diniz, "Partial-update NLMS algorithms with data-selective updating," IEEE Transactions on Signal Processing, vol. 52, no. 4, Apr. 2004.

[4] Tandon, M. N. S. Swamy, and M. O. Ahmad, "Partial-update l∞-norm based algorithms," *IEEETransactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 2, pp. 411–419, 2007.

[5] G. Deng, "Partial update and sparse adaptive filters," *IET Signal Processing*, vol. 1, no. 1, pp. 9–17, 2007.

[6] K. Dogancay, *Partial-Update Adaptive Signal Processing: Design, Analysis and Implementation*, Academic Press, Elsevier, 2008.

[7] S. Werner and P. S. R. Diniz, "Set-membership affine projection algorithm," *IEEE Signal Processing Letters*, vol. 8, no. 8, pp. 231–235, August 2001.

[8] A. Cheffi, M. Djendi, and A. Guessoum A new correlated set-membership partial-update NLMS (SM-PU-NLMSCOR) algorithm for acoustic noise reduction. In: 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B) 2017, Boumerdes, Algeria; 2017.

[9] P. S. R. Diniz and S. Werner, "Set-membership bi-normalized LMS data-reusing algorithms," IEEE Trans. Signal Processing, vol. 51, pp. 124–134, Jan. 2003.

[10] L. Guo and Y. F. Huang, "Set-membership adaptive filtering with parameter-dependent error bound tuning," in *Proc. IEEE ICASSP*, Mar. 2005, vol. 4, pp. iv/369–iv/372.

[11] P. S. R. Diniz, R. P. Baraga, and S. Werner, "Set-membership affine projection algorithm for echo cancellation," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2006.

[12] Djendi M, Scalart P, Gilloire A. Analysis of two-sensors forward BSS structure with post-filters in the presence of coherent and incoherent noise. Speech Comm. 2013; 55(10):975–987.M.

[13] Sayoud.A, Djendi,M Efcient subband fast adaptive algorithm based-backward blind source separation for speech intelligibility enhancement. International Journal of Speech Technology. 2020; 23:471–479.

[14] Djendi M, Gilloire A, Scalart P. New frequency domain post-filters for noise cancellation using two closely spaced microphones, Proc. EUSIPCO, Poznan, 3-8 Sep. 2007, vol. 1, pp. 218-221.

Speech Emotion Recognition in Arabic and Beyond: A Review

1st Hadjadji Imene
LCPTS, USTHB
Algiers, Algeria
ihadjadji@usthb.dz

2nd Falek Leila
LCPTS, USTHB
Algiers, Algeria
lfalek@usthb.dz

3rd Hamza Amina
LCPTS, USTHB
Algiers, Algeria
ahamza1@usthb.dz

***Abstract*—Speech Emotion Recognition (SER) has garnered increasing attention in recent years due to its wide-ranging applications in human-computer interaction, affective computing, and healthcare. Despite significant progress, developing robust SER systems for Arabic speech remains a challenge, primarily due to the linguistic diversity and distinct acoustic characteristics of the Arabic language. This paper presents a comprehensive review of recent advancements in Arabic SER, emphasizing key aspects such as datasets, feature extraction methods, classification techniques, and emotion detection strategies specific to Arabic speech. We address the challenges associated with the scarcity of large-scale Arabic emotional speech datasets, highlighting notable contributions like ArabEmo, QCRI-ARABIC, and Tashkeela that aim to bridge this gap. Additionally, we analyze the role of prosodic and spectral features—including pitch, energy, duration, and MFCCs—and the increasing adoption of deep learning approaches, such as CNNs and RNNs, in enhancing emotion classification performance. The paper also explores the impact of Arabic dialectal variations and advocates for the development of dialect-specific models to improve recognition accuracy. Lastly, we discuss future directions for Arabic SER, underscoring the critical need for diverse and representative datasets, as well as the integration of advanced machine learning techniques, to achieve robust and scalable performance across various Arabic dialects.**

***Index Terms*—speech emotion recognition, Arabic speech processing, deep learning, Arabic emotion datasets, feature extraction, dialect-specific models**

I. Introduction

Speech Emotion Recognition (SER) is an interdisciplinary field that focuses on identifying and categorizing emotions from speech signals. By analyzing vocal features such as tone, pitch, intensity, and rhythm, SER systems provide a deeper understanding of human emotional states, enabling applications in mental health monitoring, customer service automation, and human-computer interaction. These systems serve as a bridge between human emotional expression and machine understanding, offering valuable insights for enhancing both technology-driven and human-centered communication [1]. As emotions form a core component of communication, SER has the potential to transform how machines interpret and respond to human needs, fostering more empathetic and intuitive interactions [2].

SER research encompasses various components, including datasets, feature extraction techniques, and classification methods. While there has been significant progress in developing SER systems, much of this research has been conducted in a limited number of languages, such as English, Mandarin, and German [3]. Language and cultural diversity play a critical role in shaping emotional expression, making it essential to extend SER research to underrepresented languages [4]. Multilingual SER research not only enhances the inclusivity of these systems but also strengthens their robustness and generalizability across diverse populations [5].

Among the underrepresented languages in SER research, Arabic is particularly noteworthy. Spoken by over 400 million people globally, Arabic is characterized by a rich phonetic structure, complex grammatical rules, and significant dialectal variations that reflect the language's vast cultural diversity [6], [7]. These features pose unique challenges to the development of SER systems. Furthermore, cultural norms play a pivotal role in shaping the ways emotions are expressed and perceived in Arabic speech, further complicating the task of emotion recognition [8]. Despite the global importance of Arabic, it remains underexplored in SER research, highlighting a significant gap in the field [9].

This review aims to address this gap by providing a comprehensive analysis of existing SER research with a particular focus on the Arabic language. It evaluates the strengths and limitations of current studies, compares Arabic SER research to advancements in other languages, and identifies key challenges and opportunities for future work [10]. By shedding light on these aspects, the paper seeks to guide the development of more inclusive and culturally sensitive SER systems that account for the linguistic and cultural nuances of Arabic speakers [11].

The remainder of this paper is organized as follows. Section II provides an overview of SER, discussing its core concepts, commonly used techniques, and the challenges faced in this field. Section III delves into the datasets, features, and classification methods employed in SER research, with a specific focus on Arabic. Finally, Section IV outlines future directions and concludes the paper.

II. Overview of Speech Emotion Recognition (SER)

Speech Emotion Recognition (SER) is the process of identifying and categorizing emotions from speech signals. This field aims to bridge the gap between emotional expression in human speech and machine understanding, enabling machines

to interpret emotions from voice data. SER involves several key components: datasets, feature extraction, classification methods, and the types of emotions detected.

Fig. 1 demonstrates the organigram of an SER.

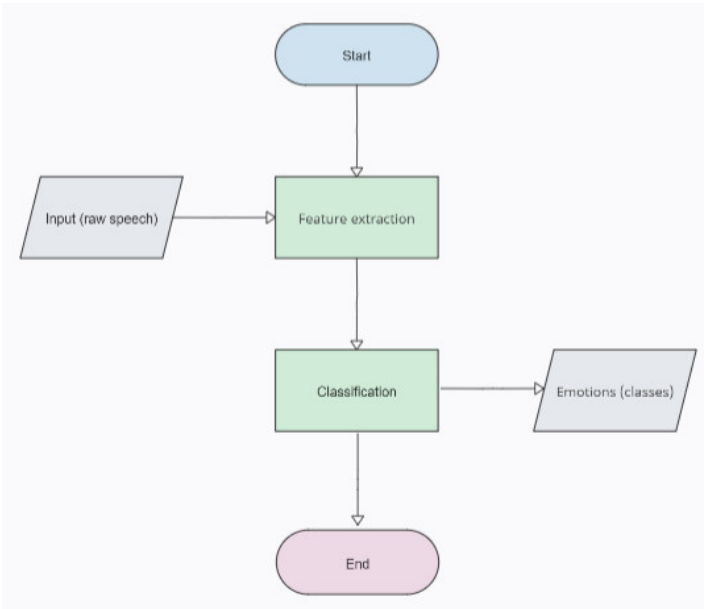


Fig. 1. Speech Emotion Recognition Flowchart.

A. Datasets

The foundation of any SER system lies in the quality and variety of the datasets used for training and evaluation. These datasets contain speech samples along with labeled emotional states, providing the necessary data for learning and testing. High-quality, diverse datasets are crucial to developing robust SER systems that can generalize across different speakers, languages, and cultures [12].

Common datasets for SER include the Emo-DB (Berlin Database of Emotional Speech) [13], RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) [14], and TESS (Toronto Emotional Speech Set) [15], among others. These datasets provide labeled speech samples in various emotional categories such as anger, joy, sadness, and fear, among others. Some datasets also include non-linguistic features such as prosody (tone, pitch, duration) to enrich emotion detection. In multilingual SER research, datasets specific to particular languages such as [16] for Arabic have been developed to capture the unique linguistic and phonetic characteristics of the language, as well as the cultural nuances associated with emotional expression. For example, the dataset described in [16] focuses on expressive Arabic speech synthesis, utilizing diphone concatenation to generate sentences in both "neutral" and "sadness" expressions. Such datasets not only facilitate emotion recognition tasks but also contribute to the development of more culturally and linguistically sensitive SER systems. To provide an overview of some of the commonly used datasets in SER, table I [18] highlights selected datasets across various languages. This table emphasizes the size and diversity of the datasets, along with the emotional categories they cover, illustrating the range of resources available for SER research.

B. Feature Extraction

Feature extraction is a critical step in SER, as it transforms raw speech signals into meaningful information that can be used by machine learning models. The goal is to identify the most relevant characteristics of the speech signal that are indicative of emotional states [12].

Commonly used features in SER include prosodic features such as pitch, intensity, and duration, which capture variations in speech rhythm, loudness, and frequency [17]. Formant frequencies, which are resonant frequencies in the vocal tract, also provide valuable insight into emotional states. Mel-frequency cepstral coefficients (MFCCs) are widely used features, representing the power spectrum of speech signals and encoding information about the timbre of the voice [18]. Other features include zero-crossing rate (ZCR), short-time energy, and spectral features, which help capture more fine-grained details of the speech signal that might reflect emotional variations [19].

Feature selection techniques are often employed to reduce the dimensionality of the feature set and retain only the most informative features [19]. This ensures that the classification algorithms are not overwhelmed with irrelevant data, which could impact their performance.

C. Classification Methods

After feature extraction, the next critical step in SER is the classification of the speech signal into one of the predefined emotional categories. Various machine learning and deep learning methods are used for classification, depending on the complexity of the task and the nature of the dataset [20].

Traditional machine learning algorithms used in SER include Support Vector Machines (SVMs), K-Nearest Neighbors (KNN), and Random Forests [19]. These methods typically perform well with hand-crafted features and smaller datasets, as they can effectively classify emotion categories with high accuracy.

In recent years, deep learning approaches have gained prominence in SER due to their ability to automatically learn relevant features from raw speech data, bypassing the need for manual feature extraction. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks, are commonly used for SER tasks [21]. CNNs are particularly effective for capturing spatial patterns in spectrograms of speech, while RNNs are suited for handling the temporal nature of speech signals. More recently, transformer-based architectures such as BERT and wav2vec have shown promising results for emotion recognition by leveraging large amounts of data and advanced training techniques [21].

Ensemble methods, such as combining multiple classifiers to improve the robustness of the system, are also widely used

TABLE I
OVERVIEW OF EMOTION SPEECH DATASETS ACROSS DIFFERENT LANGUAGES [18]

Database Name	Language	Size	Emotions Covered
Berlin Emotional Database (EMO-DB)	German	7 emotions, 10 utterances, 10 speakers	Neutral, anger, sadness, fear, boredom, happiness, disgust, boredom
Surrey Audio-Visual Expressed Emotion (SAVEE)	English	7 emotions, 4 speakers, 120 utterances	Surprise, anger, fear, disgust, sadness, neutral, happiness
RECOLA Speech Database	French	7 hours of speech, 46 speakers	5 social behaviors, valence, and arousal
FAU Aibo Emotion Corpus	German	51 children talking to robot dog Aibo, 9 hours of speech	Bored, joyful, helpless, touchy, anger, reprimanding, emphatic, surprised, neutral, motherese, rest
Oriya Emotion Speech Dataset	Odia/Oriya	35 speakers, recoded text from Oriya drama scripts	Astonishment, sadness, fear, anger, happiness, neutral
Multilingual Database	Japanese	LEGO emotion database, EMO-DB, UUDB, SAVEE	Angry, slightly angry, very angry, neutral, friendly, nonspeech, and others
Persian Emotion Speech Dataset	Persian	748 utterances from PDREC	Happiness, anger, sadness, surprise, fear, boredom, neutral, disgust

in SER to enhance performance, particularly in real-world applications [22].

D. Emotions Detected

The emotions typically detected by SER systems are chosen based on the dataset used and the specific application of the system. The most commonly recognized emotions in SER are anger, joy, sadness, fear, disgust, and surprise, which correspond to basic human emotions [23]. These emotions are usually mapped to affective states that reflect varying levels of intensity, such as positive, negative, or neutral.

Some SER systems focus on a smaller set of emotions, often referred to as the ”big six” basic emotions: anger, joy, sadness, fear, disgust, and surprise. Others, particularly those that employ deep learning methods, may aim to detect more nuanced emotional states or include additional emotions such as boredom, shame, or frustration [20]. For instance, some Arabic SER studies focus on detecting a subset of emotions that are more culturally relevant or expressed differently in the Arabic language and cultural context [23].

In cross-lingual SER research, it is essential to account for cultural differences in emotional expression. For instance, certain emotions may be more difficult to detect or may be expressed differently in various linguistic and cultural groups. Consequently, emotion detection models must be trained with data from diverse cultural contexts to ensure their accuracy and generalizability [12].

III. LITERATURE IN ARABIC SPEECH

The table II summarizes notable contributions in the domain of SER for Arabic speech. Each work highlights the dataset characteristics, features used, employed classifiers or models, targeted emotions, recognition accuracy, and any specific notes or observations. This compilation aims to provide insights into the advancements and challenges faced in this specialized area.

A. Discussion

The presented table highlights the diverse approaches used in SER for Arabic speech, revealing a mix of successes and challenges across different studies:

1) Datasets and Dialects: Several studies focus on Modern Standard Arabic (MSA), while others explore specific dialects such as Egyptian or Saudi. This diversity is crucial as dialects can significantly influence speech patterns and emotion expression.

The use of acted versus spontaneous datasets also impacts performance. For example, [Hadjadji et al., 2020] showed improved recognition in spontaneous speech scenarios. Features and Techniques:

2) Features: Features such as acoustic, spectral (e.g., MFCC), and prosodic elements dominate the field, reflecting their effectiveness in capturing emotion-specific variations. Advanced feature combinations, such as wavelet analysis in [Abdel-Hamid, 2020], indicate efforts to enhance classifier inputs for better accuracy.

3) Classifiers and Models: Traditional classifiers like SVM, KNN, and Random Forest remain popular, with newer models like CNNs and hybrid architectures (e.g., CNN-BLSTM-DNN in [Hifny et al., 2020]) demonstrating significant performance improvements. [Meftah et al., 2014] uniquely tested a wide range of classifiers, showcasing the variability in their application to Arabic SER. Recognition Performance:

The accuracy varies widely depending on dataset characteristics, feature selection, and model complexity. For example, [Klaylat et al., 2018] achieved a high 95.5% accuracy using

TABLE II
SUMMARY OF EMOTION RECOGNITION IN ARABIC SPEECH WORKS [27]

Reference	Dataset (Dialect/Speakers)	Features	Classifiers/Models	Emotions	Recognition Accuracy	Notes
[Hadjadji et al., 2020] [19]	MSA (non-professionals) vs German (acted)	Acoustic, statistical features	KNN, SVM, ExtraTrees, Random Forest, Gradient Boosting, CNN, MLP	Anger, Happiness, Neutral, Sadness	64% (MSA)	Improved recognition for spontaneous speech.
[Hifny et al., 2020] [24]	MSA (acted)	Spectral features (MFCC, mel spectrogram)	CNN-BLSTM-DNN, CNN	Various emotions	87.2% (CNN-BLSTM-DNN), 85% (CNN)	Attention-based CNN-LSTM-DNN model showed 2.2% improvement over baseline CNN.
[Abdel-Hamid, 2020] [25]	Egyptian (acted)	Prosodic, spectral, wavelet	SVM, KNN	Various emotions	66.8% (SVM), 61.7% (KNN)	Speech from Egyptian Arabic dialect.
[Klaylat et al., 2018] [26]	Mixed dialects	Prosodic features	SMO	Happy, Angry, Sad, Neutral	95.5%	Used mixed-dialect datasets.
[ALJUHANI et al., 2021] [27]	Saudi Dialect Corpus	Spectral features (MFCC)	SVM, KNN, MLP	Anger, Happiness, Sadness, Neutral	77.14% (SVM)	Saudi Arabic dataset using YouTube data.
[Meftah et al., 2014] [28]	MSA	Acoustic, linguistic	35 tested classifiers	Happy, Angry, Surprised	Varied	Naturalistic data from Arabic TV shows.
[Hadjadji et al., 2019] [17]	MSA (students)	Acoustic (cepstral, shimmer, jitter, duration)	MLP	Joy, Sadness, Anger, Neutral	98% (intra-speaker), 54.75% (inter-speaker)	Highlighted system's dependency on speaker variability.
[Hadjadji et al., 2023] [29]	MSA (students)	Spectrogram (Linear, Logarithmic and Mel)	CNN	Joy, Sadness, Anger, Neutral	76%	Highlighted the effect of the transmission channel on the emotion recognition in speech.

mixed dialects, while [Hadjadji et al., 2019] faced challenges with inter-speaker variability (54.75%).

4) Challenges and Observations: Speaker variability, dataset quality, and emotional granularity remain significant challenges, as highlighted in studies like [Hadjadji et al., 2019].

Leveraging naturalistic datasets, as in [Meftah et al., 2014], reflects a growing interest in real-world applications, although such datasets often introduce noise and complexity. This summary emphasizes the ongoing advancements in Arabic SER while shedding light on areas requiring further research, such as addressing inter-speaker variability, enhancing spontaneous speech recognition, and expanding datasets to include underrepresented dialects.

IV. CONCLUSION

This review has highlighted the current state of Speech Emotion Recognition (SER) with a focus on Arabic, a language characterized by its rich dialectal variation and cultural diversity. Although significant advancements have been achieved in SER research for widely spoken languages, Arabic SER remains underexplored, presenting unique challenges and opportunities for development.

We examined the essential components of Arabic SER, including datasets, feature extraction methods, and classification techniques, as well as the influence of dialectal and cultural nuances on emotion detection. While datasets like ArabEmo and QCRI-ARABIC represent valuable progress, the scarcity of large-scale, diverse resources continues to hinder the development of robust SER systems for Arabic. Addressing dialectal variations and creating culturally sensitive datasets are crucial for improving system accuracy and generalizability.

Feature extraction techniques have advanced from traditional prosodic and spectral features, such as pitch, energy, and MFCCs, to deep learning-based methods like CNNs and RNNs, which excel in capturing complex acoustic and temporal patterns in speech. Similarly, hybrid models combining machine learning algorithms with deep learning approaches are increasingly effective for emotion classification in Arabic speech.

Moving forward, the development of larger, dialectally diverse, and culturally representative datasets must be prioritized to support the creation of robust SER systems. Research should also focus on innovative feature extraction and classification techniques, along with deeper exploration of the cultural and contextual factors influencing emotional expression in Arabic-speaking communities. These efforts are critical to bridging the gap between linguistic complexity and technological advancement.

In summary, Arabic SER remains a challenging yet promising field. With growing interest and technological progress, there is a solid foundation for further research and innovation. Collaboration among linguists, data scientists, and machine learning experts will be essential in overcoming existing barriers and unlocking the full potential of SER in Arabic speech. Such advancements could drive impactful applications in human-computer interaction, healthcare, customer service, and social robotics, contributing to more inclusive and effective emotion recognition systems.

ACKNOWLEDGMENT

The authors thank the University of Science and Technology Houari Boumediene for supporting this research.

REFERENCES

[1] Ingale, A. B., & Chaudhari, D. S. (2012). Speech emotion recognition. International Journal of Soft Computing and Engineering (IJSCE), 2(1), 235-238.

[2] Al-Dujaili, M. J., & Ebrahimi-Moghadam, A. (2023). Speech emotion recognition: a comprehensive survey. Wireless Personal Communications, 129(4), 2525-2561.
[3] Lee, S. W. (2019, May). The generalization effect for multilingual speech emotion recognition across heterogeneous languages. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5881-5885). IEEE.
[4] Heracleous, P., & Yoneyama, A. (2019). A comprehensive study on bilingual and multilingual speech emotion recognition using a two-pass classification scheme. PloS one, 14(8), e0220386.
[5] Hozjan, V., & Kačič, Z. (2003). Context-independent multilingual emotion recognition from speech signals. International journal of speech technology, 6, 311-320.
[6] Iben Nasr, L., Masmoudi, A., & Hadrich Belguith, L. (2024). Survey on Arabic speech emotion recognition. International Journal of Speech Technology, 27(1), 53-68.
[7] Ouali, S., & El Garouani, S. (2024). Arabic Speech Emotion Recognition using Convolutional Neural Networks. Journal of Electrical Systems, 20(3), 2700-2708.
[8] Alhumoud, S. O., & Al Wazrah, A. A. (2022). Arabic sentiment analysis using recurrent neural networks: a review. Artificial Intelligence Review, 55(1), 707-748.
[9] Singh, Y. B., & Goel, S. (2022). A systematic literature review of speech emotion recognition approaches. Neurocomputing, 492, 245-263.
[10] Thimmaiah, S. (2024). RETRACTED ARTICLE: A review on emotion recognition from dialect speech using feature optimization and classification techniques. Multimedia Tools and Applications, 83(29), 73793-73793.
[11] Alqadasi, A. M. A., Abdulghafor, R., Sunar, M. S., & Salam, M. S. B. H. (2023). Modern standard Arabic speech corpora: a systematic review. Ieee Access, 11, 55771-55796.
[12] Swain, M., Routray, A., & Kabisatpathy, P. (2018). Databases, features and classifiers for speech emotion recognition: a review. International Journal of Speech Technology, 21, 93-120.
[13] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. In Ninth European Conference on Speech Communication and Technology
[14] S. R. Livingstone and F. A. Russo, The ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English, PLoS ONE, vol. 13, no. 5, May 2018, Art. no. e0196391.
[15] Dupuis, K., & Pichora-Fuller, M. K. (2010). Toronto emotional speech set (tess)-younger talker happy.
[16] Demri, L., Falek, L., & Teffahi, H. (2015). Contribution to the design of an expressive speech synthesis system for the Arabic language. In Speech and Computer: 17th International Conference, SPECOM 2015, Athens, Greece, September 20-24, 2015, Proceedings 17 (pp. 178-185). Springer International Publishing.
[17] Hadjadji, I., Falek, L., Demri, L., & Teffahi, H. (2019, November). Emotion recognition in Arabic speech. In 2019 international conference on advanced electrical engineering (ICAEE) (pp. 1-5). IEEE.
[18] Wani, T. M., Gunawan, T. S., Qadri, S. A. A., Kartiwi, M., & Ambikairajah, E. (2021). A comprehensive review of speech emotion recognition systems. IEEE access, 9, 47795-47814.
[19] Hadjadji, I., Falek, L., Lounnas, K., & Lichouri, M. (2020, December). Enhancement of the interlocutor emotion recognition rate from non-professionals speakers in Arabic database. In 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS) (pp. 1-6). IEEE.
[20] Koolagudi, S. G., & Rao, K. S. (2012). Emotion recognition from speech: a review. International journal of speech technology, 15, 99-117.
[21] Khalil, R. A., Jones, E., Babar, M. I., Jan, T., Zafar, M. H., & Alhussain, T. (2019). Speech emotion recognition using deep learning techniques: A review. IEEE access, 7, 117327-117345.
[22] Zhao, J., Mao, X., & Chen, L. (2019). Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomedical signal processing and control, 47, 312-323.
[23] Wani, T. M., Gunawan, T. S., Qadri, S. A. A., Kartiwi, M., & Ambikairajah, E. (2021). A comprehensive review of speech emotion recognition systems. IEEE access, 9, 47795-47814.
[24] Y. Hifny and A. Ali, Ef cient Arabic emotion recognition using deep neural networks, in Proc. IEEEInt. Conf. Acoust., Speech Signal Process. (ICASSP), May 2019, pp. 67106714.
[25] L. Abdel-Hamid, Egyptian Arabic speech emotion recognition using prosodic, spectral and wavelet features, Speech Commun., vol. 122, pp. 1930, Sep. 2020.
[26] S. Klaylat, Z. Osman, L. Hamandi, and R. Zantout, Emotion recognition in Arabic speech, Analog Integr. Circuits Signal Process., vol. 96, no. 2, pp. 337351, Aug. 2018.
[27] Aljuhani, R. H., Alshutayri, A., & Alahdal, S. (2021). Arabic speech emotion recognition from Saudi dialect corpus. IEEE Access, 9, 127081-127085.
[28] A.Meftah,Y.Alotaibi, and S.-A. Selouani, Designing, building, and analyzing an Arabic speech emotional corpus, inProc.WorkshopFree/Open Source Arabic Corpora Corpora Process. Tools Workshop Programme, 2014, p. 22.
[29] Hadjadji, I., & Falek, L. (2024). Evaluating degradation in emotional speech quality over a 4G telephone network. Multimedia Tools and Applications, 83(20), 58075-58102.

Emotion Recognition in Speech Using CNNs and Mel Spectrograms: Classifying Seven Emotions

1st Hadjadji Imene
LCPTS, USTHB
Algiers, Algeria
ihadjadji@usthb.dz

2nd Falek Leila
LCPTS, USTHB
Algiers, Algeria
lfalek@usthb.dz

3rd Hamza Amina
LCPTS, USTHB
Algiers, Algeria
ahamza1@usthb.dz

***Abstract*—This paper presents a Convolutional Neural Network (CNN) model for emotion recognition in speech, using Mel spectrograms as input features. The model classifies seven emotional states—anger, neutral, joy, anxiety, disgust, boredom, and sadness—using the EmoDB database, which contains emotional speech samples recorded by ten actors. The CNN architecture achieved an accuracy of 81.3%, outperforming previous studies and demonstrating the effectiveness of deep learning techniques in capturing spectral and temporal features of speech. The results highlight the potential of CNNs for emotion recognition in speech, with applications in human-computer interaction and automated systems.**

***Index Terms*—Mel spectrogram, CNN, emotion recognition, speech processing, deep learning**

I. Introduction

Emotion recognition in speech signals has gained increasing attention in recent years due to its potential applications in human-computer interaction, mental health assessment, and customer service automation. Accurately classifying emotions such as joy, sadness, fear, and anger from speech remains a challenging problem due to the variability of emotional expression across speakers, languages, and contexts [1].

Emotion recognition traditionally relies on various feature extraction techniques to capture both low-level and high-level acoustic features, including prosodic features (pitch, energy) and spectral features (Mel-frequency cepstral coefficients, MFCCs). Recently, Mel spectrograms have been widely adopted as they provide time-frequency representations of speech signals that are more directly interpretable by deep learning models, especially in capturing temporal patterns associated with emotional states [2].

Deep learning, particularly using Convolutional Neural Networks (CNNs), has significantly advanced the state of the art in emotion recognition [3]. CNNs are well-suited for capturing spatial and frequency patterns in spectrogram representations of audio signals, making them a powerful tool for analyzing speech-based emotion recognition tasks. Unlike traditional methods that rely on handcrafted features, CNNs automatically learn hierarchical representations from raw input data, effectively extracting meaningful patterns that correspond to emotional cues.

By leveraging the structure of spectrograms, CNNs can identify localized patterns such as energy distribution and frequency variations that are indicative of different emotional states. These models excel at learning both low-level features, such as edges and contours in the spectrogram, and higher-level abstract features that emerge in deeper layers. Studies have consistently demonstrated the efficacy of CNNs in emotion recognition, attributing their success to their ability to focus on spatial and frequency domains without requiring additional temporal modeling [4], [7].

Recent advancements have explored various CNN architectures to enhance performance further. Techniques such as deeper networks, attention mechanisms, and optimized filter sizes have been used to improve the network's ability to generalize across diverse datasets and capture subtle variations in emotional expression. CNN-based methods offer a computationally efficient alternative to hybrid approaches, focusing on spatial and frequency domain analysis while maintaining competitive classification accuracy [8].

Mel spectrograms have been increasingly used as input features for deep learning models in emotion recognition, as they provide a richer representation of the speech signal's spectral content compared to traditional features like MFCCs. This representation aligns well with human perception of sound frequencies, making it an effective feature for emotion detection in speech [9].

Feature Extraction: The Mel spectrogram, with its compressed frequency scale and higher sensitivity to human voice range, allows for capturing nuanced emotional cues that might be missed in traditional spectral representations [10]. Emotion-Specific Patterns: Different emotions tend to exhibit distinct patterns in the time-frequency domain, which Mel spectrograms effectively capture. For instance, joy may be associated with higher pitch and energy, whereas sadness may exhibit lower pitch and energy variations [11].

Several databases, such as the German Emotional Speech Database (Emo-DB), are widely used as benchmarks for evaluating emotion recognition models. Emo-DB includes recordings in German with clearly labeled emotion classes, such as joy, sadness, fear, and anger, providing a standardized dataset for testing and comparing various emotion recognition algorithms [12].

In previous research [13], we investigated the application of Convolutional Neural Networks (CNNs) on Mel spectrogram representations for emotion recognition in speech. This approach demonstrated promising results, showcasing

the effectiveness of CNNs in capturing spatial features from Mel spectrograms. However, our previous system was limited to recognizing only four emotion classes, whereas real-life emotion recognition in speech often involves a much broader and more diverse range of emotional expressions.

To address this gap, the present work explores the extension of the emotion recognition system to handle seven distinct emotion classes: anger, joy, sadness, fear, disgust, surprise, and neutrality. This expansion aligns more closely with real-world scenarios, where emotional expressions are diverse and nuanced.

The proposed system incorporates Mel spectrograms as input, given their ability to capture detailed spectral content and emotional nuances in speech. The CNN architecture is designed to enhance feature extraction and classification performance by utilizing multiple convolutional layers to capture localized patterns, such as pitch variations and energy shifts, which are indicative of emotional states.

By expanding the number of emotion classes and optimizing the CNN framework, this work aims to improve the applicability of emotion recognition systems in real-life scenarios, such as customer service, therapeutic settings, and intelligent virtual assistants. The findings are expected to demonstrate the scalability and robustness of CNNs when applied to larger and more diverse emotional datasets.

The remainder of the paper is organized as follows: In Section 2, we present the linguistic material used in the study, describing the EmoDB dataset and its partitioning for training and testing. Section 3 details the classification method, starting with an introduction to Convolutional Neural Networks (CNNs) and followed by a description of the specific CNN architecture employed for this study. Section 4 presents the results and discusses the performance of the proposed model in comparison with previous work. Finally, Section 5 concludes the paper, summarizing the main findings and suggesting potential directions for future research.

II. Linguistic Material

The database EmoDB comprises emotional speech samples performed by ten actors (five female, five male) who simulated seven different emotions, including anger, sadness, joy, and neutral state. Each actor recorded ten German sentences in an anechoic chamber, producing approximately 800 utterances. The recordings were evaluated in a perception test to assess the recognizably and naturalness of each emotion. Sentences with over 80% recognition accuracy and more than 60% perceived naturalness were further labeled for detailed phonetic analysis, covering voice quality and articulatory features. This curated data supports various studies on emotional cues in speech, accessible online at http://www.expressive-speech.net/emodb/ [16].

The entire dataset utilized in this study encompasses a diverse range of seven distinct emotional categories, namely anger, neutral, joy, anxiety, disgust, boredom, and sadness. Each of these emotions represents unique characteristics and variations, making the dataset both comprehensive and challenging for classification tasks. To prepare the data for model development and evaluation, we carefully partitioned the dataset into two subsets. Specifically, 70% of the total samples were designated for the training process, allowing the model to learn and fine-tune its parameters effectively. The remaining 30% of the samples were set aside for testing purposes, ensuring an unbiased evaluation of the model's generalization capabilities. This division was designed to strike a balance between providing the model with sufficient data for robust training and maintaining a reliable test set to assess its performance accurately.

III. Classification method

A. CNNs

Convolutional Neural Networks (CNNs) are a specialized type of artificial neural network designed to process and analyze data with a grid-like topology, such as images or spectrograms. Inspired by the visual cortex in the human brain, CNNs are particularly adept at capturing spatial hierarchies in data through their layered structure. The primary building blocks of a CNN are convolutional layers, pooling layers, and fully connected layers, each serving a distinct purpose in the learning process.

The convolutional layers form the core of a CNN and are responsible for feature extraction. They use learnable filters (or kernels) that slide across the input, performing element-wise multiplications to detect patterns such as edges, textures, or more complex structures as the network deepens. These filters enable the CNN to learn spatial features from the data, maintaining the relationships between pixels or elements. The extracted feature maps from the convolutional layers are passed through activation functions like ReLU (Rectified Linear Unit), which introduce non-linearity, allowing the model to capture more intricate relationships [8].

Pooling layers, such as max pooling or average pooling, are typically interspersed between convolutional layers to reduce the spatial dimensions of the feature maps. This dimensionality reduction helps decrease computational complexity while retaining the most significant features, thereby making the model more robust to slight variations and noise in the data [2].

At the final stage, the feature maps are flattened and passed through fully connected layers, where high-level features are transformed into a prediction output. This layer serves as the decision-making component of the network, mapping the learned features to the desired output, such as class probabilities in classification tasks. Together, these components make CNNs a versatile and powerful tool for tasks involving visual and structured data, capable of learning both simple and complex representations autonomously [1].

B. Classification method used:

The classification method employed in this study involves a Convolutional Neural Network (CNN) designed for effective feature extraction and classification. Spectrograms, generated from audio signals and represented in Mel scales, serve as the input to the CNN. Each spectrogram is of size 128×128 and undergoes feature extraction through two convolutional layers. The first convolutional layer has 32 filters, while the second layer contains 128 filters, both using appropriate activation functions and pooling mechanisms for dimensionality reduction and feature enhancement.

The features extracted by the CNN are then flattened and passed into a fully connected layer with 512 neurons, enabling the model to learn higher-level representations. Finally, the output layer predicts seven distinct classes, corresponding to the categories in the dataset. This architecture effectively captures the spatial structure of the Mel spectrograms, ensuring robust performance for the classification task.

IV. Results and Discussion

Table I summarizes the model's performance metrics.

TABLE I
PERFORMANCE METRICS

Measure	Precision	Recall	F1 Score	Accuracy
Results	0.791	0.823	0.795	0.813

The CNN model achieves a classification accuracy of 81.3%, outperforming the results reported in Pham et [16] with 76% accuracy and Chauhan et [17] with 72% accuracy, highlighting its competitive performance compared to recent studies. However, the accuracy is slightly lower than the [13] result from my previous work, which can be explained by the increased complexity of the current classification task. Unlike [13], where a smaller number of classes was considered, this study involves the classification of seven distinct classes, introducing greater variability and overlapping features.

The higher number of classes poses additional challenges for feature differentiation and model generalization. Nonetheless, the achieved accuracy demonstrates the effectiveness of our CNN-based approach in addressing this more demanding problem, showcasing its robustness in handling diverse and complex classification scenarios.

V. Conclusion

This study presents an in-depth exploration of emotion recognition in speech signals using a Convolutional Neural Network (CNN) architecture. By leveraging Mel spectrogram representations as input features, the proposed approach effectively captures the spatial and frequency patterns associated with different emotional states. The study emphasizes the strength of CNNs in automatically learning hierarchical features, eliminating the need for manual feature engineering and significantly advancing the field of emotion recognition in speech.

Through the careful design of the CNN model, including two convolutional layers with increasing filter dimensions, pooling mechanisms for dimensionality reduction, and a fully connected layer for classification, the proposed methodology achieves competitive results. The classification accuracy of 81.3% highlights the efficacy of the approach, particularly when addressing the challenges posed by a diverse dataset encompassing seven distinct emotional categories. These results underscore the robustness and generalization capability of CNNs in handling complex tasks with significant variability in emotional expression.

The findings of this study are particularly noteworthy when compared to prior works in the field. The proposed CNN model outperforms recent studies, such as those by Pham et al. and Chauhan et al., demonstrating its superiority in terms of classification accuracy. However, it is acknowledged that the slightly lower accuracy compared to our previous work [13] is attributable to the increased complexity of the current task. Specifically, the inclusion of a larger number of emotional classes introduces challenges in feature separation and class differentiation, which are critical for achieving higher accuracy.

Despite these challenges, the achieved results confirm the potential of CNNs as a reliable and computationally efficient solution for emotion recognition in speech. Furthermore, the use of Mel spectrograms as input features aligns with human auditory perception, enabling the model to capture subtle emotional cues often overlooked by traditional feature extraction methods.

Future work could explore further enhancements to the CNN architecture, such as incorporating advanced techniques like attention mechanisms or residual connections, to improve the model's ability to capture nuanced patterns in the data. Additionally, extending the study to multilingual datasets or real-world scenarios with spontaneous speech could further validate the robustness and applicability of the approach.

In conclusion, this work contributes significantly to the growing body of research on deep learning-based emotion recognition, offering a robust framework for accurately classifying emotions from speech signals. The insights gained from this study not only advance the state of the art but also pave the way for practical applications in human-computer interaction, mental health monitoring, and customer service automation. The findings reaffirm the promise of CNNs as a cornerstone technology in the development of intelligent systems capable of understanding and responding to human emotions.

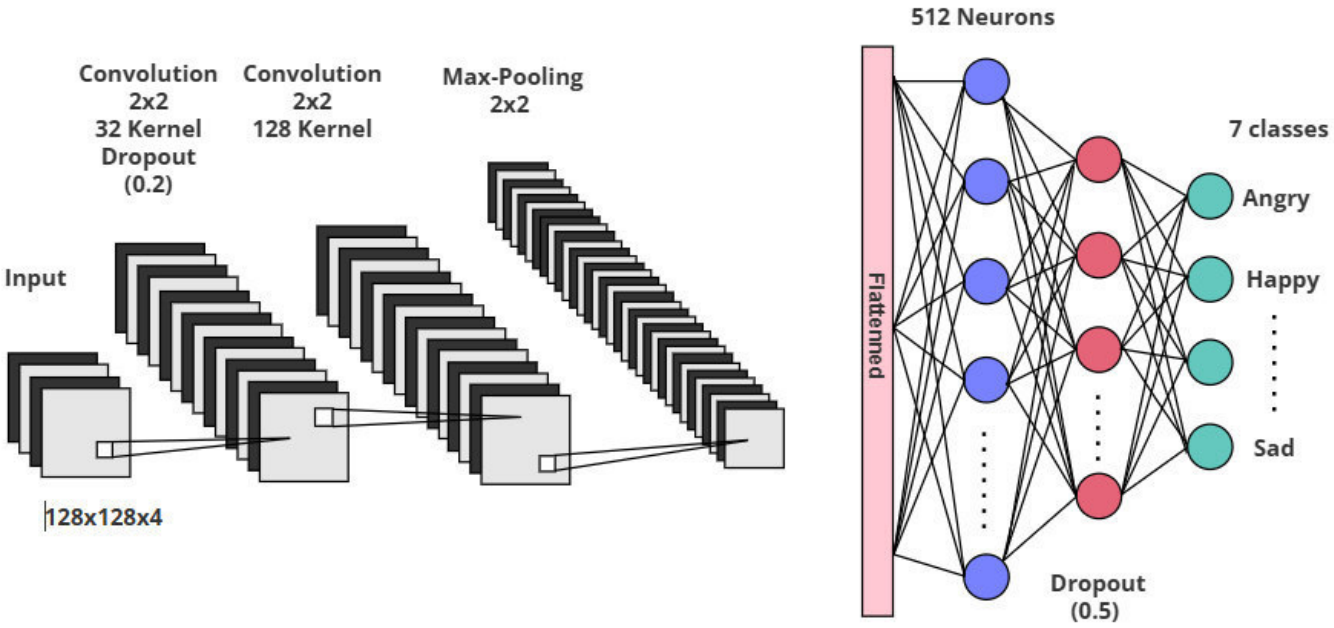


Fig. 1. Proposed CNN model architecture.

ACKNOWLEDGMENT

The authors thank the University of Science and Technology Houari Boumediene for supporting this research.

REFERENCES

[1] Liu, S., Gao, P., Li, Y., Fu, W., & Ding, W. (2023). Multi-modal fusion network with complementarity and importance for emotion recognition. Information Sciences, 619, 679-694.

[2] Meng, H., Yan, T., Yuan, F., & Wei, H. (2019). Speech emotion recognition from 3D log-mel spectrograms with deep learning network. IEEE access, 7, 125868-125881.

[3] Fan, Y., Lu, X., Li, D., & Liu, Y. (2016, October). Video-based emotion recognition using CNN-RNN and C3D hybrid networks. In Proceedings of the 18th ACM international conference on multimodal interaction (pp. 445-450).

[4] Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative study of CNN and RNN for natural language processing. arXiv preprint arXiv:1702.01923.

[5] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 33(12), 6999-7019.

[6] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). Recent advances in convolutional neural networks. Pattern recognition, 77, 354-377.

[7] Lim, W., Jang, D., & Lee, T. (2016, December). Speech emotion recognition using convolutional and recurrent neural networks. In 2016 Asia-Pacific signal and information processing association annual summit and conference (APSIPA) (pp. 1-4). IEEE.

[8] Nfissi, A., Bouachir, W., Bouguila, N., & Mishara, B. L. (2022, December). CNN-n-GRU: end-to-end speech emotion recognition from raw waveform signal using CNNs and gated recurrent unit networks. In 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 699-702). IEEE.

[9] Meng, H., Yan, T., Yuan, F., & Wei, H. (2019). Speech emotion recognition from 3D log-mel spectrograms with deep learning network. IEEE access, 7, 125868-125881.

[10] Koduru, A., Valiveti, H. B., & Budati, A. K. (2020). Feature extraction algorithms to improve the speech emotion recognition rate. International Journal of Speech Technology, 23(1), 45-55.

[11] Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., ... & Wu, Y. (2018, April). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4779-4783). IEEE.Burkhardt,

[12] F.,Paeschke,A.,Rolfes,M.,Sendlmeier,W.F.,&Weiss, B. (2005).Adatabase of German emotional speech.In Ninth European Conference on Speech Communicationand Technology.

[13] Hadjadji, I., & Falek, L. (2024). Evaluating degradation in emotional speech quality over a 4G telephone network. Multimedia Tools and Applications, 83(20), 58075-58102.

[14] Salem, F.M. (2022). Gated RNN: The Gated Recurrent Unit (GRU) RNN. In: Recurrent Neural Networks. Springer, Cham. https://doi.org/10.1007/978-3-030-89929-5 5.

[15] Mangal, S., Joshi, P., & Modak, R. (2019). LSTM vs. GRU vs. Bidirectional RNN for script generation. arXiv preprint arXiv:1908.04332.

[16] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. In Ninth European Conference on Speech Communication and Technology

[17] Pham, M. H., Noori, F. M., Torresen, J. (2021, December). Emotion recognition using speech data with convolutional neural network. In 2021 IEEE 2nd international conference on signal, control and communication (SCC) (pp. 182-187). IEEE.

[18] Chauhan, K., Sharma, K. K., Varma, T. (2021, March). Speech emotion recognition using convolution neural networks. In 2021 international conference on artificial intelligence and smart systems (ICAIS) (pp. 1176-1181). IEEE.

A COMPUTATIONAL FRAMEWORK FOR AUTOMATIC GROUND CONTROL POINT IDENTIFICATION IN SATELLITE IMAGERY

Oussama Mezouar, Fatiha Meskine

RCAM Laboratory, Djillali Liabes University, Sidi Bel Abbes, Algeria

ABSTRACT

Ground Control Points (GCPs) play a crucial role in remote sensing applications, particularly for accurate image processing. However, obtaining GCPs through traditional methods such as ground surveys or GPS measurements is often impractical in inaccessible or remote regions due to logistical challenges and limited ground access. This study presents an innovative framework for the automatic detection and extraction of GCPs directly from satellite imagery, addressing these challenges. The proposed method comprises two key phases: data pre-processing and GCP detection and extraction. In the pre-processing phase, image quality is enhanced using advanced band selection and contrast optimization techniques. In the second phase, GCP identification and extraction leverage a hybrid approach combining the strengths of SURF and SIFT algorithms for feature matching. The framework was validated using ALSAT-2 satellite imagery under complex conditions, demonstrating its efficiency and accuracy in identifying and extracting GCPs. These findings underscore the potential of the proposed method to streamline remote sensing workflows and improve image rectification in challenging environments.

Index Terms— Remote sensing images, GCPs, SURF, SIFT.

1. INTRODUCTION

Remote sensing imagery has become increasingly essential for a variety of applications, such as land use and land cover mapping, agriculture, forestry, urban planning, archaeological exploration, military reconnaissance, and geological surveying [1]. Remote sensing, defined as the process of collecting information about objects or phenomena without direct contact, offers a wealth of data stored within its captured imagery. However, the effective use of this data is often hindered by challenges in processing raw images, extracting meaningful information with precision, and applying it to practical scenarios. Poor image quality or suboptimal feature extraction can significantly reduce the impact and utility of remote sensing applications[2].

The raw data obtained from remote sensing platforms is not inherently geospatial due to geometric distortions introduced by sensor movements and acquisition conditions. To ensure the data aligns with real-world coordinate systems, geometric corrections—geo-referencing—are critical. This pre-processing step transforms satellite images into reliable geospatial datasets suitable for accurate location-based analyses. A cornerstone of this process is the identification and utilization of Ground Control Points (GCPs), which are key features that serve as reference points for geometric correction[3].

GCP selection is pivotal for improving the accuracy of satellite imagery. However, manually identifying GCPs often relies on expert judgment, which can vary between individuals, and this process may be infeasible in regions with challenging conditions such as cloud cover, oceans, or rugged terrains. Automatic detection and extraction algorithms have emerged as a solution, enabling efficient and consistent identification of GCPs without human intervention[4].

Modern feature detection and matching techniques form the backbone of automated GCP extraction. These methods, such as SIFT (Scale-Invariant Feature Transform) and SURF (Speeded-Up Robust Features)[5], are known for their robustness, efficiency, and adaptability to changes in scale, rotation, and illumination. SIFT has been widely employed in remote sensing for its resilience to variations in lighting and its suitability for real-time applications[6]. Meanwhile, SURF offers faster computation, making it a competitive alternative in feature extraction tasks.

In this work, we propose an automated framework for detecting and extracting GCPs by leveraging the complementary strengths of SIFT and SURF algorithms. The process begins with pre-processing to enhance satellite image quality using CLAHE (Contrast Limited Adaptive Histogram Equalization). The hybrid SURF-SIFT approach is then applied to extract and match key points, followed by the

RANSAC (Random Sample Consensus) algorithm to refine results and eliminate false matches[7,8].

2. METHODOLOGY FRAMEWORK

This study proposes an automated framework for detecting and extracting Ground Control Points (GCPs) from remote sensing data. The methodology is divided into two primary phases: Data Pre-processing and Identification and Extraction of GCPs. The structure of the proposed framework is illustrated in Figure 1.

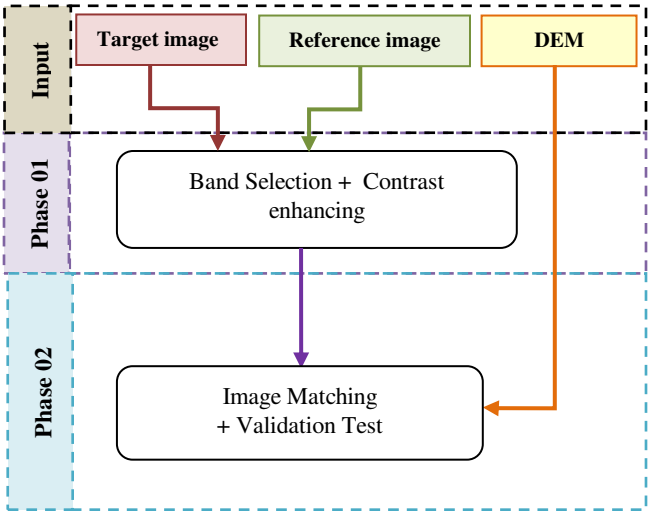


Fig.1. The proposed framework.

2.1 Dataset Description (Input)

The research utilizes the ORAN dataset, acquired from ALSAT-2, Algeria's first high-resolution satellite. This dataset poses significant challenges for GCP detection due to the presence of mountainous terrain, which introduces variability in elevation and terrain features. These variations require accurate detection and matching of GCPs to ensure reliable geo-referencing[9].

The ORAN dataset is multispectral with a spatial resolution of 10 meters and dimensions of 1750 × 1750 pixels.

To provide reliable reference data for GCP extraction, LANDSAT images with a spatial resolution of 15 meters are employed. Known for their high-resolution ortho-image references, LANDSAT images facilitate accurate alignment and verification. By leveraging multispectral ALSAT-2 data and referring to LANDSAT images, this framework addresses the challenges posed by complex terrain features in the ORAN dataset.



Fig.2. ALSAT2 target image

2.2 Data pre-processing (Phase 01)

This phase enhances the quality of the satellite image to optimize GCP detection and extraction. It involves two key steps: Band Selection and Contrast Enhancement.

a) Band Selection:

Optimal bands (channels) are selected to maximize contrast and improve the visibility of key features. The Contrast Ratio (CR) is computed for each RGB band using Peli's formula: [10].

$$CR = \sqrt{\frac{1}{M \times N}\sum_{y=1}^{N}\sum_{x=1}^{M}(I(x,y)-\beta)^2} \quad (1)$$

Where x,y is the index pixel in the image I with M, N dimension, β defines the brightness:

$$\beta = \frac{1}{M \times N}\sum_{y=1}^{N}\sum_{x=1}^{M}(I(x,y)) \quad (2)$$

This band selection step aims to identify the bands that provide the most distinguishable features or contrasts in the images, as these are crucial for accurately detecting and extracting the GCPs in the subsequent phases of the study.

b) Contrast Enhancing

To further enhance image quality, two techniques are applied:

b.1) Contrast Limited Adaptive Histogram Equalization (CLAHE): Improves local contrast by limiting contrast amplification in small regions (tiles), enhancing edges and fine details[11].

b.2) Histogram Equalization: Adjusts the intensity distribution in the image histogram to enhance overall contrast, particularly in low-contrast areas[12].

These methods improve gray-level distribution, preparing the ORAN dataset for effective GCP extraction.

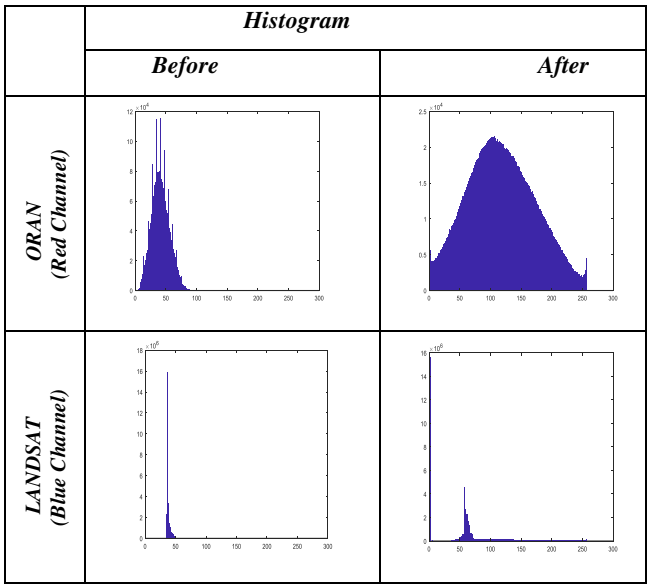


Fig.3. Histogram results.

2.3 Identification and Extraction of GCPs (Phase 02)

The second phase focuses on detecting and extracting GCPs using a mixed algorithm combining SURF (Speeded-Up Robust Features) for detection and SIFT (Scale-Invariant Feature Transform) for descriptor matching. [13,14-17].

a) Feature Detection and Matching

SURF Detector: Selected for its robustness against scale, blur, illumination changes, and partial invariance to rotation. Its speed and efficiency make it ideal for processing remote sensing images[18].
SIFT Descriptor: Used for its high accuracy in matching key points under varying conditions[19].

b) Validation and Outlier Removal

Lowe's Ratio Test: Filters out unreliable matches by retaining only those with sufficiently distinct matches.
RANSAC Algorithm: Further refines matches by identifying and removing outliers, ensuring only accurate GCPs are selected[20].

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1 Contrast Ratio Analysis

The contrast ratio for the three bands of the input images was analyzed, and the results are presented in Table 1. For the ALSAT-2 dataset, the red band exhibited the highest contrast ratio, making it the most suitable for GCP detection. Conversely, for the LANDSAT dataset, the blue band provided the highest contrast ratio, suggesting its use for the reference data.

Table 1.Contrast ratio values of different bands.

Datasets		ALSAT2 ORAN	LANDSAT
Contrast Ratio	**Red**	**0.0581**	0.0578
	Green	0.0362	0.0584
	Blue	0.0263	**0.0657**

3.2 Pre-Processing Results

Figure 3 illustrates the results of the pre-processing phase, showcasing a significant improvement in image quality after applying CLAHE and Histogram Equalization. The histogram representation confirms a more balanced and uniform spread of gray levels, resulting in enhanced image details. The improvements made during this phase provide a solid foundation for accurate GCP detection in the subsequent steps.

3.3 Key Point Detection and Descriptor Matching

To determine the optimal combination for key point detection and matching, the study evaluated various descriptor algorithms, including SIFT, BRIEF, and BRISK, alongside the SURF detector. A brute-force matcher was used to find correspondences between key points.

The SURF-SIFT combination was identified as the most effective, as it demonstrated superior performance in detecting valid GCPs, particularly in challenging conditions. Table 2 compares the results of the SURF-SIFT and SURF-SURF combinations across all datasets.

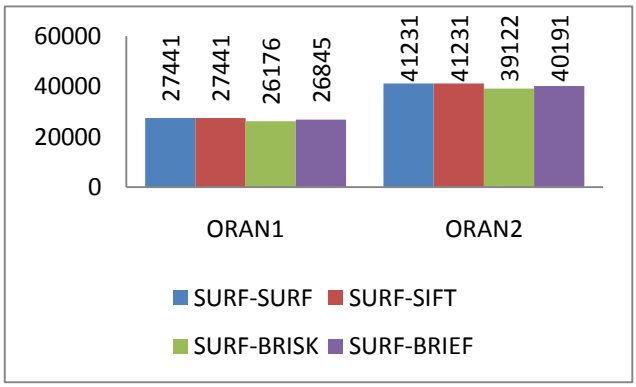


Fig. 4. Number of matching points using different descriptor

Table 2. Inliers detected points.

	Competitive method (SURF-SURF)		Proposed method (SURF-SIFT)	
	Inliers	**Outliers**	**Inliers**	**Outliers**
ORAN 2	34	95	31	12

While both methods showed satisfactory performance in the ORAN dataset, the SURF-SIFT approach achieved a robust balance between accuracy and computational efficiency. Across all datasets, the SURF-SIFT method detected 36.5% more valid points compared to the conventional SURF approach, underscoring its overall superiority.

3.4 GCP Detection in ORAN Dataset

In the ORAN dataset, the proposed SURF-SIFT algorithm demonstrated comparable results to the conventional SURF method, detecting 31 valid points compared to 34 points detected by SURF. This close performance suggests that both methods are effective for GCP extraction in mountainous terrains, with SURF-SIFT maintaining an edge in challenging scenarios due to its robustness and ability to filter outliers effectively.

3.5 Visual Results of GCP Detection

Figure 5 shows the GCPs detected in the ORAN dataset. The proposed framework successfully identifies GCPs with high accuracy, ensuring reliable geo-referencing even in complex terrain conditions.

4. CONCLUSION

This paper presented an automated framework for the detection and extraction of Ground Control Points (GCPs) from remote sensing data. The proposed method consisted of two main phases:

Data Pre-processing Phase: Advanced histogram techniques, including Contrast Limited Adaptive Histogram Equalization (CLAHE) and Histogram Equalization, were applied to enhance image contrast, significantly improving the reliability of the GCP detection process.

GCP Identification and Extraction Phase: A combination of the SURF detector and SIFT descriptor was employed to accurately match key points between ALSAT-2A and LANDSAT satellite images.

The experimental results highlight the effectiveness of the proposed method in automatically detecting a sufficient number of reliable GCPs, even in challenging conditions such as mountainous terrain. This method provides a practical alternative to conventional approaches that require physical access to collect GCPs, reducing time and labor-intensive processes.

The developed framework demonstrates great potential for remote sensing applications, particularly for geo-referencing tasks. By automating the GCP detection process, the proposed method enhances accuracy and efficiency, making it a valuable tool for addressing challenges in complex and hard-to-reach environments.

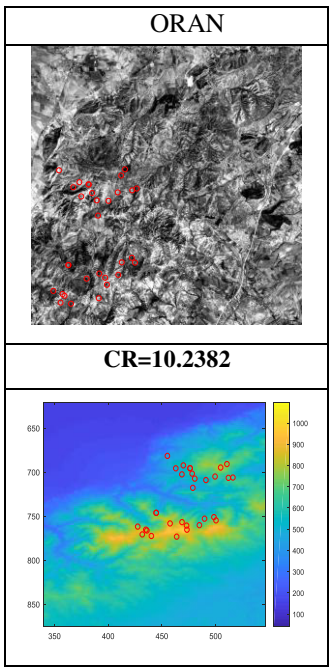


Fig.5. The detected GCPs on the image.

REFERENCES

[1] Freek D. Van der Meer, "Remote Sensing and Digital Image Processing", Springer 2023.

[2] Anju Asokan, J. Anitha, Monica Ciobanu, Andrei Gabor, Antoanela Naaji and D. Jude Hemanth, "Image Processing Techniques for Analysis of Satellite Images for Historical Maps Classification--An Overview", *Applied Science*, Vol.10, N.12, 2020.

[3] Thanh Nguye, "Optimal ground Control points for Geometric Correction Using Genetic Algorithm with Global Accuracy", *European Journal of Remote Sensing*, Vol.48, pp. 101-120, 2015.

[4] X. Dai, and K. Siamak, "The Effects of Image Misregistration on the Accuracy of Remotely Sensed Change Detection, *IEEE Trans. Geoscience & Remote Sensing*, Vol.*36*, pp.1566–1577, 1998.

[5] Xiangjun Wang, Yang Li, Hong Wei and Feng Liu, "An ASIFT-Based Local Registration Method for Satellite Imagery", *Remote Sensing,*, Vol.*7*, pp. 7044-7061, 2015. doi:10.3390/rs70607044

[6] Tinne Tuytelaars and Krystian Mikolajczyk, "Local Invariant Feature Detectors: A Survey", Foundations and trends in computer graphics and vision, vol. 3, no. 3, pp. 177–280, 2008.

[7] Z. Yi, C. Zhiguo, X. Yang, X, "Multi-spectral remote image registration based on SIFT", *Electronic Letters*, Vol.*44*, pp. 107–108, 2008.

[8] Beril Sırmaçek, and Cem ÜnsalanUrban, "Area and Building Detection Using SIFT Keypoints and Graph Theory", *IEEE Trans. Geoscience & Remote Sensing*, VOL. 47, NO. 4, APRIL 2009.

[9] A. Elashry, B., Sluis, C. Toth, "Improving RANSAC Feature Matching based on Geometric Relation", *The International Archives of the Photogrammetry, Remote Sensing and*

Spatial Information Sciences, Volume XLIII-B2-2021, pp. 321-327, 2021.

[10] E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A*, vol. 7, no. 10, pp. 20-32, Oct. 1990, doi: 10.1364/JOSAA.7.002032.

[11] I. Boukerch, B. Takarli, A. Boudaouie, and A. Chaachoue, " A Framework for Geometric Quality Evaluation and Enhancement of Alsat-2A Satellite Imagery", in *Sensors, Systems, and Next-Generation Satellites XXIII*,, SPIE, 2019, pp. 501-508.

[12] S. Pizer, R. Johnston, J. Ericksen, B. Yankaskas, and K. Muller, "Contrast-limited Adaptive Histogram Equalization: Speed and Effectiveness," in *Proc. 1st Conf. Visual. Biomed. Comput*., May 1990, pp. 337– 345.

[13] B. Kalantar, N. Ueda, H. A. H Al-Najjar, H. A. H., H. Moayedi, A. A Halin, and S. Mansor, "UAV and Lidar Image Registration: A SURF-Based Approach for Ground Control Points Selection", Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, pp. 413–418, 2019.

[14] D. Liu, G. Zhou, D. Zhang, X. Zhou and C. Li, "Ground Control Point Automatic Extraction for Spaceborne Georeferencing Based on FPGA," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3350-3366, 2020

[15] Y. Angel, D. Turner, S. Parkes, Y. Malbeteau, A. Lucieer, MF. McCabe, "Automated Georectification and Mosaicking of UAV-Based Hyperspectral Imagery from Push-Broom Sensors", *Remote Sensing,* Vol.12(1):34, 2020.

[16] L. Rasmy, I. Sebari, M. Ettarid, "Automatic Sub-Pixel Co-Registration of Remote Sensing Images Using Phase Correlation and Harris Detector", *Remote Sensing*. 2021; 13(12):2314.

[17] M.N Alrajhi, K. Jacobsen, C. Heipke, "A Procedure for Semi-Automatic Ortho-photo Generation from High Resolution Satellite Imagery", In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 40 (2013), Nr. 7W2, S. 7-11.

[18] C. Alhwarin, C. Wang, D. Ristic-Durrant, and A. Graser, "Improved SIFT-Features Matching for Object Recognition," in Proc. Int. Academic Conf. Vision of Computer Science-BSC, 2008, pp. 179–190.

[19] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in *Proc. Eur. Conf. Comput. Vision*, Graz, Austria, May 2006, pp. 404–417.

[20] I. Misra, S. M. Moorthi, D. Dhar, and R. Ramakrishnan, "An Automatic Satellite Image Registration Technique Based on Harris Corner Detection and Random Sample Consensus (RANSAC) Outlier Rejection Model," 1st International Conference on Recent Advances in Information Technology (RAIT) 2012, India, pp. 68–73.

A Dual Forward Correlated Set-Membership NLMS Algorithm for Noise Reduction

Abdelhak Cheffi[1], Akila Sayoud[2] , Rahima Henni [2] and Mohamed Djendi[1]

[1]Signal Processing and Image (LATSI) Laboratory, University of Blida 1, Blida 09000, Algeria. Emails: cheffi.abdelhak@yahoo.com, m_djendi@yahoo.fr
[2]LCPTS Laboratory, Faculty of Electrical Engineering, USTHB University, Algiers, Algeria. Emails: Akila.sayoud@gmail.com, henni93rahima@gmail.com

Abstract—**This work addresses the problem of acoustic noise reduction and speech enhancement employing set membership adaptive filtering algorithms combined with the forward blind source separation (FBSS) structure. We propose a new dual forward set membership algorithm for acoustic noise reduction and speech quality enhancement application. This proposed algorithm is established on the use of the cross-correlation of the output error signal and the noisy to control the variable-step-size in time domain. We have focused on the comparison between our proposed algorithm with it conventional version dual forward set membership NLMS (DF-SM-NLMS), in terms of speed convergence and stability. In order to examine the good performance of this proposed algorithm, we have evaluated it through three objective criteria: System Mismatch (SM), Segmental Mean Square Error (SegMSE) and the Segmental SNR (SegSNR).**

Keywords—speech enhancement, adaptive algorithm, convergence speed, correlation, set membership, MSE.

1. INTRODUCTION

In general, several telecommunication systems such as hand-free telephone, conferencing systems and hearing aids employ the adaptive filter approach [1-2]. Among these adaptive algorithms that are often used in these applications are the least mean square (LMS), normalized least mean square (NLMS), affine projection (AP) and recursive least squares (RLS) algorithms. In real time applications, the computational complexity of these adaptive algorithms play a very important role and have to be taken in consideration. The families of adaptive filtering algorithms introduced so far present a compromise between the speed of convergence and the computational complexity. Several approaches have been proposed to the problem of the computational complexity. In our work, we have carried on one of the most important approaches that reduce the computational complexity of any adaptive algorithm which is the set membership filtering (SMF) [1]. As has been announced earlier, the SMF reduces computational complexity in adaptive filtering, since the filter coefficients are only updated when the estimate of the output error is greater than a predetermined upper limit. Several techniques have been proposed to deal with the problem of acoustic noise reduction and speech enhancement applications. For instance, several single and dual microphones based on adaptive techniques are proposed to correct the distortions of the speech signal. One of the most popular techniques for acoustic noise reduction (ANR) is the blind source separation technique (BSS) [3-4]. This latter is used to restore the original source signals using only noisy observations as information. In this paper, we propose a new version of the dual forward correlated set membership normalized least mean square (DFC - SM-NLMS) algorithm for ANC and speech enhancement applications. This paper is organized as follows: section 2 describe the dual mixing model. The mathematical derivation of the proposed DFC-SM-NLMS algorithm is given in section 3. The simulations results of the proposed DFC-SM-NLMS algorithm are presented in Section 4, and finally a conclusion is given in Section 5.

2. DUAL MIXING MODEL

In this section, we present the dual mixing model. It allows to mix two uncorrelated sources that are the speech signal $s(k)$ and the punctual noise $b(k)$ as shown in Fig. 1 [5-6].

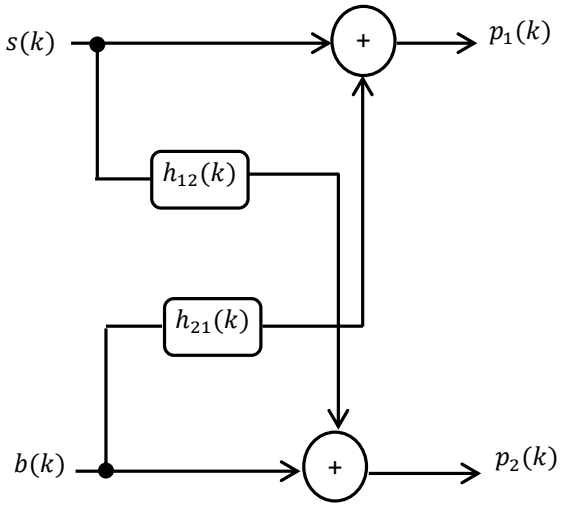


Figure 1.The dual mixing model [5-6]

Where $h_{12}(k)$ and $h_{21}(k)$ represent the mutual coupling effects. The observed signals $p_1(k)$ and $p_2(k)$ of this model are given by:

$$p_1(k) = s(k) + b(k) * h_{21}(k) \quad (1)$$

$$p_2(k) = b(k) + s(k) * h_{12}(k) \quad (2)$$

3. PROPOSED DFC-SM-NLMS

The mathematical derivation of the proposed DFC-SM-NLMS algorithm is presented in this section. In the first, the dual mixing signals $p_1(k)$ and $p_2(k)$ obtained by Fig.1 will be the inputs of the forward BSS structure given in Fig 2. According to this figure, the outputs signal $u_1(k)$ and $u_2(k)$ are given by the following relations:

$$u_1(k) = p_1(k) - p_2(k) * w_{21}(k) \quad (3)$$

$$u_2(k) = p_2(k) - p_1(k) * w_{12}(k) \quad (4)$$

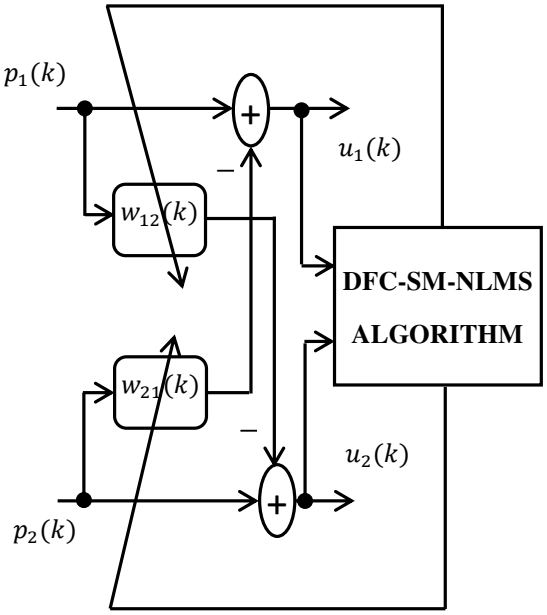


Figure 2.The forward structure

The insertion of equations (1) and (2) in (3) and (4) respectively, and using the optimality assumption for both adaptive filters, $i.e. w_{12}^{opt}(k) = h_{12}(k)$ and $w_{21}^{opt}(k) = h_{21}(k)$ led to the following equations [7]

$$u_1(k) = s(k) * [\delta(k) - h_{12}(k) * w_{12}(k)] \quad (5)$$

$$u_2(k) = b(k) * [\delta(k) - h_{21}(k) * w_{21}(k)] \quad (6)$$

According to these last relations, we show clearly that the output signals $u_1(k)$ and $u_2(k)$ converge respectively to the original signals $s(k)$ and $b(k)$. We have noticed a small distortion.by using the post-filters at the output of the forward BSS structure, we have avoid these distortions [7].

The aim of the proposed DFC-SM-NLMS algorithm is to minimize the Euclidean distance $\|\boldsymbol{w}_{21}(k) - \boldsymbol{w}_{21}(k-1)\|^2$ and $\|\boldsymbol{w}_{12}(k) - \boldsymbol{w}_{12}(k-1)\|^2$, subject to constraint of zero a posterior errors. An additional constraint is to perform a test to check if the previous estimate of $\boldsymbol{w}_{21}(k)$ and $\boldsymbol{w}_{12}(k)$ are respectively outside the constraint sets $\boldsymbol{H_1}(k)$ and $\boldsymbol{H_2}(k)$ that are given by the following relations:

$$\boldsymbol{H_1}(k) = \{\boldsymbol{w}_{21}(k) \in R^N : |u_1(k)| \le \gamma_1 \} \quad (7)$$

$$\boldsymbol{H_2}(k) = \{\boldsymbol{w}_{12}(k) \in R^N : |u_2(k)| \le \gamma_2 \} \quad (8)$$

Where N represente the number of coefficients filter, and the two constants γ_1 and γ_2 represent the threshold error.

In the proposed algorithm, the main idea is to control the variable step-sizes $\mu_{cor1}(k)$ and $\mu_{cor2}(k)$ by using respectively the cross-correlation between the noisy signal $\boldsymbol{p}_1(k)$ and the output error signal $u_2(k)$, and the cross-correlation between the noisy signal $\boldsymbol{p}_2(k)$ and the output error signal $u_1(k)$. The expressions of the variable step-sizes are given by the following relations:

$$\mu_{cor1}(k) = \begin{cases} 1 - \frac{\gamma_1}{\|u_2(k)\boldsymbol{p}_1^T(k)\|} & if \quad |u_2(k)| > \gamma_1 \\ 0 & else \end{cases} \quad (9)$$

$$\mu_{cor2}(k) = \begin{cases} 1 - \frac{\gamma_2}{\|u_1(k)\boldsymbol{p}_2^T(k)\|} & if \quad |u_1(k)| > \gamma_2 \\ 0 & else \end{cases} \quad (10)$$

The new updating relations are given by the following relations:

$$\boldsymbol{w}_{12}(k) = \boldsymbol{w}_{12}(k-1) + \mu_{cor1}(k)\boldsymbol{\beta}_1(k) \quad (11)$$

$$\boldsymbol{w}_{21}(k) = \boldsymbol{w}_{21}(k-1) + \mu_{cor2}(k)\boldsymbol{\beta}_2(k) \quad (12)$$

Where

$$\boldsymbol{\beta}_1(k) = \frac{u_2(k)\boldsymbol{p}_1(k)}{\|\boldsymbol{p}_1(k)\|^2} \quad (13)$$

$$\boldsymbol{\beta}_2(k) = \frac{u_1(k)\boldsymbol{p}_2(k)}{\|\boldsymbol{p}_2(k)\|^2} \quad (14)$$

4. SIMULATION RESULTS

The analysis simulation results of the new proposed algorithm are presented in this section. We are interesting on the comparison between the proposed DFC-SM-NLMS algorithm and the classical DF-SM-NLMS [2, 8] algorithm in terms of stability and convergence speed. Two types of noise are used in our simulations (i.e. White Gaussian noise WGN, and United States of America Standard Institute (USASI) noise). All the obtained simulations results are performed at a sampling frequency of $Fs = 8\,kHz$.

Figure 3 illustrate two examples of the real impulse responses $h_{12}(k)$ and $h_{21}(k)$ that we had used in our simulations. The length of each impulse response is 64. The original speech signal and its manual voice activity detector (MVAD), and an example of the mixture signal $\boldsymbol{p}_1(k)$ that we have used in these simulations are presented in fig.4.

Table 1. Simulation parameters of DF-SM-NLMS and Proposed algorithm.

Algorithms	Simulation parameters
DF-SM-NLMS [2]	**Initial step-sizes :** $\mu_1 = \mu_2 = 0.1$ **Adaptive filter length** $\boldsymbol{w}_{12}, \boldsymbol{w}_{21}$: $N = 64$ **Fixed threshold error** $\gamma_1 = \gamma_2 = 15$ and 90 Input $SNR1 = 3$ dB, Input $SNR2 = 3$ dB
Proposed algorithm [In this paper]	**Initial step-sizes :** $\mu_1 = \mu_2 = 0.1$ **Adaptive filter length** $\boldsymbol{w}_{12}, \boldsymbol{w}_{21}$: $N = 64$ **Fixed threshold error** $\gamma_1 = \gamma_2 = 15$ and 90 Input $SNR1 = 3$ dB, Input $SNR2 = 3$ dB

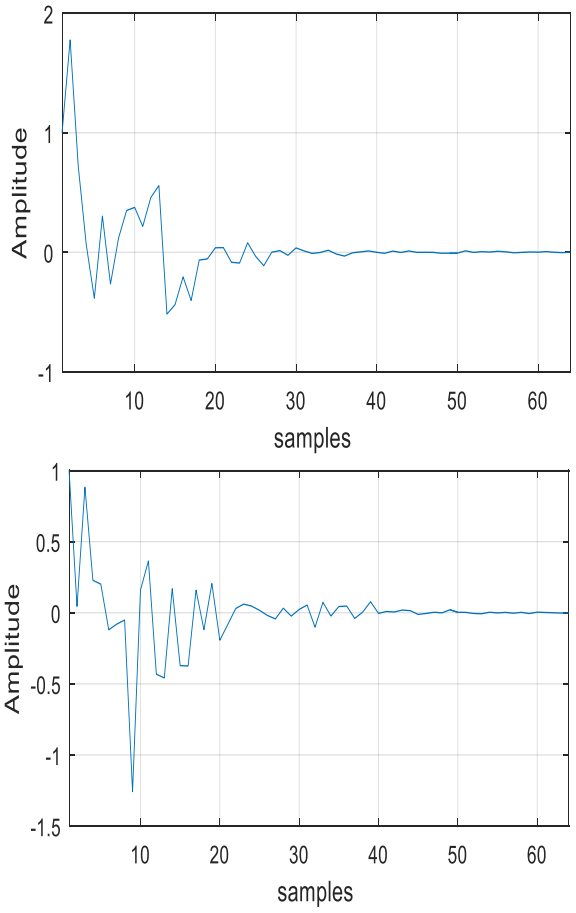


Figure 3. Examples of Impulse responses $h_{12}(k)$,$h_{21}(k)$ [from top to bottom].

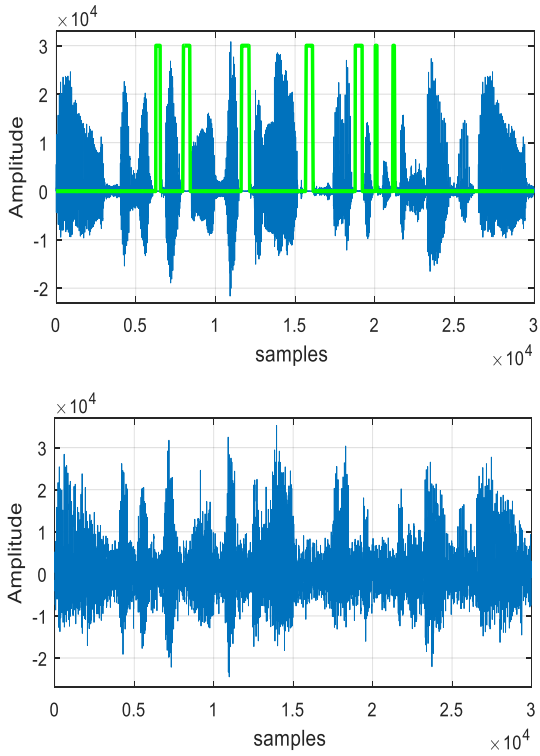


Figure 4. Original speech signal and its MVAD, and noisy signal [from top to bottom].

Temporal evaluation

The time evolution of the output speech signal of the proposed algorithm have been evaluated, and the obtained results are reported on Fig. 5.

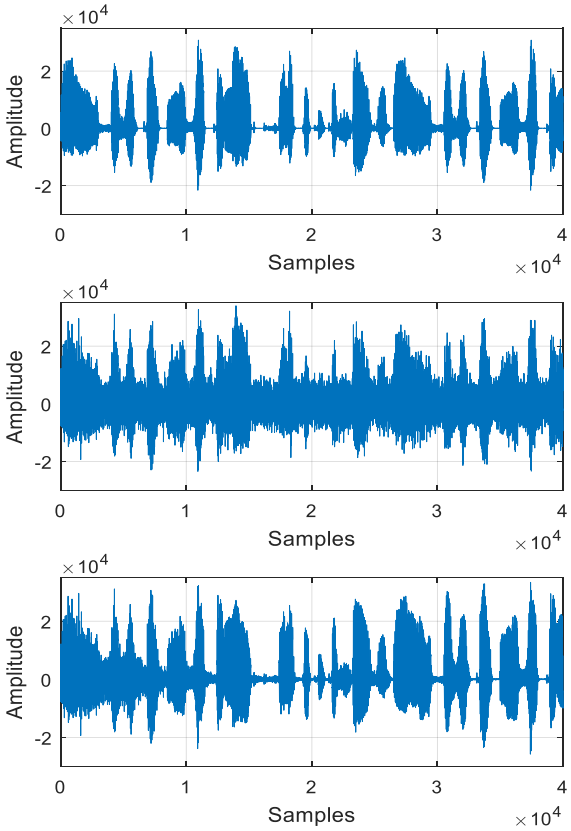


Figure 5. Time evolution of the output signal. From top to bottom: original speech signal, mixture signal, the output speech signal obtained by Proposed DFC-SM-NLMS.

All simulation parameters are summarized in table 1. We notice that the noise is sufficiently canceled by the proposed algorithm. So the obtained results show the good performance of this algorithm for noise cancellation and the speech signal enhancement.

System mismatch (SM) evaluation

In this subsection, the evaluation of the system mismatch criterion (SM) for the both algorithms (.i.e. proposed algorithm and the classical DF-SM-NLMS) is presented. We carried only on the output $u_1(k)$ and the adaptive filter $\boldsymbol{w}_{21}(k)$.The SM is given by the following relation:

$$SM_{dB} = 20Log_{10}\left(\frac{\|\boldsymbol{h}_{21}(k)-\boldsymbol{w}_{21}(k)\|}{\|\boldsymbol{h}_{21}(k)\|}\right) \qquad (20)$$

Where $h_{21}(k)$ and $w_{21}(k)$ represent respectively the real impulse response and adaptive filters vectors and $\|.\|$ represents the norm-2. All simulation parameters of each algorithms are summarized in Table 1.The obtained results are reported on Fig. 6.

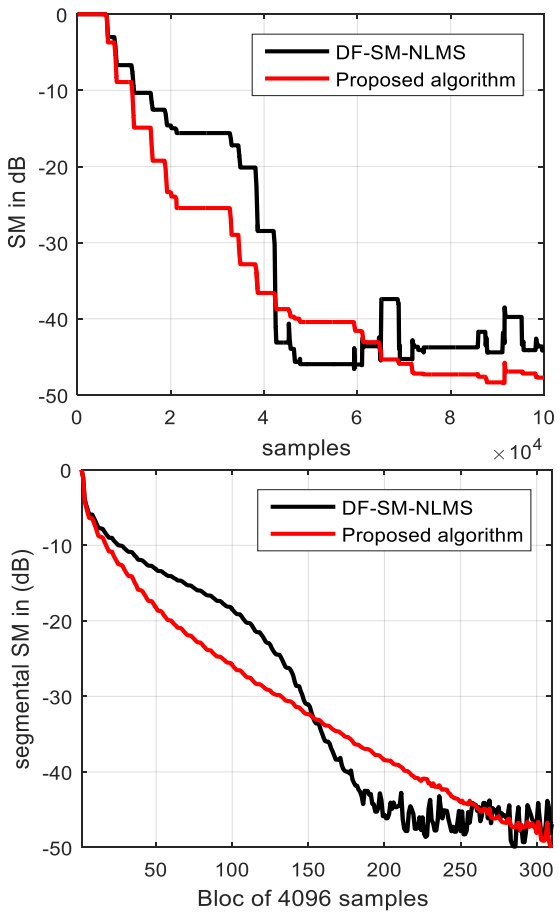


Figure 6. System mismatch evaluation of DF-SM-NLMS and Proposed algorithm. Input signals: speech signal and WGN and speech signal in the top. Input signals: speech signal and USASI.

According to the evaluation of the system mismatch (SM), we can clearly observe the good behavior of the proposed algorithm in comparison with the classical DF-SM-NLMS ones.

Segmental Mean Square Error SegMSE evolution

This subsection illustrate another criterion. The evaluation of segmental mean square error SegMSE is presented in this simulation, comparing the proposed algorithm with the original DF-SM-NLMS version. The SegMSE criterion is computed by the following relation;

$$SegMSE_{dB} = 20Log_{10}(\sum_{k=0}^{M-1} u_1(k)) \qquad (21)$$

Where M symbolize the time averaging frame length of output $u_1(k)$. The presence of the voice activity detector $VAD(k)$ means that the SegMSE criterion is estimated only in the absence of speech segments. The obtained results are reported on Fig 7.

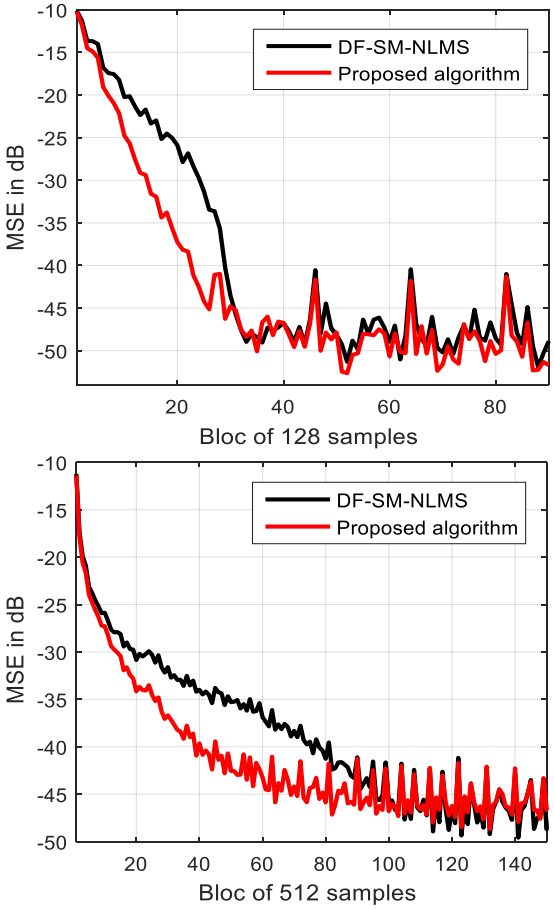


Figure 7. Mean square error evaluation of DF-SM-NLMS and Proposed algorithm. Input signals: speech signal and WGN and speech signal in the top. Input signals: speech signal and USASI.

The simulation results of the segmental mean square evaluation (SegMSE) given by fig 7, shows the best performance of the proposed DFC-SM-NLMS algorithm in term of speed convergence.

Output Segmental SNR evolution

In this subsection, we describe the analysis of the output segmental signal to noise ratio (SegSNR), comparing the proposed algorithm with the original version in term of stability. Figure 8 represents the evaluation of the SegSNR output of the both algorithms (i.e. DF-SM-NLMS, and Proposed DFC-

SM-NLMS algorithm).The SegSNR criterion is given by the following relation:

$$SNR_{dB} = \frac{1}{Q}\sum_{i=0}^{Q-1} 10Log_{10}\left(\frac{\sum_{k=0}^{M-1}|s_i(k)|^2}{\sum_{k=0}^{M-1}|s_i(k)-u_{1,i}(k)|^2}\right) (22)$$

Where $s(k)$ and $u_1(k)$ are the original speech signal and the enhanced one, respectively. The parameter M is the mean averaging value of the output SNR. The parameter Q is the number of only-speech periods.

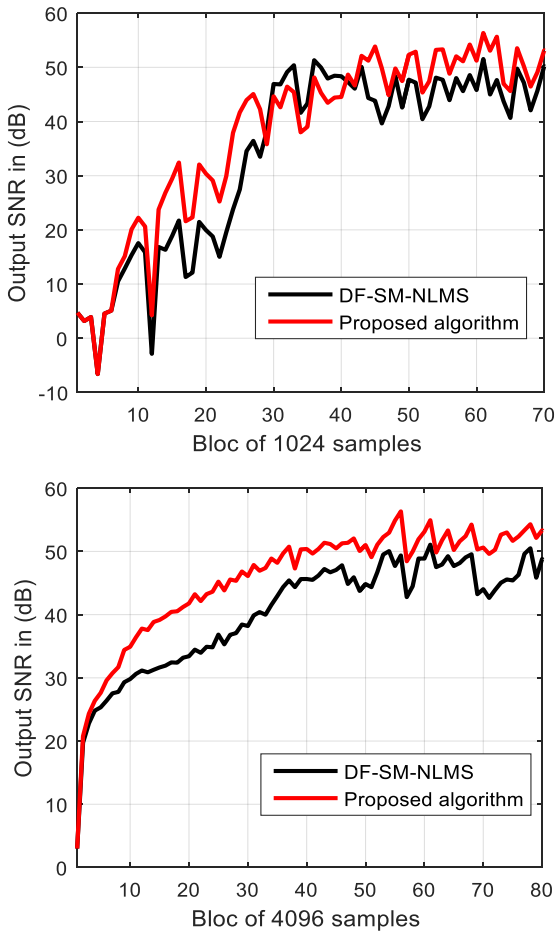


Figure 8. Output signal to noise evolution evaluation of DF-SM-NLMS and Proposed algorithm. Input signals: speech signal and WGN and speech signal in the top. Input signals: speech signal and USASI.

5. CONCLUSION

In this paper, we have proposed new version of two-channel forward partial update algorithm for acoustic noise reduction and speech quality enhancement application. We have carried out a comparison between the classical DF-SM-NLMS algorithm and the proposed algorithm to show clearly the performance of convergence speed and stability. In order to do a fair comparison between the both algorithms, three objective criteria have used, these criteria are: the system mismatch (SM), the segmental mean square error (SegMSE) and of the output segmental signal to noise ratio (SegSNR). According to the obtained results, we can well see that the proposed algorithm behaves more efficiently than the other one. All the obtained results confirmed the superiority of the proposed algorithm in speech enhancement and acoustic noise reduction application.

REFERENCES

[1] P. S. R. Diniz, "Data-selective adaptive filtering", *Adaptive Filtering*, Springer, 2008.

[2] P. S. R. Diniz, R. P. Baraga, and S. Werner, "Set-membership affine projection algorithm for echo cancellation," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2006.

[3] Djendi M, Scalart P, Gilloire A. Analysis of two-sensors forward BSS structure with post-filters in the presence of coherent and incoherent noise. Speech Comm. 2013; 55(10):975–987.M.

[4] Djendi M, Bendoumia R. A new adaptive filtering subband algorithm for two channel acoustic noise reduction and speech enhancement. ComputElectr Eng. 2013; 39(8):2531–2550.

[5] A. Cheffi, M. Djendi, and A. Guessoum A new correlated set-membership partial-update NLMS (SM-PU-NLMSCOR) algorithm for acoustic noise reduction. In: 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B) 2017, Boumerdes, Algeria; 2017.

[6] Abdelhak Cheffi, Mohamed Djendi. New Two Channel Set-Membership Partial-Update NLMS Algorithms for Acoustic Noise Reduction. In: 4th international conference on Signal, Image, Vision and their Applications (SIVA'18), at Guelma University. Algeria, November 2018.

[7] Abdelhak Cheffi, Mohamed Djendi, Abderrezak Guessoum. New efficient two channel forward set-membership partial-update NLMS algorithms for blind speech enhancement and acoustic noise reduction. Appl Acoust 2018; 141:322–332.

[8] Diniz, P.S.R. Adaptive Filtering: Algorithms and Practical Implementaion, 4th ed.; Springer: New York, NY, USA, 2013.

New Symmetric Set Membership Adaptive Filtering for Blind Source Separation

Abdelhak Cheffi[1], Akila Sayoud[2], Rahima Henni [2] and Mohamed Djendi[1]

[1]Signal Processing and Image (LATSI) Laboratory, University of Blida 1, Blida 09000, Algeria. Emails: cheffi.abdelhak@yahoo.com, m_djendi@yahoo.fr

[2]LCPTS Laboratory, Faculty of Electrical Engineering, USTHB University, Algiers, Algeria. Emails: Akila.sayoud@gmail.com, henni93rahima@gmail.com

Abstract—**this paper addresses the problem of the forward blind source separation (FBSS) structure by using set membership adaptive filtering algorithms. We propose a new two sensor forward set membership algorithm for FBSS application. This proposed algorithm focus on the cross-correlation of the output error signal and the noisy to control the variable-step-size in time domain. A comparison between our proposed algorithm with it conventional version two sensors forward set membership NLMS (TS-SM-NLMS) has done to improve the best performance of the proposed one. In order to examine the good performance of this proposed algorithm, we have evaluated it through three objective criteria: System Mismatch (SM), Segmental Mean Square Error (SegMSE) and the Segmental SNR (SegSNR).**

Keywords—blind source separation, adaptive algorithm, convergence speed, correlation, set membership filtering.

1. INTRODUCTION

In general, several telecommunication systems, such as hands-free telephones, conferencing systems, and hearing aids, employ the adaptive filter approach [1-2]. Adaptive algorithms commonly used in these applications include the Least Mean Square (LMS), Normalized Least Mean Square (NLMS), Affine Projection (AP), and Recursive Least Squares (RLS) algorithms. In real-time applications, the computational complexity of these adaptive algorithms plays a critical role and must be taken into consideration. The families of adaptive filtering algorithms introduced so far represent a trade-off between convergence speed and computational complexity. Several approaches have been proposed to address the problem of computational complexity.

In our work, we focus on one of the most effective approaches to reducing the computational complexity of adaptive algorithms: Set Membership Filtering (SMF) [1]. As mentioned earlier, SMF reduces computational complexity in adaptive filtering by updating the filter coefficients only when the estimate of the output error exceeds a predetermined upper limit.

Various techniques have been proposed to address the problem of acoustic noise reduction and speech enhancement. For instance, several single- and two-microphone adaptive techniques have been proposed to correct distortions in speech signals. One of the most popular techniques for acoustic noise reduction (ANR) is Blind Source Separation (BSS) [3-4], which restores the original source signals using only noisy observations as input.

In this paper, we propose a new version of the two-sensor forward correlated Set Membership Normalized Least Mean Square (TSC-SM-NLMS) algorithm for ANR and speech enhancement applications. The paper is organized as follows: Section 2 describes the two-sensor mixing model. The mathematical derivation of the proposed TSC-SM-NLMS algorithm is provided in Section 3. Simulation results of the proposed algorithm are presented in Section 4, and finally, Section 5 presents the conclusion.

2. MIXING MODEL

In this section, we present the two sensor mixing model. It allows to mix two uncorrelated sources that are the speech signal $s(k)$ and the punctual noise $b(k)$ as shown in Fig. 1 [5-6].

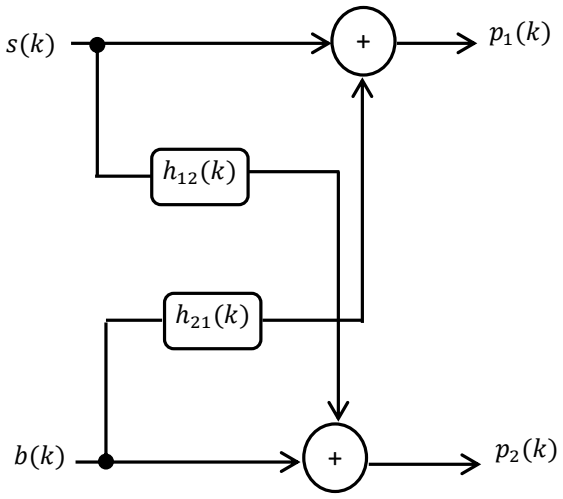


Figure 1.The mixing model [5-6]

Where $h_{12}(k)$ and $h_{21}(k)$ represent the mutual coupling effects. The observed signals $p_1(k)$ and $p_2(k)$ of this model are given by:

$$p_1(k) = s(k) + b(k) * h_{21}(k) \quad (1)$$

$$p_2(k) = b(k) + s(k) * h_{12}(k) \quad (2)$$

3. PROPOSED TSC-SM-NLMS

The mathematical derivation of the proposed algorithm is presented in this section. In the first, the dual mixing signals $p_1(k)$ and $p_2(k)$ obtained by Fig.1 will be the inputs of the forward BSS structure given in Fig 2. According to this figure, the outputs signal $u_1(k)$ and $u_2(k)$ are given by the following relations:

$$u_1(k) = p_1(k) - p_2(k) * w_{21}(k) \quad (3)$$

$$u_2(k) = p_2(k) - p_1(k) * w_{12}(k) \quad (4)$$

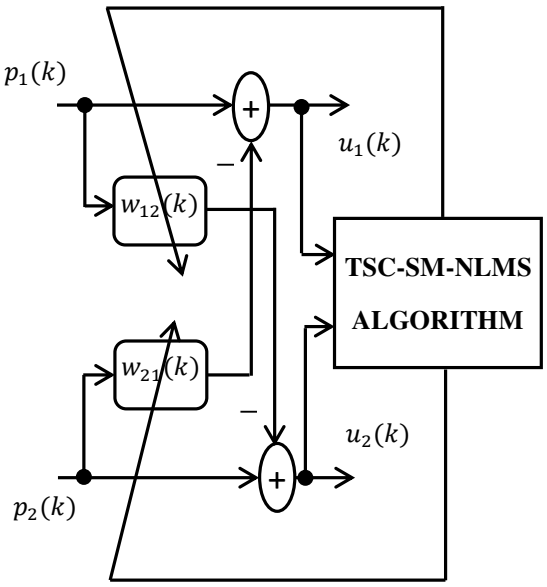


Figure 2.The proposed forward structure

The insertion of equations (1) and (2) in (3) and (4) respectively, and using the optimality assumption for both adaptive filters, $i.e. w_{12}^{opt}(k) = h_{12}(k)$ and $w_{21}^{opt}(k) = h_{21}(k)$ led to the following equations [6]

$$u_1(k) = s(k) * [\delta(k) - h_{12}(k) * w_{12}(k)] \quad (5)$$

$$u_2(k) = b(k) * [\delta(k) - h_{21}(k) * w_{21}(k)] \quad (6)$$

According to these last relations, we show clearly that the output signals $u_1(k)$ and $u_2(k)$ converge respectively to the original signals $s(k)$ and $b(k)$. We have noticed a small distortion.by using the post-filters at the output of the forward BSS structure, we have avoid these distortions [6].

The aim of the proposed TSC-SM-NLMS algorithm is to minimize the Euclidean distance subject to constraint of zero a posterior errors. An additional constraint is to perform a test to check if the previous estimate of $\boldsymbol{w}_{21}(k)$ and $\boldsymbol{w}_{12}(k)$ are respectively outside the constraint sets $\boldsymbol{\psi_1}(k)$ and $\boldsymbol{\psi_2}(k)$ that are given by the following relations:

$$\boldsymbol{\psi_1}(k) = \{\boldsymbol{w}_{21}(k) \in R^N : |u_1(k)| \leq \gamma_1 \} \quad (7)$$

$$\boldsymbol{\psi_2}(k) = \{\boldsymbol{w}_{12}(k) \in R^N : |u_2(k)| \leq \gamma_2 \} \quad (8)$$

Where N represent the number of coefficients filter, and γ_1 and γ_2 are the threshold error.

In the proposed TSC-SM-NLMS, our goal is to control the variable step-sizes $\mu_{cor1}(k)$ and $\mu_{cor2}(k)$ by using respectively the cross-correlation between the noisy signal $\boldsymbol{p}_1(k)$ and the output error signal $u_2(k)$, and the cross-correlation between the noisy signal $\boldsymbol{p}_2(k)$ and the output error signal $u_1(k)$. The expressions of the variable step-sizes are given by the following relations:

$$\mu_{cor1}(k) = \begin{cases} 1 - \frac{\gamma_1}{\|u_2(k)\boldsymbol{p}_1^T(k)\|} & if \quad |u_2(k)| > \gamma_1 \\ 0 & else \end{cases} \quad (9)$$

$$\mu_{cor2}(k) = \begin{cases} 1 - \frac{\gamma_2}{\|u_1(k)\boldsymbol{p}_2^T(k)\|} & if \quad |u_1(k)| > \gamma_2 \\ 0 & else \end{cases} \quad (10)$$

The updating relations are given by:

$$\boldsymbol{w}_{12}(k) = \boldsymbol{w}_{12}(k-1) + \mu_{cor1}(k)\boldsymbol{\varphi}_1(k) \quad (11)$$

$$\boldsymbol{w}_{21}(k) = \boldsymbol{w}_{21}(k-1) + \mu_{cor2}(k)\boldsymbol{\varphi}_2(k) \quad (12)$$

Where

$$\boldsymbol{\varphi}_1(k) = \frac{u_2(k)\boldsymbol{p}_1(k)}{\|\boldsymbol{p}_1(k)\|^2} \quad (13)$$

$$\boldsymbol{\varphi}_2(k) = \frac{u_1(k)\boldsymbol{p}_2(k)}{\|\boldsymbol{p}_2(k)\|^2} \quad (14)$$

1. SIMULATION RESULTS

The analysis simulation results of the proposed algorithm are presented in this section. We are interesting on the comparison between the proposed TSC-SM-NLMS algorithm and the classical TS-SM-NLMS [2] algorithm in terms of stability and convergence speed. Two types of noise are used in our simulations (i.e. White Gaussian noise WGN, and United States of America Standard Institute (USASI) noise). All the obtained simulations results are performed at a sampling frequency of $Fs = 8\,kHz$.

Table 1. Simulation parameters.

Algorithms	Simulation parameters
TS-SM-NLMS [2]	**Initial step-sizes :** $\mu_1 = \mu_2 = 0.1$ **Adaptive filter length** $\boldsymbol{w}_{12}, \boldsymbol{w}_{21}: N = 64$ **Fixed threshold error** $\gamma_1 = \gamma_2 = 15$ and 90 Input $SNR1 = SNR2 = 3$ dB
Proposed algorithm [In this paper]	**Initial step-sizes :** $\mu_1 = \mu_2 = 0.1$ **Adaptive filter length** $\boldsymbol{w}_{12}, \boldsymbol{w}_{21}: N = 64$ **Fixed threshold error** $\gamma_1 = \gamma_2 = 15$ and 90 Input $SNR1 = SNR2 = 3$ dB

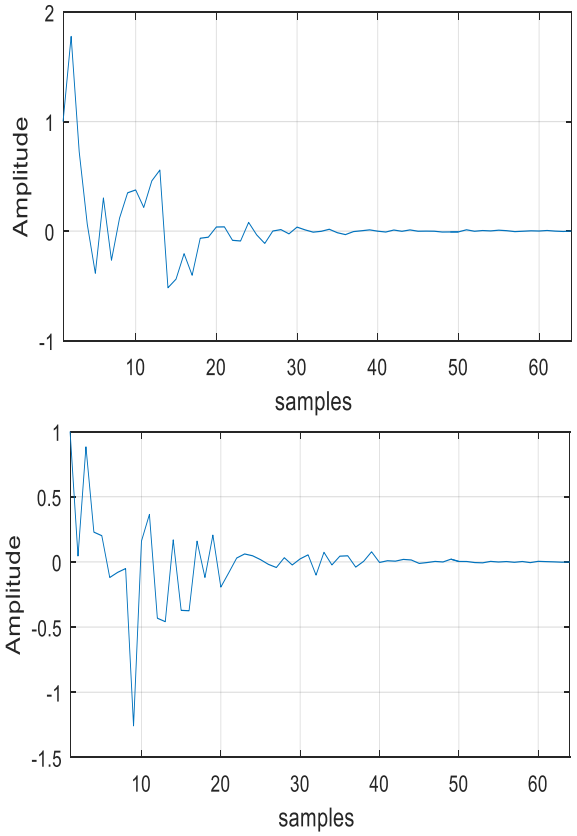


Figure 3. Examples of Impulse responses $h_{12}(k)$,$h_{21}(k)$ [from top to bottom].

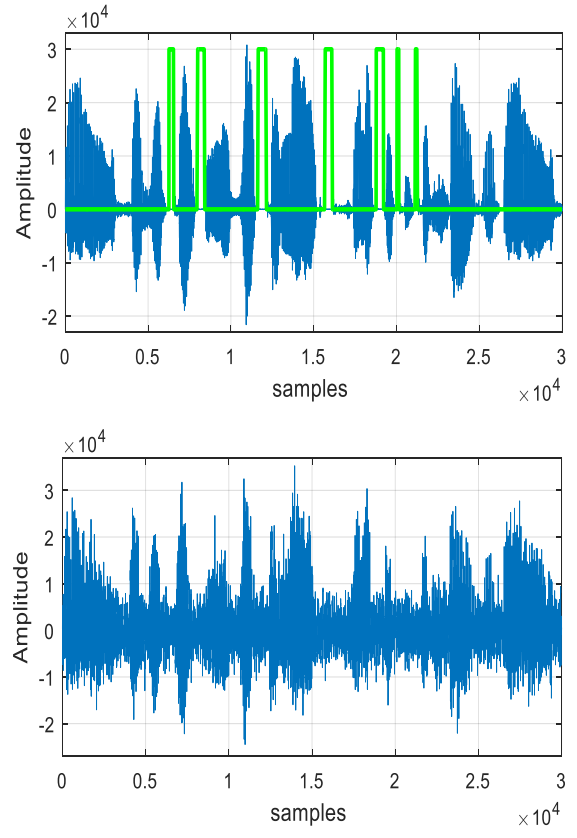


Figure 4. Original speech signal and its MVAD, and noisy signal [from top to bottom] [5].

Figure 3 illustrate two examples of the real impulse responses $h_{12}(k)$ and $h_{21}(k)$ that we had used in our simulations. The length of each impulse response is 64. The original speech signal and its manual voice activity detector (MVAD), and an example of the mixture signal $\boldsymbol{p}_1(k)$ that we have used in these simulations are presented in fig.4.

The evaluation of the system mismatch (SM) and the mean square error (MSE) criteria [6] for the both algorithms (.i.e. proposed algorithm and the classical TS-SM-NLMS) is presented in fig 5 and fig 6. All simulation parameters of each algorithms are summarized in Table 1.

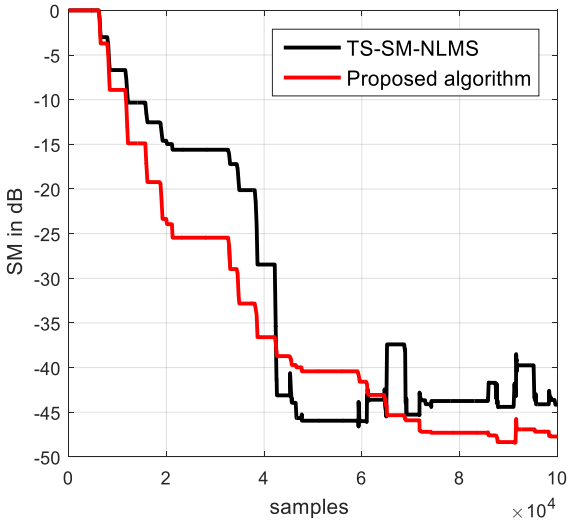


Figure 5. System mismatch evaluation of TS-SM-NLMS and Proposed algorithm, with WGN.

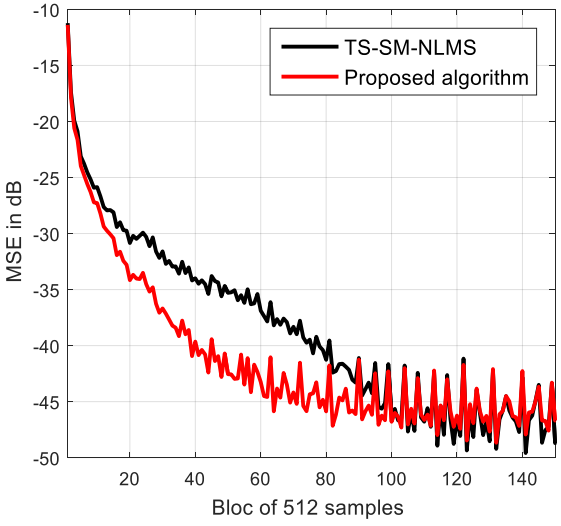


Figure 6. Mean square error evaluation of TS-SM-NLMS and Proposed algorithm, with USASI.

According to the evaluation of the two criteria (SM and MSE) presented in fig 5 and fig 6, we can clearly observe the good behavior of the proposed algorithm in comparison with the classical TS-SM-NLMS ones.

Table 2. Output SNR evaluation for different noise types.

Noise Type		WGN			USASI			Babble			Street		
Input SNR in dB		-6	0	6	-6	0	6	-6	0	6	-6	0	6
Output SNR in dB	TS-SM-NLMS [7-8]	41,91	42,41	46,39	40,55	45,23	50,81	39,89	42,43	48,77	41,41	43,55	46,11
	Proposed algorithm	48,69	49,12	50,32	46,48	48,25	53,89	44,93	47,50	52,43	47,85	49,25	52,33

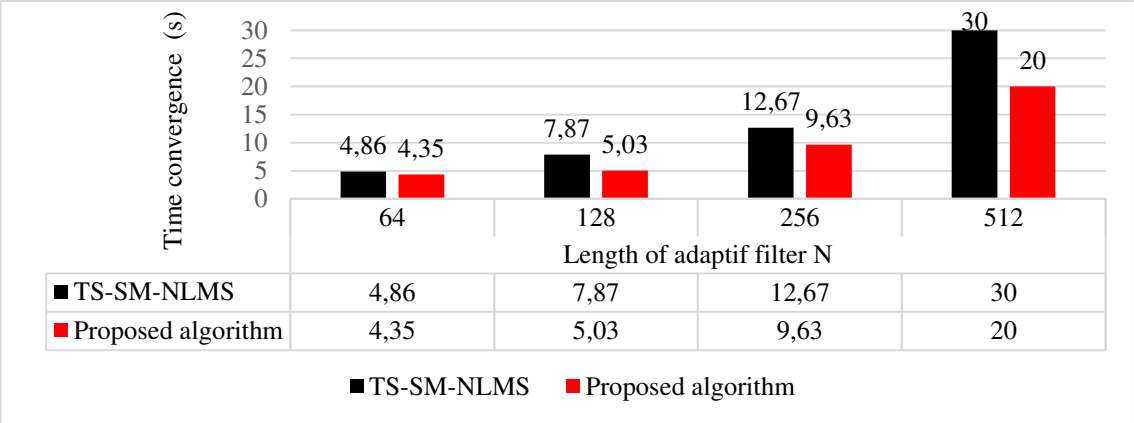


Figure 7. Time convergence for different adaptive length filter.

Table 2 illustrate the evaluation of the output SNR [6] for different type of noise and different input SNR. The time convergence in second is presented in fig 7. From these simulation results, the proposed algorithm show the best performance in term of speed convergence and quality of enhanced speech signal. This superiority is confirmed for different noisy conditions.

5. CONCLUSION

In this paper, we propose a new version of the two-sensor forward Set Membership algorithm for acoustic noise reduction and speech quality enhancement applications. We compare the classical TS-SM-NLMS algorithm with the proposed algorithm to clearly demonstrate differences in convergence speed and stability. To ensure a fair comparison between the two algorithms, three objective criteria are used: system mismatch (SM), segmental mean square error (SegMSE), and output segmental signal-to-noise ratio (Seg-SNR). The results show that the proposed algorithm performs more efficiently than the classical algorithm. All obtained results confirm the superiority of the proposed algorithm in speech enhancement and acoustic noise reduction applications.

6. REFERENCES

[1] P. S. R. Diniz, "Data-selective adaptive filtering", *Adaptive Filtering*, Springer, 2008.

[2] P. S. R. Diniz, R. P. Baraga, and S. Werner, "Set-membership affine projection algorithm for echo cancellation," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2006.

[3] Djendi M, Scalart P, Gilloire A. Analysis of two-sensors forward BSS structure with post-filters in the presence of coherent and incoherent noise. Speech Comm. 2013; 55(10):975–987.M.

[4] Djendi M, Bendoumia R. A new adaptive filtering subband algorithm for two channel acoustic noise reduction and speech enhancement. ComputElectr Eng. 2013; 39(8):2531–2550.

[5] A. Cheffi, M. Djendi, and A. Guessoum A new correlated set-membership partial-update NLMS (SM-PU-NLMSCOR) algorithm for acoustic noise reduction. In: 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B) 2017, Boumerdes, Algeria; 2017.

[6] Sayoud.A, Djendi,M Efcient subband fast adaptive algorithm based-backward blind source separation for speech intelligibility enhancement. International Journal of Speech Technology. 2020; 23:471–479.

[7] Diniz, P.S.R. Adaptive Filtering: Algorithms and Practical Implementaion, 4th ed.; Springer: New York, NY, USA, 2013.

[8] Diniz, P.S.R. Adaptive Filtering: Algorithms and Practical Implementaion, 5th ed.; Springer: New York, NY, USA, 2020.

Optimizing the Structural, Morphological, and Optical Properties of $(PbS)_{0.5}(SnS)_{0.5}$ Composite Thin Films via Spray Pyrolysis for Advanced Optoelectronic Applications

MAKHLOUF Oussama
Laboratoire d'Elaboration et de Caracterisation des Materiaux
Department Electronique, University Djilali Lyabes
BP 89 Sidi Bel Abbes 22000-Algérie
makhloufoussama72@gmail.com

Pr KHADRAOUI Mohammed
Laboratoire d'Elaboration et de Caracterisation des Materiaux
Department Electronique, University Djilali Lyabes
BP 89 Sidi Bel Abbes 22000-Algérie
khadraoui_hm@yahoo.fr

Pr Redouane MILOUA
Laboratoire d'Elaboration et de Caracterisation des Materiaux
Faculte des Sciences de la Nature et de la Vie, Universite Ibn Khaldoun
14000, Tiaret, Algeria
mr_lecm@yahoo.

Abstratc : This study investigates the impact of incorporating SnS into PbS-based composite thin films, focusing on their structural, morphological, and optical properties. Thin films were synthesized using the spray pyrolysis technique, and their characteristics were analyzed to understand how SnS influences the crystal structure, surface morphology, and optical behavior. Results reveal significant modifications in band gap energy, crystallinity, and surface features, offering valuable insights for optimizing composite thin films for energy and optoelectronic applications.

Keywords: Composite thin films, Lead sulfide (PbS), Tin sulfide (SnS), Spray pyrolysis, Optical properties*)*

I. INTRODUCTION

Thin film materials, particularly lead sulfide (PbS), are widely studied for their applications in optoelectronics and photovoltaics due to their tunable band gap and high absorption coefficient. However, their performance often faces limitations, necessitating enhancements through material modification. Incorporating tin sulfide (SnS) into PbS matrices is a promising approach to improve properties such as band gap tunability, crystallinity, and surface morphology.

Previous research has shown that Sn-doped PbS thin films exhibit enhanced microstructure and electrical properties, making them suitable for infrared photodetectors [1]. Chemically deposited PbS films further demonstrate how deposition parameters affect their structural and optical behavior [2]. Similarly, studies on doped PbS films highlight the role of secondary materials in enhancing optical and photovoltaic performance [3].

This study focuses on the effect of SnS on the structural, morphological, and optical properties of PbS-based composite thin films, using spray pyrolysis to provide insights for advanced energy and optoelectronic applications.

II. EXPÉRIMENTAL DÉTAILS

The spray pyrolysis technique is used to prepare a mixture of thin films $(SnS2)_x(PbS)_{1-x}$ on microscope glasses (75×25 mm^2), we used $Pb(NO_3)_2$ powder, which was dissolved solely in deionized water, and the tiouréa $CS(NH_2)_2$ were prepared with a solvent that contains an appropriate ratio of deionized water, First, the substrates were cleaned in a water bath, and then they were dipped successively in con HCl, acetone, and ethanol. Afterwards, the substrates were washed with deionized water and then dried using a hot air oven. The preparation process utilized compressed air as the carrier gas, which was maintained at a pressure of 2 bar. The process was conducted under the following conditions: a substrate temperature of 350°C and a solution flow rate of 8 ml/min. The distance between the spray nozzle and the heating plaque was 29 cm. It was also observed that the solutions that were prepared were sprayed immediately to prevent any potential chemical reactions over time

III. RESULTS AND DISCUTION

A. Structural studies

X-ray diffraction (XRD) analysis was conducted to examine the crystallographic structure and composition of PbS, SnS, and their composite thin films with PbS:SnS ratios of 7:3, 5:5, and 3:7 (Figure 1). The polycrystalline nature of the films was confirmed by multiple peaks corresponding to PbS (JCPDS 05-0592) and SnS (JCPDS 39-0354). PbS shows a cubic structure with prominent peaks at 26.10° (111), 30.27° (200), and 43.18° (220), while SnS has an orthorhombic structure with peaks at 21.72° (110), 27.49° (021), and 31.92° (111). Composite films displayed peaks from both phases, confirming their successful formation.The pattern for $(PbS)_{0.5}(SnS)_{0.5}$ shows peaks from both phases, including the SnS peak at 21.65° (110), indicative of an almost equal mixture. The Williamson-Hall (W-H) method analyzes strain and crystallite size from XRD data, separating peak broadening effects using the following equation:

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

$$\beta cos\theta = \frac{k\lambda}{D} + 4\varepsilon sin\theta \quad (1)$$

β is the FWHM (in radians), θ the Bragg angle, λ the X-ray wavelength, D the crystallite size, ϵ the microstrain, and k the shape factor (typically 0.9). Plotting βcosθ vs. sinθ (Figure 4) gives kλ/D (crystallite size) as the intercept and 4ϵ (strain) as the slope.

The dislocation density (δ) represents the number of dislocations per unit area and is calculated using the grain size D:

$$\delta = \frac{1}{D^2} \quad (2)$$

Where D is the average grain size in nanometers (from W-H).This formula highlights that smaller grain sizes correspond to higher dislocation densities, indicating greater structural defects

Table 1 provides the calculated parameters of the films, including lattice constants, dislocation density, grain size, and strain. The structural analysis reveals significant changes in grain size, dislocation density, and strain with varying compositions of $(PbS)_x(SnS)_{1-x}$ thin films

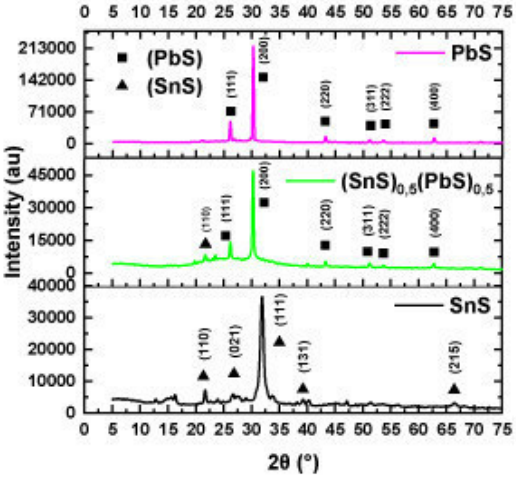


Figure.1. XRD patterns of composite thin films $(PbS)_x(SnS)_{1-x}$for different x values $0 \leq x \leq 1$ prepared by the spray pyrolysis method at 350◦C

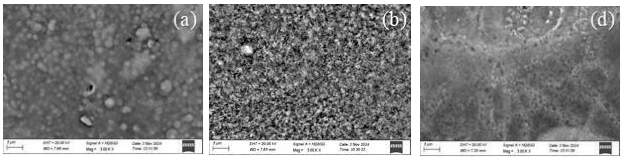
Table 1 Results of crystallite size D, microstrain, dislocation density using W-H methods of $(PbS)_x(SnS)_{1-x}$ thin film

Composition $(PbS)_X(SnS)_{1-X}$	Lattice constants (Å)	D_{W-H} (nm)	Strain W-H 10^{-3}	Dislocation Density (δ) (10^{-3} line/nm^2)
PbS	5.90	38.61	1.28	0.51
$(PbS)_{0.5}(SnS)_{0.5}$	5.89	13.50	-0.46	0.83
SnS	a 4.43 b 11.54 c 3.77	9.68	-2.79	6.57

B. Morphological analyses

The surface features of the deposited thin films, analyzed via SEM, reveal significant morphological differences across compositions. Pure PbS (Figure 2-a) shows a granular morphology with larger grains, indicating a rougher texture. In contrast, pure SnS (Figure 5-b) exhibits a fine-grained, smooth surface. Meanwhile, the $(PbS)_{0.5}(SnS)_{0.5}$ composition (Figure 5-d) shows mixed morphology with medium-sized grains and rougher regions, reflecting a balance between the two phases. Overall, higher PbS content increases surface roughness, while higher SnS content results in smoother, compact surfaces.

Figure. 2.MEB images of $(PbS)_x$ $(SnS)_{1-x}$ composite thin film: (a)PbS, (b) SnS, (d) $(PbS)_{0.5}$ $(SnS)_{0.5}$



C. Optical properties

A UV–Visible–NIR spectrophotometer was used to measure transmittance and reflectance, analyzing the optical properties of PbS, SnS, and $(PbS)_x(SnS)_{1-x}$ thin films. Figure 3 shows optical transmittance varying with wavelength, with average transmittance in the visible range from 0.56% to 10.78%. Figure 3 displays low reflectance for compositions like x = 0, 5, highlighting the films' suitability as light-absorbing materials for solar cells. The optical band gap was calculated using the Tauc relation.

$$(\alpha h\gamma)^m = An.(h\gamma - E_g) \quad (3)$$

Where An represents a constant that is not influenced by energy variations. The exponent m, on the other hand, is contingent upon the type of transition taking place. For instance, when m equals 2, it signifies a direct allowed transition, whereas when m is set to 1/2, it corresponds to an indirect allowed transition, The direct band gap energies for the thin films of SnS, PbS and $(PbS)_x(SnS)_{1-x}$ were determined by extending the linear section of the $(\alpha h\upsilon)^2$ graph to intersect the horizontal axis, hυ, as illustrated in Figure 4, The graph shows the Tauc plot used to determine the energy band gap (Eg) of PbS−SnS thin films and their mixtures. The calculated values of $(PbS)_x(SnS)_{1-x}$ optical band gaps are listed in Table, The band gap energies decreases from 1.66 eV for pure PbS to 1.16 eV for pure SnS, with intermediate value of, 1.38 eV for $(PbS)_{0.5}(SnS)_{0.5}$, This trend highlights the tunable band gap property of the PbS−SnS system, making it suitable for applications requiring tailored optical absorption, such as photovoltaics or optoelectronic devices

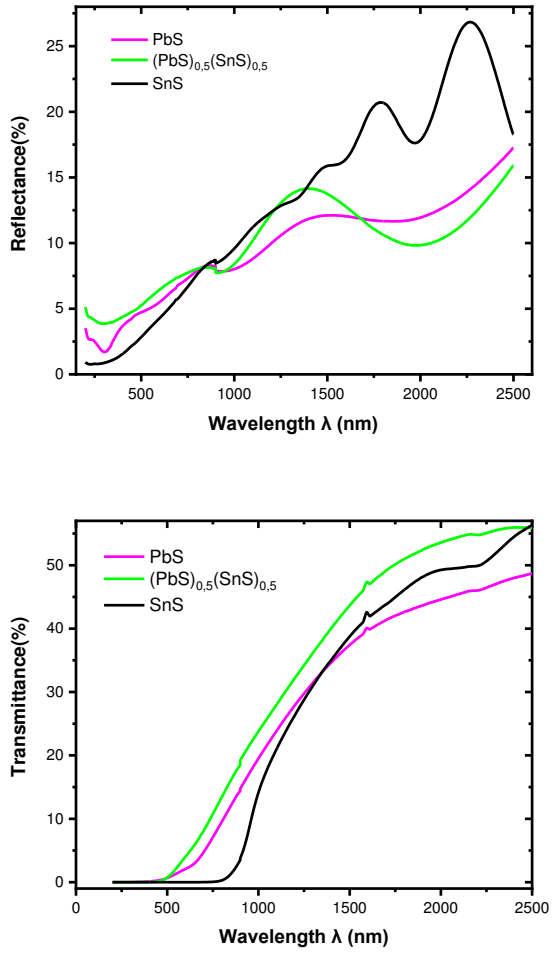


Figure.3. The transmittance and reflectance curves of $(PbS)_x(SnS)_{1-x}$ composite thin film

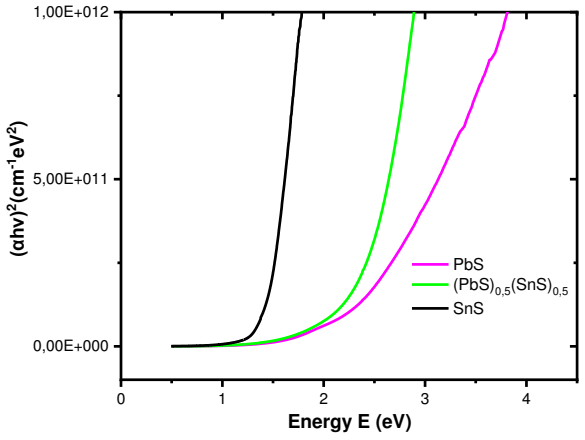


Figure.4. Variation of (αhv)2 versus energy of $(PbS)_x(SnS)_{1-x}$ composite thin films

IV . Conclusion

This study successfully demonstrated the potential of (PbS)x(SnS)1−x composite thin films, synthesized through the spray pyrolysis technique, in advancing energy and optoelectronic applications. By varying the composition ratios, significant modifications in crystallinity, grain size, and band gap energy were observed, highlighting the tunability of these materials. The incorporation of SnS into the PbS matrix effectively enhanced optical absorption properties, crucial for applications such as photovoltaics. Additionally, morphological analyses revealed a smooth-to-rough transition with varying PbS and SnS concentrations, offering further insights into the design of films tailored for specific applications. These findings underscore the versatility and potential of PbS–SnS composites in developing next-generation optoelectronic devices. Future work could explore extended applications and further optimizations in the deposition parameters to enhance film performance.

References

[1] Sn-doped PbS thin films and their application in infrared photodetectors. *Applied Physics A: Materials Science & Processing*, vol. 127, no. 6, pp. 456–464, June 2021.

[2] Structural and optical properties of chemically deposited PbS thin films. *Advances in Materials Research*, vol. 9, no. 3, pp. 310–320, September 2023.

[3] Optical, electrical, and photovoltaic properties of PbS thin films with dopants. *Journal of Materials Science*, vol. 52, no. 14, pp. 8625–8636, July 2017.

[4] P. Vidhya, K. Shanmugasundaram, P. Thirunavukkarasu, T. Govindaraj, V. Balasubramani, B. Yogeswari, and M. Karuppusamy, “Enhancement of optoelectronic properties of PbS thin films grown by Jet nebulizer spray pyrolysis technique for photodetector applications: an impact of substrate temperature,” J. Mater. Sci.: Mater. Electron., vol. 34, no. 4, pp. 2755–2767, Apr. 2023.

[5] T. Deshmukh and N. Y. Dzade, “Spray Pyrolysis: Thin Film Coating,” in *Simple Chemical Methods for Thin Film Deposition*, Springer, 2023, ch. 8, pp. 275–302.

[6] M. Mohanraj, S. Aejitha, T. Govindaraj, T. Niyitanga, H. Kim, and M. Shkir, “Enhancement of physical properties of nebulizer spray-pyrolyzed PbS thin films for optoelectronic device application: an effect of Ag doping,” J. Mater. Sci.: Mater. Electron., vol. 34, no. 2, pp. 1171–1185, Feb. 2023.

Machine learning approach for automated gender recognition

Amina Hamza
Speech and Signal Processing Lab
University of Sciences and Technology, Houari Boumediene
Algiers, Algeria
ahamza1@usthb.dz

Imene Hadjadji
Speech and Signal Processing Lab
University of Sciences and Technology, Houari Boumediene
Algiers, Algeria
ihadjadji@usthb.dz

***Abstract*— Gender recognition through speech has emerged as a pivotal application in artificial intelligence and human-computer interaction, offering solutions for personalized services, emotion recognition, and speech-driven technologies. This study examines the effectiveness of acoustic parameters, particularly the fundamental frequency, in gender classification using two machine learning approaches: Support Vector Machines (SVM) and Multi-Layer Perceptron (MLP) neural networks. By analyzing a dataset of 3,168 speech samples (1,584 male and 1,584 female), key acoustic features such as mean frequency, spectral centroid, and modulation index were extracted for training and validation. Results show that the MLP model, utilizing dense layers with Tanh activation, achieved the highest accuracy of 95.11%, surpassing SVMs employing various kernels. The study also employed K-fold cross-validation, further demonstrating the superior adaptability of the MLP model for speech-based gender recognition tasks. This research underscores the potential of deep learning architectures in enhancing gender classification accuracy and highlights critical considerations in feature selection and activation function design for speech-based machine learning applications.**

Keywords—gender recognition, acoustic parameters, SVM, MLP.

I. INTRODUCTION

Speech signals are a rich source of information, serving as a key medium for understanding various human characteristics such as gender, age, and emotions [1]. In the realm of artificial intelligence and human-computer interaction, recognizing gender through speech has become an increasingly valuable application, enabling personalized user experiences, optimizing speaker recognition systems, and enhancing the efficiency of voice-driven technologies. The fundamental frequency, as one of the most prominent acoustic features [2], plays a vital role in capturing gender-specific vocal characteristics, making it a crucial feature for accurate gender classification. Coupled with machine learning techniques such as Support Vector Machines (SVMs) [3], fundamental frequency offers a robust foundation for building reliable and efficient gender recognition systems.

Over the years, researchers have explored various approaches to improve the accuracy and robustness of gender recognition systems. Yasmin et al. (2021) [4] demonstrated that the combination of acoustic features selected using Rough Set Theory (RST) and features automatically learned by deep neural networks provides enhanced performance in gender classification, particularly when transgender categories are included alongside male and female genders. Their work highlights the significance of feature selection and hybrid systems in addressing practical challenges in speech-based gender recognition. Furthermore, La Mura and Lamberti (IEEE) [5] emphasized the importance of gender recognition for personalizing human-machine interaction systems, showcasing its application in domains such as virtual assistants, smart homes, and emotion recognition systems.

Deep learning models, particularly Convolutional Neural Networks (CNNs) with specialized modules, have shown remarkable progress in extracting salient temporal and spatial features from speech spectrograms. Tursunov et al. (2021) [6] introduced a multi-attention module within a CNN framework, enabling efficient extraction of time and frequency domain features for gender and age recognition. Their model achieved superior accuracy in both gender and age classification tasks, underscoring the potential of deep learning architectures in speech processing applications.

In addition to neural network-based methods, traditional machine learning techniques such as SVMs continue to play a pivotal role in speech-based gender recognition. Alkhawaldeh (2019) [7] explored various feature selection algorithms and demonstrated the efficacy of SVMs in achieving high classification accuracy when optimal features are employed. Similarly, Bisio et al. (2013) [8] highlighted the significance of gender information in improving emotion recognition systems, illustrating how gender recognition serves as a precursor to other speech-based applications.

Another critical aspect of speech-based recognition systems lies in the integration of utterance-level representations, as explored by Wang and Tashev (2017) [9]. Their work utilized deep neural networks to learn robust representations for age, gender, and emotion recognition, showcasing the synergy between feature extraction and classification techniques. The findings further emphasize the complementary nature of traditional and deep learning approaches in addressing the complexities of speech processing tasks.

In the other hand, other studies have explored the integration of gender recognition systems to enhance speech processing applications. Sallam (2018) [11] demonstrated that gender-based modeling significantly improves speech recognition accuracy by leveraging distinct features such as fundamental and formant frequencies. Similarly, Chachadi and Nirmala [12] proposed a 1-D CNN-based approach, achieving high accuracy in gender classification by extracting and analyzing MFCC, Mel spectrogram, and Chroma features from speech signals using the Mozilla voice dataset. These advancements

highlight the potential of gender recognition to refine the performance of speech-related technologies.
This study aims to leverage the fundamental frequency as the primary acoustic feature for gender recognition and utilize SVMs for classification. By combining the insights from previous research and focusing on a simplified yet effective feature set, this work seeks to contribute to the development of efficient and accessible gender recognition systems. The integration of methodologies from Yasmin et al., La Mura and Lamberti, Tursunov et al., Alkhawaldeh, Bisio et al., and Wang and Tashev provides a comprehensive foundation for exploring the interplay between acoustic features and machine learning techniques in achieving accurate gender classification.

II. DATASET AND SYSTEM ARCHITECTURE

A. *Dataset*

1584 male and 1584 female voice parameters make up the dataset used to train and validate the model. A single CSV file containing 20 parameters for every record in the dataset has been created. Voice samples that lasted no more than three seconds were used to extract the features.

Table I contains the features used and their descriptions:

Acoustic features	
Feature	**description**
Meanfreq	Mean Frequency: Male voices tend to have lower average frequencies compared to female voices because of differences in vocal fold length and mass [13].
Sd	Standard Deviation of Frequency: Measures variability in frequencies; higher variability may indicate speech dynamics, often found in female speech (fig. 1).
Median	Median Frequency: Like the mean, it reflects pitch characteristics. Female voices typically exhibit higher median frequencies [13].
Q25 and Q75	First Quartile and Third Quartile: Indicates the spread of frequencies in the lower and upper ranges, capturing pitch dynamics. Male voices often have a narrower spread in lower ranges [14].
IQR	Interquartile Range: An audio recording's IQR is its frequency range between Q25 and Q75, and its value is the difference between Q25 and Q75 [14].
Skew	Skewness: Indicates asymmetry in frequency distribution; skewed towards lower frequencies in male speech. [15]
Kurt	Kurtosis: High kurtosis indicates sharper peaks in frequency distribution, more common in female speech due to higher harmonics [16].
Sp.Ent	Spectral Entropy: Measures the randomness in the spectrum; lower values in male speech due to more uniform energy distribution in lower frequencies [17].
Sfm	Spectral Flatness Measure: Indicates noisiness; male speech tends to have less flatness as it is more harmonic-dominant [18].
Mode	Mode Frequency: Represents the most frequent pitch value; typically, lower in male voices.
Acoustic features	
Feature	**description**
Centroid	Spectral Centroid: Describes the center of spectral energy; higher in female voices due to greater energy in higher frequencies [18].
Meanfun, minfun, maxfun	Mean, Min, and Max Fundamental Frequencies: Directly captures pitch range; higher values are characteristic of female voices [19].
Meandom, mindom, maxdom, dfrange	Mean, Min, and Max Dominant Frequencies and Range: Dominant frequencies are associated with the strongest harmonic components; males often have stronger dominance in lower ranges [19].
Modindx	Modulation Index: Measures variations in amplitude; typically, more varied in female speech.

Figure 1 presents the first four acoustic parameters utilized in this framework, highlighting the differences between male and female voice features.

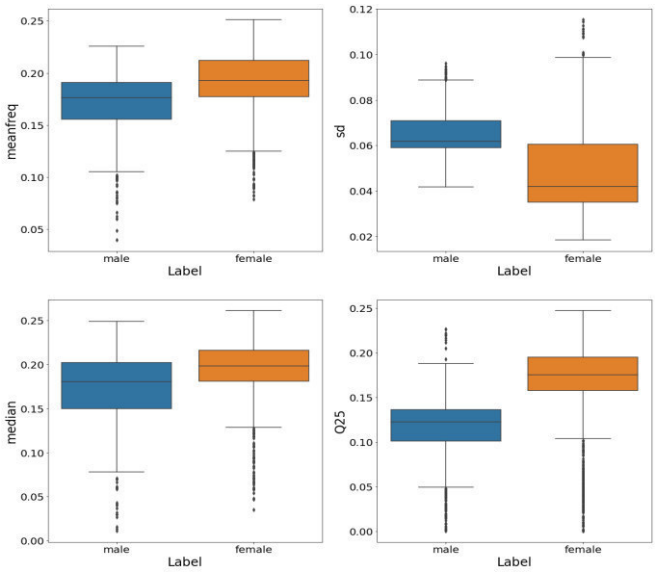


Fig. 1. Example of a figure caption

B. *Framework architecture*

The goal of this study is to determine which model—voice features, SVM, or MLP—is more effective at recognizing gender.

1) Support Vector Machine:

This architecture aims to effectively separate the examples of two classes by optimizing the margin or distance between them and the hyperplane. This idea, sometimes referred to as the "maximum margin classifier," guarantees that the hyperplane keeps the classes as far apart as possible [20]. We evaluated the SVM classifier's performance in our experiments using radial basis function (RBF), linear, and polynomial kernels. To identify the best kernel, evaluation metrics like accuracy, precision, recall, and F1 score are applied to test or validation datasets. The complexity of the problem and the properties of the data determine which kernel is best. Finding the kernel that best fits the specified classification task can be accomplished by testing out different kernels and evaluating how well they perform.

2) Artificial Neural Nertwork:

Multi Layer Perceptron (MLP) is a computational model inspired by the structure and functioning of biological neural networks, such as the human brain. It is composed of interconnected layers of nodes, called neurons or units, which process information in a manner similar to biological neurons. MLPs are used to learn patterns and relationships in data, making them a fundamental tool in machine learning and artificial intelligence.

In this framework, The MLP model is implemented in Keras with multiple dense layers, including intermediate layers of 128, 64, 32, 16, and 8 neurons, all using the ReLU then tanh activation function. The output layer consists of a single neuron with a sigmoid activation function, suitable for binary classification tasks. The model uses the binary_crossentropy loss function for optimization, the Adam optimizer with a learning rate of 0.0001, and accuracy as the evaluation metric to assess performance.

III. RESULTS

A. *Simple validation*

The database was split into 70% for training and 30% for testing. The dataset is already labeled as "male" and "female" to perform binary classification (0 or 1). For the SVM method, tests were conducted using three kernels: linear, polynomial, and Gaussian (RBF). The neural network was trained for 250 epochs. The results are summarized in the following table:

TABLE I. GENDER RECOGNITION BY SVM AND MLP

Model		Accuracy (%)
SVM	Linear kernel	91.69
	Polynomial kernel	52.79
	RBF kernel	67.30
MLP		95.11

The results indicate that the neural network (MLP) outperformed the SVM models, achieving the highest accuracy of 95.11%. Among the SVM kernels, the linear kernel provided the best performance with an accuracy of 91.69%, while the polynomial kernel performed the worst at 52.79%, showing its unsuitability for this task. The RBF kernel achieved moderate accuracy at 67.30%, indicating that it struggled to capture the patterns in the dataset as effectively as the linear kernel or the MLP. Overall, the MLP demonstrated superior performance, likely due to its ability to model complex non-linear relationships in the data during its extended training (250 epochs).

B. *Cross validation*

Using K-fold cross-validation, we divide the sample set into K subsets and run a number of tests. While the remaining subsets are used for training, each subset serves as a validation set. We selected K values of 5, 10, and 20 for our experiments. Ten iterations for each fold were used in the evaluation of the SVM and MLP models. Table II provides a summary of the findings.

TABLE II. GENDER RECOGNITION BY CROSS VALIDATION (SVM AND MLP)

Model		**Classification rate (%)**		
		K=5	**K=10**	**K=20**
MLP	**Relu**	73.03	77.29	80.38
	Tanh	78.07	77.92	85.44
SVM (linear)		92.01	92.52	92.71

- The model using tanh activation performs better overall (78.07% to 85.44%) than that using ReLU (73.03% to 80.38%).
- Tanh generates outputs between -1 and 1, unlike ReLU, which produces values between 0 and ∞ (fig. 2).
- For normalized data or data centered around 0, tanh allows for better symmetry in gradient propagation, which can help the network converge faster and more efficiently.
- ReLU tends to produce zero outputs for negative inputs (“dead neuron” effect), which can reduce learning in certain layers if some of the neurons never activate. This is particularly problematic if the data do not cover the entire input range or if the initial weights are not optimal. Tanh, on the other hand, is always active for all input values (positive or negative), which favors better gradient propagation.
- Returning to cross validation, with K=20, performance increases for both models, probably due to better exploitation of larger data sets or more representative subgroups.
- SVM with a linear hyperplane performs better, suggesting that the acoustic data are linearly separable, or at least that a linear model is sufficient.

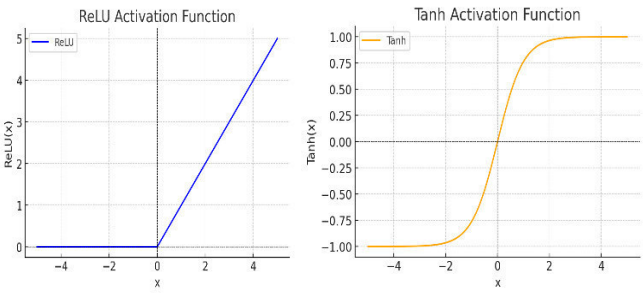


Fig. 2. Relu and tanh activation functions.

IV. CONCLUSION

This study demonstrates the importance of leveraging acoustic features and advanced machine learning models for accurate gender recognition from speech signals. The analysis reveals that the Multi-Layer Perceptron (MLP) model outperforms Support Vector Machines (SVM), with an impressive accuracy of 95.11%, owing to its ability to capture complex, non-linear relationships in the data. Additionally, the use of the Tanh activation function in MLP offers superior gradient propagation and convergence compared to ReLU, particularly in normalized datasets. Cross-validation

experiments further validate the robustness of the models, highlighting the significance of dataset size and representative subsets for improved classification performance. These findings not only reinforce the utility of machine learning in speech-based applications but also provide a foundation for future advancements in human-computer interaction and personalized voice-driven systems. Future work can explore the integration of hybrid models and larger datasets to further refine the efficiency and scalability of gender recognition systems.

REFERENCES

[1] Md. Sadek Ali, Md. Shariful Islam, and Md. Alamgir Hossain, "Gender Recognition System Using Speech Signal," International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.1, February 2012.

[2] Anjali Pahwa and Gaurav Aggarwal, "Speech Feature Extraction for Gender Recognition," J. Image, Graphics and Signal Processing, Vol. 9, September 2016, pp. 17–25. DOI: 10.5815/ijigsp.2016.09.03.

[3] Bhagyalaxmi Jena, Anita Mohanty, and Subrat Kumar Mohanty, "Gender Recognition of Speech Signal Using KNN and SVM," International Conference on IoT based Control Networks and Intelligent Systems (ICICNIS 2020).

[4] Yasmin, G., Das, A. K., Nayak, J., Vimal, S., & Dutta, S. (n.d.). A rough set theory and deep learning-based predictive system for gender recognition using audio speech. St. Thomas College of Engineering and Technology, Kolkata, India; Indian Institute of Engineering Science and Technology, Shibpur, Howrah, India; Aditya Institute of Technology and Management, Tekkali, India; Ramco Institute of Technology, Rajapalayam, India; Institute of Engineering and Management, Saltlake, Kolkata, India.

[5] La Mura, M., & Lamberti, P. (n.d.). Human-Machine Interaction Personalization: A Review on Gender and Emotion Recognition Through Speech Analysis. *IEEE*.

[6] Tursunov, A., Mustaqeem, Choeh, J. Y., & Kwon, S. (2021). Age and gender recognition using a convolutional neural network with a specially designed multi-attention module through speech spectrograms. *Sensors*, *21*(17), 5892. https://doi.org/10.3390/s21175892

[7] Alkhawaldeh, R. S. (2019). DGR: Gender recognition of human speech using one-dimensional conventional neural network. *Scientific Programming*, *2019*, Article ID 7213717. https://doi.org/10.1155/2019/7213717

[8] Bisio, I., Delfino, A., Lavagetto, F., Marchese, M., & Sciarrone, A. (2013). Gender-driven emotion recognition through speech signals for ambient intelligence applications. *IEEE Transactions on Emerging Topics in Computing*, *1*(2), 244–257. https://doi.org/10.1109/TETC.2013.2274797

[9] Héctor A. Sánchez-Hevia, Roberto Gil-Pita, Manuel Utrilla-Manso, and Manuel Rosa-Zurera, "Age Group Classification and Gender Recognition from Speech with Temporal Convolutional Neural Networks," Artificial Intelligence for Physical Agents, Vol. 81, 2022, pp. 3535–3552.

[10] Wang, Z.-Q., & Tashev, I. (2017). Learning utterance-level representations for speech emotion and age/gender recognition using deep neural networks. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5150–5154). https://doi.org/10.1109/ICASSP.2017.7953138

[11] Amer Sallam, "Effect of Gender on Improving Speech Recognition System," International Journal of Computer Applications (0975 – 8887), Vol. 179, No. 14, January 2018.

[12] Kavita Chachadi and S. R. Nirmala, "Gender Recognition from Speech Signal Using 1-D CNN," Speech Recognition Applications and Research, published with experiments on Mozilla voice dataset.

[13] Jena, B., Mohanty, A., & Mohanty, S. K. (2020). Gender recognition of speech signal using knn and svm.

[14] Nair, R. R., & Vijayan, B. (2019). Voice based gender recognition. *International Research Journal of Engineering and Technology*, *6*.

[15] Albiero, V., Ks, K., Vangara, K., Zhang, K., King, M. C., & Bowyer, K. W. (2020). Analysis of gender inequality in face recognition accuracy. In *Proceedings of the ieee/cvf winter conference on applications of computer vision workshops* (pp. 81-89).

[16] Hammami, I., Salhi, L., & Labidi, S. (2020). Voice pathologies classification and detection using EMD-DWT analysis based on higher order statistic features. *Irbm*, *41*(3), 161-171.

[17] Hu, J. (2018). An approach to EEG-based gender recognition using entropy measurement methods. *Knowledge-Based Systems*, *140*, 134-141.

[18] Qadri, S. A. A., Gunawan, T. S., Wani, T., Alghifari, M. F., Mansor, H., & Kartiwi, M. (2019, August). Comparative Analysis of Gender Identification using Speech Analysis and Higher Order Statistics. In *2019 IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)* (pp. 1-6). IEEE.

[19] Yasmin, G., Das, A. K., Nayak, J., Vimal, S., & Dutta, S. (2022). A rough set theory and deep learning-based predictive system for gender recognition using audio speech. *Soft Computing*, 1-24.

[20] A. Ganapathiraju, JE Hamaker and J. Picone, Applications of support vector machines to speech recognition, Signal Processing, IEEE Transactions on, 52(8), pp.2358-2355, 2004

A Review of Automated Voice Pathology Detection: Databases, Acoustic Features, and Classification Algorithms

Amina Hamza
Speech and Signal Processing Lab
University of Sciences and Technology, Houari BoumedieneB
Algiers, Algeria
ahamza1@usthb.dz

Djamel Addou
Speech and Signal Processing Lab
University of Sciences and Technology, Houari Boumediene
Algiers, Algeria
daddou@usthb.dz

Imene Hadjadji
Speech and Signal Processing Lab
University of Sciences and Technology, Houari Boumediene
Algiers, Algeria
ihadjadji@usthb.dz

***Abstract*—Speech, a cornerstone of human interaction, can be compromised by voice disorders. Early detection of these disorders is crucial to mitigate emotional and psychological impact. Computer-aided diagnosis offers a cost-effective and efficient solution over traditional methods. This review delves into recent advancements in voice pathology detection, focusing on machine learning and deep learning techniques. We examine essential aspects, including voice pathology taxonomy, feature extraction methods, and the utilization of standard databases like the Arabic Voice Pathology Database, Saarbruecken Voice Database, Nemours, UASpeech, and Massachusetts Eye and Ear Infirmary. A key emphasis is placed on the extraction of both static and dynamic features from speech signals. These features, once extracted, are transformed into suitable representations for input into machine learning and deep learning models. By providing a comprehensive overview of the field, this review aims to contribute to the development of accurate and reliable automated voice pathology detection systems.**

Keywords—pathological voice, speech disorder, feature extraction, speech classifier, deep learning, machine learning.

I. INTRODUCTION

Humans use a variety of organs, including the brain, mouth, nose, lungs, and abdomen, to produce speech. Speech production heavily relies on the vocal cords and vocal tract. Starting at a frequency of 100 Hz, speech production can reach up to 17 kHz [1]. Any abnormality that differs from the characteristics of voices of the same age group, sex, and social group, such as "loudness, quality, pitch, and/or vocal flexibility," is referred to as a voice disorder [2,3]. According to a recently released National Center for Education Statistics survey, 20% of young adults and children ages 3 to 21 had voice issues. A flaw in the human voice generation system is the cause of voice impairment, according to research by the American Speech Language-Hearing Association [4].

Vocal abnormalities are often diagnosed with invasive surgical procedures. During endoscopic procedures like laryngoscopy, laryngeal electromyography, and stroboscopy, doctors insert a probe into the patient's mouth. Patients may experience trauma from these painful procedures. Much research has been done to find alternatives to these surgical treatments. Using voice signal processing to identify vocal pathology is one of these options [5]. To detect vocal pathology using voice signals, voice features must be extracted and analyzed. Samples of voices were recorded in a particular setting. After being extracted from the voice signals, the voice features were analyzed. After that, voice samples were separated into two categories: normal and abnormal. Voice signal identification must be taken into consideration when selecting the appropriate technology. The most popular method for recognizing and differentiating between two or more subjects is the classification algorithm [6]. The application of machine learning has demonstrated efficacy in a range of medical domains, including cancer diagnosis and classification [7], three-dimensional brain reconstruction [8, 9], and the analysis of Alzheimer's disease [10-12].

Voice features must be extracted and analyzed in order to detect voice pathology using voice signals. In a controlled setting, voice samples are gathered. Voice features are then extracted by analyzing the voice signals. Classifying voice samples into two groups—normal and pathological—is the next stage.

The right one must be carefully chosen while categorizing the voice signals. The classification algorithm is the most widely used technique. The literature has a variety of classification algorithms. The published works demonstrate that the classifier has a significant impact on the accuracy level. However, voice signal-based pathology detection systems have a lot of problems and difficulties. Choosing the right classifier tool and (a) the right voice features are two crucial considerations.

This is how the remainder of the paper is structured. A voice pathology example is given in Section 2. The vocal pathology database is explained in Section 3. Methods for obtaining vocal pathology features are covered in Section 4. Algorithms for classifying vocal pathology are presented in Section 5. This paper is concluded in Section 6.

II. VOICE PATHOLOGIES

Over the past 150 years, the term "apraxia of speech" [13] has undergone various redefinitions, yet individuals diagnosed with this condition have consistently been seen as experiencing disruptions at a higher level of motor speech organization. Clinicians distinguish these patients from those with dysarthria due to the irregular and unpredictable nature of their speech patterns, which deviate from the more consistent characteristics of motor impairments caused by paresis, ataxia, akinesia, or dyskinesia in the speech muscles (e.g., Wertz et al., 1984). Theoretically, the plausibility of a motor planning stage, along with its potential impairment, has long intrigued researchers. This notion arises from the

apparent divide between a symbolic phonological grammar and the mechanical processes of motor execution, which necessitate an efficient intermediary translation system.

In the classical dualist framework, where phonemes are seen as purely symbolic entities devoid of motor information, the existence of a "translator" becomes critical for enabling the nonlinguistic motor system to interpret phoneme strings for speech production. This idea aligns with longstanding theories in motor cognition suggesting that abstract movement representations must undergo transformation into actionable motor commands—a process that can be selectively impaired, as illustrated in motor domains outside speech (Liepmann, 1900a). This interplay between abstract motor representations and execution remains a central challenge in understanding human motor control and cognition (Prinz, Aschersleben, & Koch, 2009).

Over time, the theoretical underpinnings of the mechanisms thought to cause apraxia of speech have evolved, reflecting the psychological and physiological frameworks of different periods. This paper focuses on three dominant themes that have shaped our understanding of this condition across historical and intellectual contexts [14].

III. VOICE PATHOLOGY DATASET

One essential part of the automatic voice disorder detection (AVDD) system is the voice pathology database. The dataset contains samples of both healthy and disordered voices. These examples could be continuous speech or prolonged vowel phonation. Most studies make use of a number of standard databases, including Nemours, Torgo, UASpeech, MEEI, the SVD, and the AVPD.

A. Nemours database

The Nemours database contains 814 brief nonsense sentences, 74 of which are uttered by 11 male speakers with various levels of dysarthria [15]. The database also includes two connected-speech paragraphs spoken by each of the 11 participants. It was specifically created to evaluate the intelligibility of dysarthric speech both before and after applying various signal processing techniques and is accessible on CD-ROM. Additionally, it can be utilized to examine general features of dysarthric speech, such as patterns of production errors. The entire dataset has been annotated at the word level, with phoneme-level markings provided for sentences from 10 of the 11 speakers. Discrete Hidden Markov Model (DHMM) labeler was used to assign initial phoneme labels to the elements of the database. These techniques may be useful in the design of automatic recognition systems for persons with speech disorders, especially when limited amounts of training data are available. Each nonsense sentence in the database is of the form “The X is Ying the Z” Specific sentences were generated by randomly selecting X and 2 (X # Z) without replacement from a set of 74 monosyllabic nouns and selecting Y without replacement from a set of 37 disyllabic verbs. This process produced 37 sentences from which An additional 37 sentences were created by swapping the X and Z tokens in the original set, resulting in a total of 74 sentences. This setup ensured that each noun and verb was produced twice by every speaker. The target words (X, Y, and Z) were selected based on constraints similar to those outlined by Kent et al. (1990) [16], aimed at providing closed-set phonetic contrasts (e.g., contrasts in place, manner, and voicing) within an associated group of four to six words. All target words within a set differ by a single phoneme, making them suitable for use as alternatives in a closed-response perceptual intelligibility test. Additionally, each speaker recorded two connected-speech passages: the "Grandfather" passage and the "Rainbow" passage, both of which are widely used in speech research. As a control, one non-dysarthric speaker recorded the entire speech corpus.

B. Torgo database

It is a database of measured 2D and 3D articulatory features and aligned acoustics of English speech from speakers who have either cerebral palsy (CP) or amyotrophic lateral sclerosis (ALS), two of the most common causes of speech impairment [17]. The University of Toronto's departments of computer science and speech-language pathology, along with the Toronto-based Holland-Bloorview Kids Rehab hospital, collaborated to create this database. Disruptions in the neuro-motor interface are the cause of a group of speech impairments known as dysarthria, which includes both ALS and cerebral palsy. In most situations, these disturbances lead to atypical and largely incomprehensible speech because they skew motor commands to the speech articulators rather than impairing comprehension or cognitive aspects of natural language production. The use of conventional automatic speech recognition (ASR) software is severely hampered by this incomprehensibility, to the extent that severely dysarthric subjects may have a word-error rate of 97.5% on contemporary systems compared to 15.5% for the general population. Since other types of computer input, like keyboards or touch screens, can be particularly challenging due to the more general physical disabilities frequently associated with the causative neuro-motor disruptions, the incapacity of modern ASR to comprehend dysarthric speech is a significant issue. It is crucial to measure the articulation of speakers with dysarthria empirically because their articulation differs from that of the general population. Although the TORGO database is equally valuable to the broader ASR community, its primary purpose is to support the development of ASR models better tailored to the requirements of individuals with atypical speech production. The ability to explicitly learn "hidden" articulatory parameters automatically in computer speech models through statistical pattern recognition is a major motivation for gathering comprehensive physiological data.

C. UAspeech database

19 speakers with cerebral palsy have contributed to this database of dysarthric speech [18]. Each speaker's speech materials include 765 isolated words: 300 unique uncommon words and three repetitions of common words, computer commands, radio alphabets, and digits. One digital video camera and an array of eight microphones are used to record data. For those with neuromotor disabilities, our database offers a foundational resource for the development of automatic speech recognition. In addition to improving clinical treatments, research on articulation errors in dysarthria will advance our understanding of the neuromotor mechanisms underlying speech production. On request, data files can be accessed via secure FTP. This database includes 35 words from the grandfather passage and 30 repetitions of 46 isolated words (10 digits, 26 alphabet letters, and 10 "control" words) generated by each of six people with cerebral

palsy. This database contains 15 repetitions of the same content produced by a typical speaker. A limited set of words produced by a limited number of subjects comprise the Whitaker Database and the materials utilized in the majority of previous studies on ASR for dysarthria.

D. Saarbruecken voice database

The SVD is a publicly available database [19] maintained by the Institute of Phonetics at Saarland University. It includes sustained vowels (/a/, /i/, and /u/) recorded with various intonations, such as low-high-low, low, normal, and high. Additionally, it features a spoken German sentence, "Guten Morgen, wie geht es Ihnen?", which translates to "Good morning, how are you?" These unique attributes make the database particularly noteworthy. All voice recordings in the SVD were sampled at 50 kHz with 16-bit resolution. Being relatively new, this database has only been utilized in a limited number of studies for voice disorder detection [20].

The database provides both voice and electroglottographic (EGG) signals for each recording. These signals can be exported in their native formats (NSP and EGG) or as WAV files. Once selected, the server creates a ZIP file containing the requested data, though this process may take a few minutes. After completion, a link is provided to download the ZIP file [19].

E. Massachusetts eye and ear infirmary

The MEEI lab developed a voice and speech database. It includes the first half of the Rainbow Passage and about 1,400 voiced, sustained vowel/a/samples. This database was collected in two distinct environments and is sold by Kay Elemetrics [20,21]. Pathological samples were tested at 25 kHz or 50 kHz, while normal samples were tested at 50 kHz. The database also includes a spreadsheet containing the individuals' personal and clinical data as well as the results of the acoustic analysis of the recordings taken from Kay's MultiDimensional Voice Program (MDVP). Similar acoustic conditions were also used to record in Kay's Computerized Speech Lab (CSL). For at least three seconds, each participant was required to produce a continuous phonation of the vowel /ah/ at a comfortable pitch and loudness. Each participant underwent the procedure three times, and a speech pathologist selected the best sample for the database [22]. This database has a number of shortcomings despite its widespread use. The majority of vocal pathology identification and classification studies use the database, which captures both healthy and disordered voices using various voice sample settings and frequencies (Kay Elemetrics confirmed this data).

F. Arabic voice pathology database

Samples for this database were recorded by qualified phoneticians using a specified recording methodology over the course of multiple sessions in a sound-treated environment at the Communication and Swallowing Disorders Unit (King Abdul Aziz University Hospital, Riyadh, Saudi Arabia). [20, 23]. One of the main duties of a two-year project funded by Saudi Arabia's National Plans for Science and Technology (NPST) was compiling the database. The database protocol was designed to circumvent the shortcomings of the MEEI database [22]. The database includes a variety of samples, such as recordings of people with normal speech, sustained vowels, and speeches from people with vocal fold diseases. After a clinical examination using laryngeal stroboscopy, the vocal folds that were healthy and those that were not were identified. The frequency of voice issues in pathology cases was rated on a severity scale of 1 to 3, where 3 denoted the most severe case. Each sample was assigned a severity rating by the panel of three qualified medical professionals. The audio recording contains three different types of text: (a) continuous speech, (b) three sustained vowels with onset and offset information, and (c) Arabic words, numbers, and isolated words [20].

Every Arabic phoneme was included in the carefully selected text. To avoid burdening patients, all speakers recorded three utterances of each vowel (/a/, u/, and i/); however, isolated words and continuous speech were only recorded once. The CSL application was used to record the speech, and the database's sampling frequency was 44 kHz. The vocal issues listed in this database were examined and verified by various knowledgeable specialists from King Abdul Aziz University Hospital [20]. 366 samples of both healthy and ill patients are included in the AVPD. Fifty-one percent of the subjects are normal, and the remaining subjects are split between the five voice abnormalities (sulcus, nodules, cysts, paralysis, and polyps); sixty percent of the subjects are male, and forty percent are female. All texts, including three repeating vowels—Arabic numbers, Al-Fateha, and common terms—were recorded and stored in two distinct audio formats, WAV and NSP. The recorded samples were divided into 22 segments: 3 segments for the common word, 6 segments for vowels (three vowels plus their repetition), 11 segments for Arabic numerals (zero to ten), and 2 segments for Al-Fateha (split so that the first half can be used to train the system and the second part can be used to test the system).

IV. VOICE FEATURES USED IN VOICE PATHOLOGIES

The process of calculating a series of feature vectors that yield a condensed representation of the input speech signal is known as feature extraction. Generally speaking, this approach consists of three parts. The first step, known as the acoustic front-end or speech analysis, creates raw characteristics that reflect the power spectrum envelope of short speech intervals by temporally analyzing the speech signal spectra. The creation of a feature vector with both static and dynamic data is the second step. Before delivering the extended feature vectors to the recognizer, the last step compresses and fortifies them [24].

A. Mel frequency cepstral coefficients (MFCCS)

The most widely used feature for identifying voice disorders is Mel frequency cepstral coefficients (MFCCs). It is commonly accepted to define the human voice generation system using MFCCs. A helpful technique for identifying vocal impairment is MFCC [6].

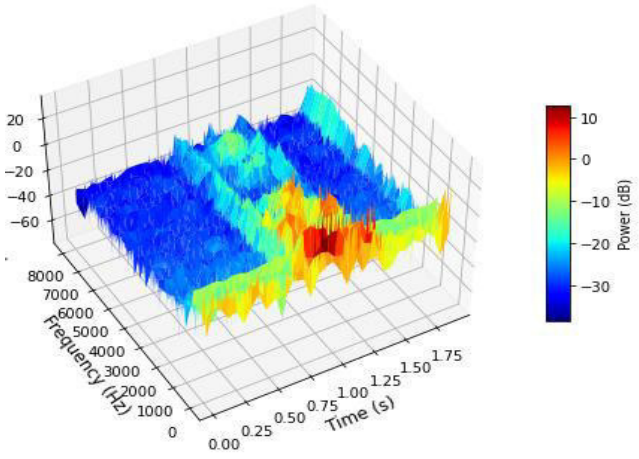
One of the most popular feature extraction techniques in speech recognition is MFCC. Features of the frequency domain, including MFCCs, are more accurate than those of the time domain because MFCC is dependent on the Mel scale, which is based on the human ear scale [25]. The true cepstral of a windowed short-time signal (FFT) is represented by MFCC, which is generated from the fast Fourier transform signal. To differentiate MFCC from the real cepstral, the auditory system activity is simulated using a nonlinear frequency scale. Furthermore, taking into account variations in speakers and recording conditions, these coefficients are consistent and trustworthy. Concurrently, MFCC is a method for feature extraction that downplays all other information

while identifying parameters from speech samples that are comparable to those humans use to hear speech [29]. Since MFCC can fully describe the geometry of the vocal tract, it has an advantage over other voice characteristics for a variety of speech samples. Once the vocal tract has been precisely defined, a precise representation of the phoneme produced by the vocal tract can be computed. The vocal tract's curvature is represented by the envelope of a short-time power spectrum, which MFCCs accurately capture [6].

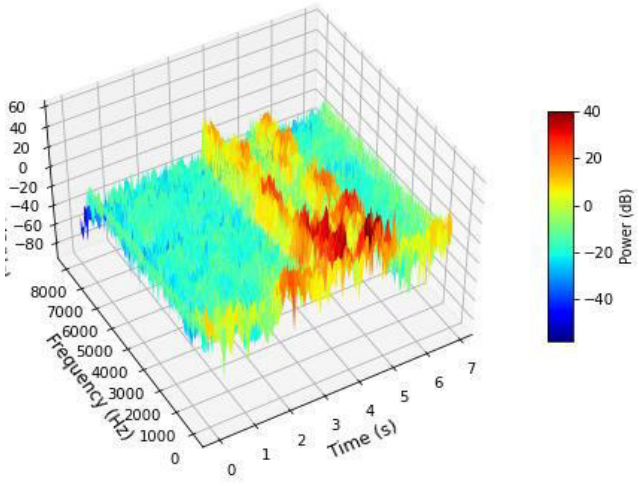
B. *Spectrograms*

A speech waveform is a series of distinct occurrences that change over time. Highly fluctuating spectral characteristics over time are correlated with this time-varying nature. As a result, STFT is utilized since a single Fourier transform is unable to capture this kind of fast time varying signal [26]. For portions of the waveform under a sliding window, the STFT consists of a distinct Fourier transform. Next, the voice signal's spectrogram is obtained from STFT by:

$$S(w) = |X(m,w_k)|^2 \quad (1)$$



(a) Spectrogram of normal voice sample.



(b) Spectrogram of pathological voice sample.

Fig .1. The voice spectrograms of normal and pathological voices for the same word.

As seen in Fig. 1, the spectrogram can be displayed as a 3-D plot that illustrates the power density distribution over time and frequency. The figure illustrates how the voice signal's power density distribution varies significantly over time and frequency, and it can be used to differentiate between pathological and normal voices. The figure also shows that the power distribution for a normal voice is consistent across time and frequency. This isn't always the case with pathological voice, though. As a result, the spectrogram is thought to be a useful tool for differentiating between pathological and normal voices.

C. *Formants*

Another significant voice characteristic that the researchers looked into was the formant frequency, or simply formant analysis. The vocal tract's resonance frequencies are known as the "formant frequencies," and they vary depending on the vocal tract configuration. The full spectral contribution of a resonance is typically referred to as the formant. The vocal tract response's formants roughly correspond to the peaks of its spectrum.

The vocal tract's resonant frequencies during speech production are represented by formants. Due to modifications in vocal tract control, they can be changed in dysarthric speech and are crucial for vowel perception. Formant frequency analysis can be used to evaluate the vowel quality and articulatory accuracy of dysarthric speech.

D. *Linear predictive coding (LPC)*

LPC was primarily developed to compress digital signals for more effective storage and transmission. But as a formant estimator, LPC has grown in popularity and is now among the most potent speech analysis methods [27]. Modeling the vocal tract as a linear all-pole infinite impulse response (IIR) filter is the foundation of the LPC technique, which is defined by:

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_p(k)\, z^{-k}}, \quad (2)$$

Where p represents the number of poles, G denotes the filter gain, and $a_p(k)$ signifies the coefficients. A speech encoder determines the appropriate excitation function, pitch period for voiced speech, gain parameter G, and the coefficients $a_p(k)$ for a short-time segment of a speech signal (i.e., *20 ms*) sampled at an 8 kHz sampling rate. The LPC is calculated using the least mean squares error approach.

E. *Perceptual linear prediction (PLP)*

PLP, which was first presented by Hermansky [28], uses the idea of the psychophysics of hearing to model human speech. PLP's primary purpose is to eliminate any unnecessary information from the speech. Unlike LPC, PLP's spectral characteristics are altered to conform to the human auditory system. As a result, PLP is better suited to human hearing than LPC. The use of two distinct kinds of transfer functions is the other primary distinction between PLP and LPC. The LPC model, for instance, makes the assumption that the vocal tract has an all-pole transfer function with a certain number of resonances within the analysis band. Conversely, the PLP transfer function is also an all-pole model, but it approximates an equal-magnitude power distribution across the analysis band's frequencies.

F. Jitter and schimmer

The voice signal's fluctuation between successive periods is reflected in jitter. The timing of the fundamental period must be detected in order to determine jitter. The average absolute difference (over N periods) between two consecutive periods is represented by jitter (local, absolute). Another common voice characteristic for detecting voice impairments is shimmer [29]. Shimmer, as opposed to Jitter, concentrates on a signal's peak values. The algorithm first determines the onset time of a signal's glottal pulses and the corresponding signal magnitude at that sample in order to calculate the Shimmer parameters. The values of each Shimmer parameter are then found using the same algorithm as for Jitter.

G. HNR

By measuring the ratio between the periodic (harmonic part) and aperiodic (noise) components, the harmonic to noise ratio gives an indication of the voice signal's overall periodicity. Rather than being a function of frequency, this parameter is typically measured as an overall signal characteristic. Because different vocal tract configurations involve different amplitudes for the harmonics, the signal's overall HNR value varies [30].

V. VOICE PATHOLOGY CLASSIFICATION ALGORITHMES

Voice signal analysis is crucial for diagnosing vocal problems by classifying a signal into predefined categories. Classifiers are used to detect vocal impairment [6], which involves three steps: feature extraction, feature selection, and feature classification. Audio samples are selected from a voice database, and required characteristics are extracted. These characteristics are optimized using an algorithm and sent to the classifier. The classifier divides the speech samples into normal and abnormal categories. Common classifiers include k-NN, convolutional neural network (CNN) and support vector machine (SVM), which are used to detect voice dysfluencies. Neural networks and the hidden Markov model (HMM) are popular nonlinear classifiers. SVMs are used because they construct an ideal wall in the feature space, dividing the two classes. SVMs maximize the difference in points between the two classes.

A few studies on pathological voice classifiers are compiled in Table I.

TABLE I. EXEMPLES OF PATHOLOCAL VOICE CLASSIFIER

Paper	Extracted features	Classifier /database	Remark	Best accuracy
[31] (2019)	MFCC	RNN, GRU, LSTM/ SVD	•Considering lengthy sequences, convolutional neural networks can offer increased accuracy. •The recurrent network is also impacted by the vanishing gradient issue.	75.64%
[32] (2019)	Spectrogram segments	CNN/ MEEI	•The CNN-GA is superior to the CNN-SA, CNN-PSO, and CNNBA algorithms when it comes to classification accuracy, sensitivity, and specificity.	99.37%
[33] (2021)	MFCC, LPCC,	FNN, CNN /SVD	•The CNN classifier yielded a noteworthy result for mixed samples. •For the /u/ vowel in men, CNN and LPCC achieved the highest accuracy of 82.69%.	82.69%
[34] (2021)	MFCC, ZCR, energy entropy, energy,	CNN, RNN/ SVD	•The findings indicate that LSTM-RNN is more powerful than DNN.	87.11%

The detection and classification of pathological speech has already been the subject of our research [35]. This study was conducted on the Nemours dysarthric speech database, where samples with a diseased voice were identified among samples with a healthy voice. The study was based on the extraction of acoustic parameters, in particular: MFCCs, pitch and its variants, HNR and jitter and schimmer with its variants as well. The study demonstrated that the MFCC coefficients exhibited the highest classification accuracy, with a rate of 99.69% when utilising RNNs. However, when compared with SVM, which reached 97% classification accuracy, the effectiveness of RNNs for the classification of the degree of severity was less pronounced.

VI. CONCLUSION

A survey of voice disability detection methods found in the literature is presented in this paper. Due to the difficulty of analyzing the voice signal, it has been demonstrated in the literature that voice disability detection is an extremely difficult task. Depending on the type of disability, the voice signals differ significantly. Numerous algorithms have been documented in the literature. None of these algorithms, however, can identify any particular kind of voice impairment. Therefore, when designing the algorithm, it is crucial to focus on a specific disability. This survey paper also demonstrates how difficult it is to select the voice samples. Since the unvoiced portion of the speech samples also shows signs of pathology detection, the researchers should concentrate on both the voiced and unvoiced components of the samples.

Researchers face challenges in selecting features from samples for voice disability detection. Frequency domain features are commonly used, but time-domain measures are less common. Acoustic features are also used, but can be sensitive to patient pathology. Multiple features analysis is common. Classification algorithms, such as Support Vector Machine (SVM), are used, but SVM is not suitable for identifying voice disability levels. Large data sets are needed for training and testing, and researchers use various tools for classification. To design an effective voice pathology detection algorithm, researchers should focus on proper voice samples and feature collection, and design a level-based algorithm suitable for specific pathologies.

REFERENCES

[1] W. B. Snow, "Audible frequency ranges of music, speech and noise," in The Bell System Technical Journal, vol. 10, no. 4, pp. 616-627, Oct. 1931, doi: 10.1002/j.1538-7305.1931.tb02334.x.

[2] AL-Dhief FT, Latiff NMAA, Malik NNNA, Sabri N, Baki MM, Albadr MAA, et al. Voice pathology detection using machine learning technique. 2020 IEEE 5th International Symposium on Telecommunication Technologies (ISTT). Manhattan, New York, USA: IEEE; 2020. p. 99–104.

[3] Mohammed MA, Abdulkareem KH, Mostafa SA, Khanapi Abd Ghani M, Maashi MS, Garcia-Zapirain B, et al. Voice pathology detection and classification using convolutional neural network model. Appl Sci. 2020;10(11):3723.

[4] Subathra MSP, Mohammed MA, Maashi MS, Garcia-Zapirain B, Sairamya NJ, George ST. Detection of focal and non-focal electroencephalogram signals using fast walsh-hadamard transform and artificial neural network. Sensors. 2020;20(17):4952.

[5] Al-Nasheri A, Muhammad G, Alsulaiman M, Ali Z, Malki KH, Mesallam TA, et al. Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions. IEEE Access. 2018;6:6961–74. doi: 10.1109/ACCESS.2017.2696056.

[6] Islam R, Tarique M, Abdel-Raheem E. A survey on signal processing based pathological voice detection techniques. IEEE Access. 2020;8:66749–76. doi: 10.1109/ACCESS.2020.2985280.

[7] Leung, W. Z., Cross, M., Ragni, A., & Goetze, S. (2024). Training Data Augmentation for Dysarthric Automatic Speech Recognition by Text-to-Dysarthric-Speech Synthesis. arXiv preprint arXiv:2406.08568.

[8] Mansoor A, Cerrolaza JJ, Idrees R, Biggs E, Alsharid MA, Avery RA, et al. Deep learning guided partitioned shape model for anterior visual pathway segmentation. IEEE Trans Med Imaging. 2016;35(8):1856–65. doi: 10.1109/TMI.2016.2535222.

[9] Shan J, Li L. A deep learning method for microaneurysm detection in fundus images. IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); 2016. p. 357–8. doi: 10.1109/ CHASE.2016.12.

[10] Fritscher K, Raudaschl P, Zaffino P, Spadea M, Sharp G, Schubert R. Deep neural networks for fast segmentation of 3D medical images. International Conference on Medical Image Computing and Computer-Assisted Intervention. Vol. 9901. 2016. p. 158–65.

[11] Cummings JL, Benson DF, Hill MA, Read S. Aphasia in dementia of the alzheimer type. Neurology. 1985;35(3):394–4. doi: 10.1212/wnl.35.3.394.

[12] Forbes K, Shanks AMV. Distinct patterns of spontaneous speech deterioration: An early predictor of Alzheimer's disease. Brain Cognition. 2002;48(2–3):356–61. doi: 10.1006/brcg.2001.1377.

[13] Ogar, J., Slama, H., Dronkers, N., Amici, S., & Luisa Gorno-Tempini, M. (2005). Apraxia of speech: an overview. Neurocase, 11(6), 427-432.

[14] Ziegler, W., Aichert, I., & Staiger, A. (2012). Apraxia of speech: Concepts and controversies.

[15] X. Menendez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio and H. T. Bunnell, "The Nemours database of dysarthric speech," Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96, Philadelphia, PA, USA, 1996, pp. 1962-1965 vol.3, doi: 10.1109/ICSLP.1996.608020.

[16] Kent, R.D., Weismer, G. Kent, J.F., and Rosenbek, J.C., 'Toward Phonetic Intelligibility Testing in Dysarthria," Journal of Speech and Hearing Disorders, 54, 482-499, 1989.

[17] Rudzicz, F., Namasivayam, A. K., & Wolff, T. (2012). The TORGO database of acoustic and articulatory speech from speakers with dysarthria. Language resources and evaluation, 46, 523-541.

[18] Kim, H., Hasegawa-Johnson, M., Perlman, A., Gunderson, J. R., Huang, T. S., Watkin, K. L., & Frame, S. (2008, September). Dysarthric speech database for universal access research. In Interspeech (Vol. 2008, pp. 1741-1744).

[19] Barry WJ, Pützer M. Saarbrucken voice database. Institute of Phonetics, University of Saarland. http://www. stimmdatenbank.coli.uni-saarland.de/

[20] Al-nasheri A, Muhammad G, Alsulaiman M, Ali Z, Mesallam T, Farahat M, et al. An investigation of multi-dimensional voice program parameters in three different databases for voice pathology detection and classification. J Voice. 2017;31:113.e9–18. doi: 10.1016/j.jvoice.2016.03.019. [online] Voice and Speech Laboratory | Mass. Eye and Ear (masseyeandear.org).

[21] Kay Elemetrics Corp., Disordered Voice Database, Version 1.03 (CD-ROM), MEEI, Voice and Speech Lab, Boston, MA; October 1994.

[22] Saenz-Lechon N, Godino-Llorente JI, Osma-Ruiz V, Gomez-Vilda P. Methodological issues in the development of automatic systems for voice pathology detection. Biomedical Signal Processing and Control. 2006;1(2):120–8.

[23] Roy N, Merrill RM, Thibeault S, Parsa RA, Gray SD, Smith EM. Prevalence of voice disorders in teachers and the general population. J Speech Lang Hear Res. 2004;47(2):281–93.

[24] Irshad, U., Mahum, R., Ganiyu, I., Butt, F. S., Hidri, L., Ali, T. G., & El-Sherbeeny, A. M. (2024). UTran-DSR: a novel transformer-based model using feature enhancement for dysarthric speech recognition. *EURASIP Journal on Audio, Speech, and Music Processing*, *2024*(1), 54.

[25] Xie L, Liu Z. A comparative study of audio features for audio to visual cobversion in MPEG-4 COMPLIANT FACIAL ANImation. Proc. of ICMLC, Dalian; 2006. p. 13–6.

[26] Soleymanpour, M., Johnson, M. T., Soleymanpour, R., & Berry, J. (2022, May). Synthesizing dysarthric speech using multi-speaker tts for dysarthric speech recognition. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 7382-7386). IEEE.

[27] O. Buza, G. Toderean, A. Nica, and A. Caruntu, ``Voice signal processing for speech synthesis," in Proc. IEEE Int. Conf. Autom., Qual. Test., Robot., Cluj-Napoca, Romania, vol. 2, May 2006, pp. 360-364, doi: 10.1109/AQTR.2006.254660.

[28] Goyani, M., Dave, N., & Patel, N. M. (2010, November). Performance analysis of lip synchronization using LPC, MFCC and PLP speech parameters. In 2010 international conference on computational intelligence and communication networks (pp. 582-587). IEEE.

[29] C. M. Vikram and K. Umarani, ``Pathological voice analysis to detect Neurological Disorder using MFCC and SVM," Int. J. Adv. Electr. Electron. Eng., vol. 2, no. 4, pp. 87-91, 2013.

[30] Teixeira, J. P., & Fernandes, P. O. (2014). Jitter, shimmer and HNR classification within gender, tones and vowels in healthy voices. Procedia technology, 16, 1228-1237.

[31] Ghoniem R. Deep genetic algorithm-based voice pathology diagnostic system deep genetic algorithm-based voice pathology diagnostic system. Researchgate. 2019;11608:220–33. doi: 10.1007/978-3-030-23281-8_18.

[32] Al-Dhief F, Latiff N, Malik N, Salim N, Baki M, Albadr M, et al. A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms. IEEE Access. 2020;8:64514–33. doi: 10.1109/ACCESS.2020.2984925.

[33] Fan Z, Wu Y, Zhou C, Zhang X, Tao Z. Class-imbalanced voice pathology detection and classification using fuzzy cluster oversampling method. Appl Sci. 2021;11:3450.

[34] Syed S, Rashid M, Hussain S, Zahid H. Comparative analysis of CNN and RNN for voice pathology detection. BioMed Res Int. 2021;2021:1–8. doi: 10.1155/2021/6635964.

[35] Hamza, A., Addou, D., & Kheddar, H. (2023, November). Machine learning approaches for automated detection and classification of dysarthria severity. In 2023 2nd International Conference on Electronics, Energy and Measurement (IC2EM) (Vol. 1, pp. 1-6). IEEE.

Design and Implementation of a Low-cost Wireless Charger for Mobile Phones

Leila SADOUKI
Electronics department
Institute of Electrical and Electronic Engineering
University M'Hamed Bougara of Boumerdes
Boumerdes, Algeria
l.sadouki@univ-boumerdes.dz

Feriel LAMINI
Electronics department
Institute of Electrical and Electronic Engineering
University M'Hamed Bougara of Boumerdes
Boumerdes, Algeria
fifilam18@gmail.com

Nourhane AYADI
Electronics department
Institute of Electrical and Electronic Engineering
University M'Hamed Bougara of Boumerdes
Boumerdes, Algeria
nhane6828@gmail.com

***Abstract*— In recent times, smart phones have become indispensable in our lives. However, their extensive use consumes a significant amount of our time; an average of 3 hours per day; consequently, frequent charging becomes necessary. This paper focuses on creating a magnetic inductance-based wireless mobile charger system that efficiently transfers power from a transmitter to a receiver, and by utilizing resonant inductive coupling, it aims to achieve cable-free charging. The successful outcome will be a functional 1.5W wireless charger prototype, along with comprehensive circuit documentation detailing the design choices, components selection, testing procedures, and the achieved results. This subject contributes to a growing field with the potential to revolutionize mobile device charging as it offers a practical experience in designing, building, and testing a complete wireless power transfer system.**

Keywords*— *Wireless, Power Transfer, Charger, Mobile Phone.

I. INTRODUCTION

As mobile phones become increasingly integrated into our daily lives, the need for convenient and efficient charging solutions grows ever stronger. Traditionally, these devices were charged using wired chargers, however, this latter has limitation, that's why scientists suggested wireless charging [1]. Wireless charging technology has transitioned from theoretical concepts to established standards, rapidly becoming a mainstream feature in consumer electronics, particularly mobile phones, and portable devices. This widespread adoption is driven by several key advantages. Firstly, wireless charging eliminates the need for fiddly cables, enhancing user convenience. Additionally, it fosters compatibility, allowing a single charger to work with various brands and models. Secondly, the contactless nature of wireless charging improves product durability by eliminating wear and tear on charging ports, contributing to features like waterproofing and dustproofing. Furthermore, wireless charging offers greater flexibility, particularly for devices were replacing batteries or connecting cables is expensive, hazardous, or impractical, such as body-implanted medical sensors. Besides, wireless charging offers the potential for on-demand power delivery, potentially reducing overcharging issues and minimizing energy consumption. This growing trend is evident, with major smartphone manufacturers like Samsung, Apple, and Huawei integrating wireless charging capabilities into their flagship devices since 2014 [2]. This paper tackles this challenge by designing and implementing a complete wireless charging system for mobile devices. It focuses on a safe and user-friendly wireless charging system adhering to established standards (like Qi). The system will consist of two key components: a transmitter pad that generates the magnetic field and a receiver coil integrated into a mobile phone that captures the energy for charging. The circuit prioritizes optimizing power transfer efficiency, minimizing energy loss during transmission. This endeavour offers valuable hands-on experience in designing, building, and testing a complete wireless power transfer system, deepening our understanding of this rapidly evolving technology [2].

II. OVERVIEW ABOUT WIRELESS POWER TRANSFER

A. Brief history of wireless power transfer

We introduce in this section the evolution of wireless power transfer throughout the years.

- 1826-1831 André-Marie Ampère formulated Ampère's circuital law. Following Ampère's work, Michael Faraday established Faraday's law of induction. This law explains how a changing magnetic field can induce an electric current in a conductor [3].
- 1891-1894 Nikola Tesla demonstrated wireless power transfer using electrostatic induction [3].
- 1899 Tesla lighted 200 bulbs by 108 volts of high frequency electric power over 25miles [2].
- 1904–1917 Tesla's Wardenclyffe Tower aimed to demonstrate transatlantic wireless communication and power transmission. Unfortunately, his strategy was not practical and after that it was shattered during World War I [4,3].
- 1980-1990 Gerald Brown's demonstration of powering a model helicopter with microwaves significantly influenced research on microwave-powered airplanes [2].
- 2007-2013 The MIT-led WiTricity research team used strongly magnetic resonance to wirelessly power a 60 W light bulb at 2 m distance with 40% power efficiency [3].

B. Wireless Charging Techniques

Wireless Power Transfer (WPT) systems are categorized based on the distance between the transmitter and receiver; This distance impacts how Electromagnetic Fields (EMFs) behave. There are three categories of transmission. First, near

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

field zones which are ideal for short-range applications like phone charging pads. It uses strong inductive or capacitive coupling in the EMFs. Second, mid field zones that offer some operational distance but have lower efficiency due to the mix of near and far-field characteristics. Finally, far field zones utilize radiating EM waves like radio waves, enabling long-range power transfer [5].

1) Magnetic Inductive Coupling: is a method of transmitting electrical energy between two coils using magnetic field induction. It occurs when a transmitter generates a magnetic field, causing a voltage/current to be induced by the receiver. The coupling's tightness depends on alignment, distance, diameter ratio, and coil shape [2].

2) Magnetic Resonant Coupling: is a method of transferring electrical energy between two resonant coils through varying magnetic fields, achieving high energy efficiency with minimal leakage. It offers immunity to neighboring environments and line-of-sight transfer requirements and allows concurrent charging of multiple devices [2].

3) Microwave Radiation: uses microwaves to carry radiant energy, propagating over space at the speed of light. The power transmission system starts with Alternating to Direct Current (AC-to-DC) conversion and DC-to-RF (Radio Frequency) conversion. The frequency ranges from 300MHz to 300GHz. Other electromagnetic waves like infrared and X-rays are not widely used due to safety concerns [6].

C. WPT standards

Various standards are set for wireless charging. Two among them are popular for significant Smartphone manufacturers, which are Qi and Alliance for Wireless Power (A4WP).

1) Qi: is a wireless charging standard created by the Wireless Power Consortium (WPC). It ensures compatibility between wireless chargers and devices, allowing for seamless data exchange alongside power transfer. Qi empowers the receiving device to control the charging process. The charger itself, as long as it's Qi-compliant, can adjust its power output based on the device's requests through a communication protocol. This technology utilizes magnetic inductive coupling, typically operating within a short range of 4 centimetres distance [8].

2) Alliance for Wireless Power (A4WP): standard takes a different approach to wireless charging compared to Qi. A4WP uses magnetic resonance to create a larger electromagnetic field, enabling several advantages. Unlike Qi, which requires precise phone placement, A4WP allows spacing and doesn't need perfect alignment. This translates to greater freedom - we can charge our device from several meters away and even place objects on the charging surface without interrupting the process. A single A4WP charger can even power multiple devices at once, with each receiving the power it needs. This flexibility in charging distance, multi-device support, and resistance to interference makes A4WP a promising contender for future wireless charging solutions [9].

D. Fundamentals of WPT

The core principles of Wireless Power Transfer are adopted from the fundamental laws of electromagnetism, namely Ampere's Law and Faraday's Law of Induction.

1) Ampere's Law: When electric current flows through a conductor, it generates a surrounding magnetic field. The strength of this magnetic field is directly proportional to the magnitude of the current. In WPT systems, the transmitter coil utilizes this principle to create a magnetic field [10].

2) Faraday's Law of Induction: A changing magnetic field can induce an electromotive force (EMF) in a nearby conductor. This EMF can then be used to create a current in the conductor. In WPT, the receiver coil obeys this principle. As the oscillating magnetic field from the transmitter coil sweeps across the receiver coil, an EMF is induced, which in turn drives current flow within the receiver circuit to charge the mobile device [10].

E. Applications of Wireless Power Transfer Systems

Wireless power transfer technology can be applied across a wide variety of applications and environments such as:

- Consumer electronics products.
- Automotive applications.
- Industrial applications.
- Medical applications.
- Military equipment.

In our paper, we employed magnetic resonant coupling to charge mobile phones.

III. WIRELESS CHARGING USING RESONANT COUPLING TECHNIQUE

Diving into the fundamental principles of wireless power transmission systems. Electrical energy is transferred through a localized magnetic field. This time-varying field is generated by the flow of Alternating Current (AC) in a coil. When the field generated by one coil reaches another coil, it induces a voltage in this latter. This voltage can then be used to power or charge a device connected to the coil. A wireless power transfer system typically consists of multiple stages to achieve the desired output. Two sections are involved in its design, a transmitter (**Tx**), and a receiver circuit (**Rx**) (see Fig.1). The transmitter is built of an oscillator, a power amplifier, and the tuned LC circuit at its end which is connected to its similar one at the receiver's input. The received signal will go through a rectifier then a voltage regulator that will provide the 5V DC needed for the battery charging [11].

1) Transmitter section ***Tx***

The transmitter circuit consists of an oscillator that generates a square wave from a DC power supply. This wave is then amplified by a power amplifier, which drives a MOSFET to operate in the switching mode. As a result, an alternating current is produced across the transmitter coil.

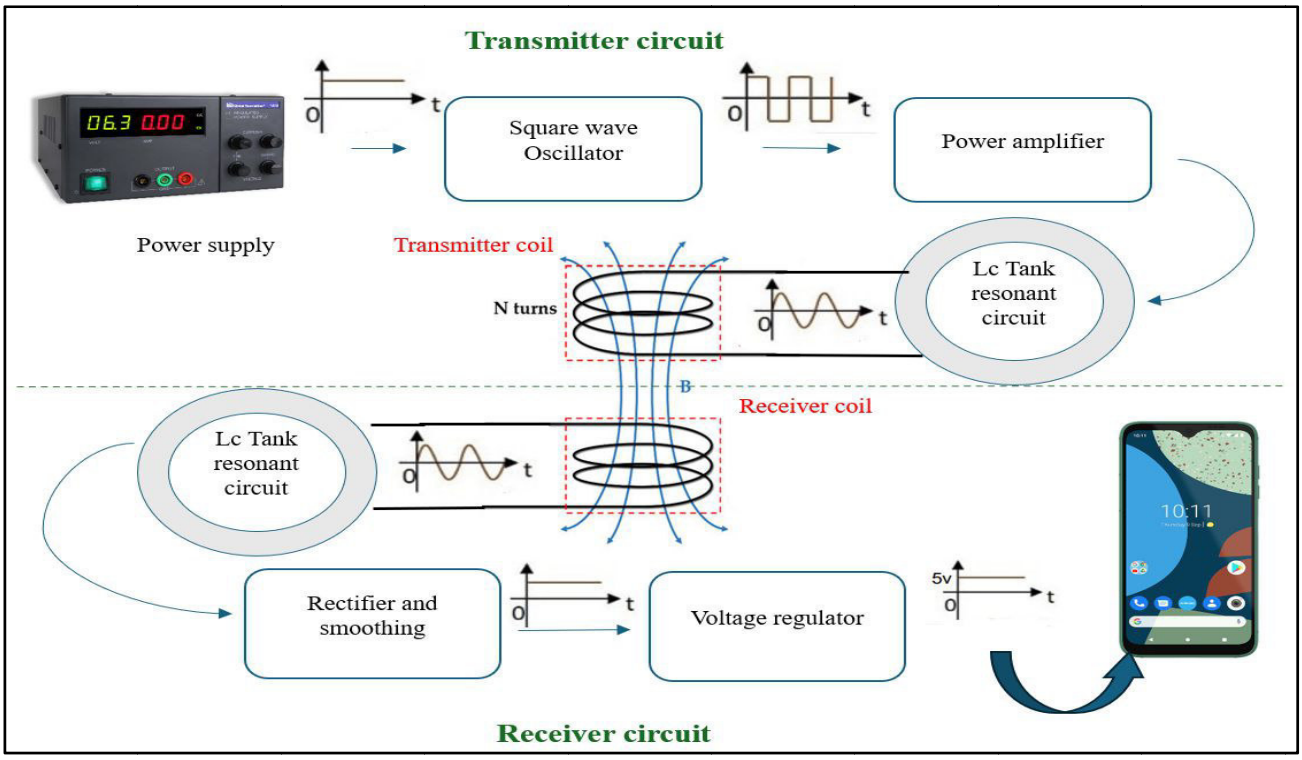


Fig. 1. A. Block Diagram of a typical wireless charging system for smart phones

- Oscillator :

The voltage regulator LM7805 will regulate this voltage to 5V DC, to power both 555 Timer and NPN transistor. The 555 Timer operates in three modes: Bistable (for toggling between states), Monostable (generating single pulses), and Astable Multivibrator mode (continuous oscillations) [12]. This latter mode was used in our circuit to generate a high frequency AC square signal. (Oscillating frequency = 18.95kHz)

$$f=1.44/((R3+2R1)*C3) \quad (1)$$

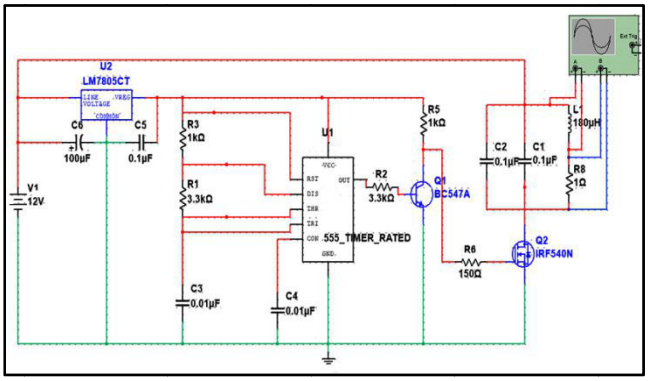


Fig. 2. Circuit design of WPT transmitter

- Power Amplifier

The square wave signal generated by the 555 timer (pin 3) controls an NPN transistor BC547. The collector was powered by 5V DC from the voltage regulator. This transistor acts as a driver, it amplify the weak signal (low-power signal) from the timer to a higher current level suitable for driving the MOSFET [13].

- LC tank circuit Tx

The amplified signal then triggers a MOSFET, known for its fast-switching ability, to create the high-frequency AC current needed to energize the transmitter coil. By applying ampere's law, a magnetic field is generated by AC current. To maximize power transfer efficiency and minimize leakage losses, the transmitter circuit is tuned to resonate at the same frequency as the receiver coil. Finally, a capacitor is connected in parallel with the transmitter coil to form a resonant circuit (tank circuit). This ensures the system to operate below the coils' self-resonant frequency, preventing inefficiencies, and the reason why parallel resonant circuit is used is because itprovides efficiency with only small input current [1][7].

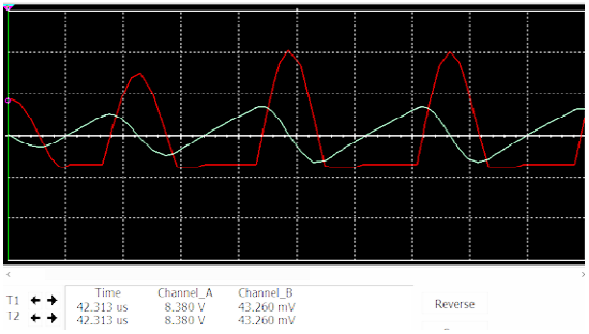


Fig. 3. Oscilloscope output voltage and current of transmitter's coil.

As we have said the resonant frequency for transmitter circuit is equal to that of the receiver, we got a high impedance matching due to the parallel tank circuit that consists of capacitance of 0.2µF and inductance of the coil of value 180 µH, this made us able to calculate the resonant frequency using this equation:

$$f = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{180*10^{-6}*0.2*10^{-6}}} \quad (2)$$

$$= 26.526\ \text{kHz}$$

We constructed the transmitter coil (denoted as L) ourselves using 24 SWG (Standard Wire Gauge) enamelled copper wire with an inner radius of 0.56 mm. To achieve the desired resonant frequency for the tank circuit, we carefully selected the values of both the capacitor and the inductor (L). To reach a target inductance of 180µH, we employed the following formula [1]:

$$L = \frac{R^2 N^2}{9R + 10l} \quad (3)$$

- Such that R is the radius of the inductance, N is the number of turns and l is the length of the coil.

In our case we took R= 2.5cm, N= 45 turns, and a 0.55cm of length.

2) Receiver circuit Rx

The wireless mobile charger circuit works by receiving an AC electromagnetic field from the transmitter coil. The receiver coil captures this electromagnetic field and converts it into a weak AC current. Then it is rectified into a pulsating DC voltage by the bridge rectifier. The capacitors smooth out this pulsating voltage into a steadier DC voltage. Finally, the voltage regulator regulates this voltage to a constant 5V DC output, which is then delivered to the mobile device through the USB connector.

- LC tank circuit Rx

The receivers coil captures the oscillating magnetic field transmitted from the transmitter circuit. The magnetic field induces a small AC voltage according to Faraday's law. To achieve resonance in the wireless power transfer system, capacitors are placed in parallel with both the transmitter and receiver coils. This capacitor-inductor combination, known as a tank circuit, creates a characteristic oscillation. The capacitor charges and discharges transferring its energy to the inductor which builds a magnetic field that strengthens as the capacitor empties. When the capacitor is fully discharged, the magnetic field collapses, and the cycle repeats. This exchange of energy sets the circuit oscillating at a specific frequency. Critically, by ensuring both the transmitter and receiver coils have the same resonant frequency, the system becomes more efficient by blocking out unwanted frequencies and allowing only the desired resonant frequency to transfer power. This selective transfer helps eliminate interference and optimizes energy delivery [14].

- As in the transmitter the resonant frequency must be the same, so we have used 195 µH receiving coil and capacitance of overall 192nF.

$$f = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{195*10^{-6}*192*10^{-9}}} = 26.01\ \text{kHz} \quad (4)$$

- Rectifier and filter

Full wave bridge rectifier converts the weak AC current from the receiver coil into a pulsating DC voltage. The filter capacitor takes the pulsating DC output from the rectifier and smooth it out into a steadier DC voltage.

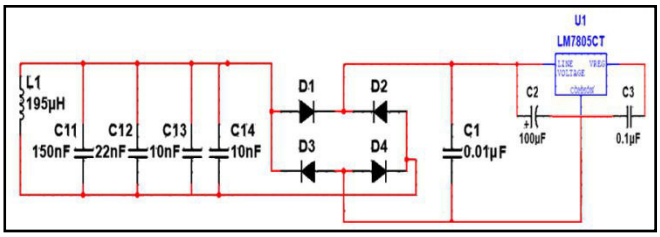


Fig. 4. The applied receiver's circuit.

IV. Implementation and Discussion

The voltage regulator IC 7805 maintains a constant voltage output for fluctuating of an input voltage and the output voltage is +5V DC. In our design it was used to power the timer and collector of the transistor from the power supply. The role of the input capacitor connected to the LM7805 is:

A. Noise Filtering:

If the LM7805 is located far from the main power supply filter, the length of the connecting wires can act as an antenna, attracting electrical noise. The input capacitor helps eliminating this noise by providing a low-impedance path for high-frequency components [15].

B. Stability:

It improves the regulator's stability by reducing voltage fluctuations caused by sudden load changes or transient events [15].

C. Preventing Oscillations:

In adjustable voltage regulators, the input capacitor helps prevent oscillations and ensures smooth operation [15].

Whereas, the purpose of the output capacitor connected to the LM7805 is:

A. Transient Response:

When the load current changes suddenly, the output capacitor helps maintain a stable output voltage. It reduces voltage spikes caused by rapid load variations [15].

B. *Filtering:*

It further filters any remaining high-frequency noise on the output voltage [15].

The NPN transistor (BC547A) receives a control signal (a pulse train from the 555-timer circuit) at its base. This transistor acts as a driver; amplifying the weaker signal from the timer into a stronger current, since it has a higher DC current gain from 110 to 800.

To create the alternating current, we have used an N-channel MOSFET in the switching mode because of:

a) Fast switching: MOSFETs exhibit rapid switching speeds, enabling high-frequency operation.

b) Current Rating: It is capable of handling relatively high continuous drain current, typically in the range of 22A to 33A.

The amplified current from the driver transistor controls the gate voltage of the IRF540N MOSFET. When the gate voltage is high, "5V", the MOSFET turns on allowing current to flow from the power supply to the transmitter coil. Conversely, when the gate voltage is low "0V", the

MOSFET switches OFF, blocking current to the coil. Concerning the capacitors in the LC tank circuit, we did not find a polyester capacitor of value 0.2µF so ; we have combined two parallel capacitors of 0.1µF. For the power supply, we noticed that in our circuit we need high current to successfully transfer power wirelessly so, Basically, a power supply less than 0.5A will not run our circuit, so, 1A power supply was used.

For the coil, we have designed them using 24 SWG enameled copper wire. And a diameter of 50mm was used. After that we have checked the value of our coil using LCR meter and we got the value of 181.7µH.

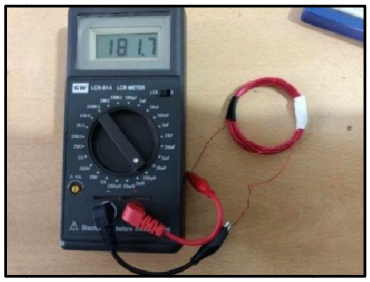


Fig. 5. The measured value of the coil on LCR meter.

We did not find a 192nF capacitor. So, 4 parallel capacitors were combined: one 150nF, one of 22nF and two of 10nF. The **Tx** and **Rx** tank circuits were tuned at the same resonant frequency to maximize power and eliminate leakage losses.The AC voltage across the coil is shown below:



Fig. 6. The voltage across Rx coil.

The circuit implementation was successful, resulting in an output voltage of 5V and an output current of 0.3A, which effectively charges smart phone. The wireless charger can operate at a distance of up to 2cm between the receiver coil and transmitter coil. Remarkably, it reaches a full charge in just 1 hour.

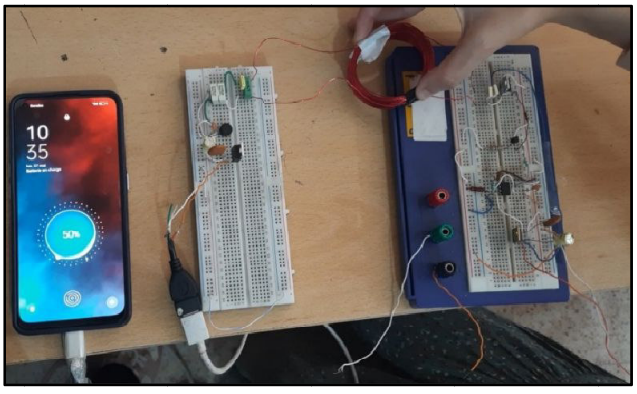


Fig. 7. The functionality of wireless charger.

The following tables collect the characteristics as well as the total cost of our circuit.

TABLE I. EXPERIMENT RESULTS

Parameter	*Value*
Input Voltage	7 V
Input Current	1 A
Input power	7 watts
Output voltage	5 Volt
Output current	0.3A
Output power	1.5 Watts
Efficiency (Output power / Input power)	0.21 = 21%

TABLE II. CIRCUIT COST

Components	*Quantity*	*Price*
MOSFET IRF540N	2	240 DA
NPN TRANSISROR BC547	1	35 DA
Resistances	5	25 DA
Bridge Rectifier	1	5 DA
Capacitors	13	185 DA
Voltage regulator LM7805	2	440 DA
Coil(using 24 SWG enamled wire)	2	170 DA
Total price	_	**1100 DA**

From the above table, it is shown that the design is cheap compared to commercial ones.

V. CONCLUSION

In the modern era, wireless technology has advanced significantly and is now widely used in various electronic applications. As wireless communication systems become more affordable, easier to implement, and smaller, an increasing number of devices can benefit from them. Wireless solutions offer time-saving and user-friendly alternatives. Additionally, wireless condition monitoring enables applications that are not feasible through traditional wired net-works.

Throughout this design, a deep understanding and a detailed comprehension of the different circuits, elements and components used were processed. Several tests and experiments were done to ensure the credibility of the outcomes.

As a first step to start, a simple circuit schematic was designed and simulated. After passing the simulation phase, it was time to build a real circuit and implement all the equipment and elements needed to finally reach the goal of this design , wireless charger.

From an efficiency standpoint, wire-less inductive power transfer is practical for general power applications only if the transmitter and receiver coils are near each other. Attempting inductive power transfer over larger distances is not feasible due to its very low efficiency. As the design process significantly influences the system's efficiency and ensures adherence to electrical specifications for voltage and current,

it should begin with careful selection of a suitable compensation topology and types of components needed in the topic.

A lot of hardships and problems have encountered us during this journey, and finally obtain satisfactory results. This made us realize what a true electronics engineer would go through.

Basically, the wireless charger is designed to charge the mobile battery. But in the future, by making some modifications, it can be used for different portable devices, for example, Laptop, iPad, digital camera, electric vehicles. With the development of the extended distance wireless charging, it is possible to charge the devices by walking outside on the street. It will work like a hotspot area like a Wi-Fi hotspot, the device gets charging while users are walking or speaking. Now, wireless charging would allow the vehicles to power up. This wireless charging vehicle would not use plugs or charging cords. Drivers would park their electric vehicles over a wireless energy source.

REFERENCES

[1] C. Garth V., and J. Deppong. "Undergraduate Application-Based Research: Design and Testing of a Wireless Phone Charger," Technology Interface International Journal, vol. 14, no. 1, pp. 45-52, Jan. 2013.

[2] X. Lu, D. Niyato, P. Wang, D. In Kim and Z. Han, "Wireless charger networking for mobile devices: fundamentals, standards, and applications," in IEEE Wireless Communications, vol. 22, no. 2, pp. 126-135, April 2015.

[3] T. Sun , X. Xie , Z. Wang .2013.Wireless power transfer for medical microsystems. Springer London.

[4] Z. Wei, and C. Chris Mi, "Compensation Topologies of High-Power Wireless Power Transfer Systems," 2015.

[5] S. Abdallah Toni, "master's Thesis in Electrical and Computer Engineering, Specialization in Telecommunications: Wireless Charger for Low Power Devices," Master's thesis University of Coimbra], 2016.

[6] https://www.britannica.com/science/electromagnetic-radiation/Microwaves. (Accessed on June 2024)

[7] P. Devendra, M. K. McDonough, J. M. Miller, B. Fahimi, and P. T. Balsara, "Wireless Power Transfer for Vehicular Applications: Overview and Challenges," 2018.

[8] https://www.wirelesspowerconsortium.com/standards/qi-wireless-charging/ (Accessed on June 2024)

[9] https://www.phonearena.com/news/Wireless-charging-standards-Qi-Powermat-A4WP-what-does-it-all-mean-and-who-will-prevail_id57371 (Accessed on June 2024)

[10] J. Talit, P. Uthansakul, and H. S. Lui, "Implementation of Wireless Charger for Mobile Phone Based on Solar Energy," 2014.

[11] H. Amanda, "Transmetteur d'énergie sans fil à haute puissance pour la recharge d'appareils électroniques," Mémoire de maîtrise, Département de génie électrique, Polytechnique Montréal , 2019.

[12] https://www.watelectronics.com/555-timer-ic-introduction-and-working-with-operating-modes/ (Accessed on June 2024)

[13] https://reversepcb.com/bc547-transistor/ (Accessed on June 2024)

[14] W. Eberhard, "Wireless Power for Mobile Devices," In 33rd IEEE International Telecommunications Energy Conference (INTELEC), 2011.

[15] https://www.quarktwin.com/blogs/integrated circuit/l7805cv-pinout-a-comprehensive-guide-to-understanding-and-using-l7805cv-voltage-regulator/245 (Accessed on June 2024)

Design and Implementation of a Digital pH Meter with Analog Conditioning Circuit Calibration

Leila SADOUKI
Electronics department
Institute of Electrical and Electronic Engineering
University M'Hamed Bougara of Boumerdes
Boumerdes, Algeria
l.sadouki@univ-boumerdes.dz

Marouane DAOUDI
Electronics department
Institute of Electrical and Electronic Engineering
University M'Hamed Bougara of Boumerdes
Boumerdes, Algeria
maroouuane.daoudi@gmail.com

Yacine BENGHARABI
Electronics department
Institute of Electrical and Electronic Engineering
University M'Hamed Bougara of Boumerdes
Boumerdes, Algeria
bengharabiyacine497@gmail.com

***Abstract*— This paper introduces a novel digital pH meter designed to enhance the accuracy, portability, and ease of use in pH measurement. The proposed device combines modern digital technology with traditional pH sensing methods to overcome the limitations of conventional analog conditioning circuits and their calibration. The study begins with a review of existing pH measurement techniques and devices, which informs the development of a microcontroller-based prototype. Extensive testing demonstrates that the device offers accuracy and reliability on par with laboratory-grade instruments. Notable features include portability, real-time data logging, automatic temperature compensation, and calibration reminders. This digital pH meter provides a cost-effective, versatile solution for diverse applications, marking a significant step forward in pH measurement technology.**

Keywords—PH meter, ESP32 microcontroller, digital, LCD display

I. Introduction

pH measurement plays a crucial role in many scientific and industrial fields, such as environmental monitoring, chemical analysis, water treatment, and food processing. Accurate pH measurement is essential to ensure product quality, compliance with regulatory standards, and safe operational conditions. Traditionally, pH meters have been analog, but recent advancements in digital technology have led to the development of digital pH meters that offer improved precision, ease of use, and additional functionality. These devices incorporate microprocessors, advanced sensors, and digital displays, making them more reliable and accessible for a range of applications.

The measurement and control of pH, which represents the hydrogen ion concentration or acidity/basicity of a solution [1], is of paramount importance across numerous scientific disciplines and industrial applications.The pH scale is defined by the negative logarithm of the hydrogen ion activity or effective concentration [2]. A pH of 7 is considered neutral, with solutions having a pH less than 7 being acidic, and solutions with a pH greater than 7 being basic or alkaline. The lower the pH, the higher the acidity; the higher the pH, the higher the alkalinity [3]. The pH value provides crucial information about the chemical nature and reactivity of a solution. It governs various chemical and biological processes, such as enzyme activity, protein structure, and solubility of molecules. Even a slight change in pH can significantly impact the properties and behaviors of substances in solution. Historically, the developmentof pH measurement has been a pivotal advancement in the fields of chemistry, biochemistry, and related areas. It was in 1909 that Danish chemist Søren Peder Lauritz Sørensen introduced the concept of the pH scale [4], which quantifies the acidity or basicity of a solution based on the negative logarithm of the hydrogen ion concentration. This groundbreaking work built upon earlier theories by pioneers such as Arrhenius, Brønsted, and Lowry, who established the foundations of acid-base chemistry Traditional methods for pH measurement, such as litmus paper and colorimetric indicators, have been widely used but often lack the desired accuracy, precision, and convenience for many applications [5]. The advent of electrochemical pH meters, which employ specialized electrodes and electronic circuitry, revolutionized the field by providing rapid, reliable, and quantitative pH measurements. However, commercial pH meters can be costly, and their performance can be influenced by factors such as temperature, ionic strength, and electrode aging [6,7].

This paper explores the design and recent innovations in digital pH meters, focusing on advancements in sensor technology, signal processing, and user interface design. We aim to provide an overview of the key considerations in the design of these devices, along with a discussion of their impact on various scientific and industrial applications..

II. PH Measurement, Theoretical Concepts

A. Definition of pH

pH is a measure of the acidity or basicity of an aqueous solution. It represents the negative logarithm of the hydrogen ion concentration (H+) in the solution. The pH scale ranges from 0 to 14, with a value of 7 representing a neutral solution, values below 7 indicating an acidic solution, and values above 7 indicating a basic (alkaline) solution [1].

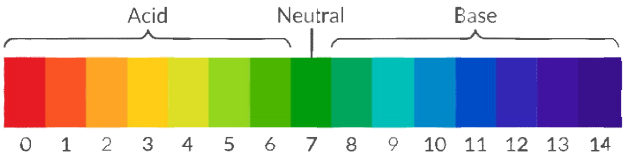


Fig. 1. PH scale [8]

A major breakthrough in pH measurement occurred in 1934 with the introduction of the glass electrode by American chemist Arnold Beckman [9]. The glass electrode, consisting of a thin glass membrane selectively permeable to hydrogen ions, revolutionized pH measurement by allowing for direct and precise potentiometric measurements. This electrode enabled the accurate determination of pH by measuring the potential difference across the glass membrane,

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

which is proportional to the hydrogen ion activity in the solution being measured.

The development of the glass electrode was a significant milestone in the field of pH measurement, as it provided a reliable and practical method for accurately measuring pH in various applications. Prior to this innovation, colorimetric methods using pH indicators were widely used, but they were limited in accuracy and required visual comparison with color charts.

B. Importance and Applications of pH Measurement

The measurement of pH is crucial in various fields, including:

- Environmental monitoring: Monitoring pH levels in water bodies, soil, and air to assess environmental quality and detect potential contamination.
- Industrial processes: Controlling and monitoring pH levels in various industrial processes, such as chemical manufacturing, food and beverage production, wastewater treatment, and pharmaceutical manufacturing.
- Agriculture: Monitoring soil pH to ensure optimal conditions for plant growth and nutrient uptake.
- Healthcare: Measuring pH levels in biological fluids (e.g., blood, urine) for medical diagnostics and treatment monitoring.

C. pH Measurement Principles

Acid-base equilibrium: The dissociation of acids and bases in aqueous solutions, resulting in the formation of hydrogen ions (H+) and hydroxide ions (OH-).

Hydrogen ion activity: The effective concentration of hydrogen ions in a solution, taking into account factors such as ionic strength and activity coefficients.

pH and pOH: The relationship between pH and pOH (the negative logarithm of the hydro

xide ion concentration) in aqueous solutions, where pH+ pOH = 14 at 25°C.

D. Glass Electrode

The most commonly used pH electrode, consisting of a thin glass membrane selectively permeable to hydrogen ions.

The glass electrode is composed of a pH-sensitive glass membrane and an internal reference electrode. The glass membrane develops a potential difference across its surface, which is proportional to the hydrogen ion activity in the solution being measured. This potential difference is measured against a reference electrode, providing a voltage that can be related to the pH value using the Nernst equation.

E. Nernst Equation and its Significance

The Nernst equation relates the potential difference (E) across the glass membrane to the pH of the solution being measured: [10]

$$E = E_0 + (RT/nF)* \ln([H+]) \quad (1)$$

Where:

E_0 : is a constant potential (reference potential).

R : is the universal gas constant.

T : is the absolute temperature.

n : is the number of electrons transferred in the electrode reaction.

F : is the Faraday constant.

[H+] : is the hydrogen ion activity in the solution

This equation forms the basis for calculating the pH value from the measured potential difference across the glass electrode.

F. Factors Affecting PH Measurement Accuracy

- Temperature effects: pH measurements are temperature-dependent, and temperature compensation is often required for accurate readings.
- Electrode aging and drift: Over time, glass electrodes can deteriorate, leading to drift in the measured potential and reduced accuracy.
- Interference from other ions: The presence of certain ions (e.g., sodium, potassium) in the solution can interfere with the pH measurement.
- Sample characteristics: Factors like ionic strength, viscosity, and suspended solids can affect the accuracy of pH measurements.

III. CIRCUIT DESIGN

The system aims to provide accurate pH measurements by integrating a high-impedance pH electrode, buffer amplifier, analog conditioning circuit, microcontroller with ADC(Analog to digital converter), and output display module. The pH electrode measures the potential difference based on hydrogen ion activity, which requires proper interfacing through a high-input resistance buffer. The analog conditioning circuit processes this signal to a suitable range for the microcontroller's ADC. The microcontroller reads the digitized voltage, applies filtering, calculates the pH using the Nernst equation, and displays the result on the output module. Simulations using Proteus software are conducted to verify the circuit designs and optimize component configurations before implementation. The following sections detail the design of each component, simulation results, and overall system integration for the digital pH meter..

A. Block Diagram

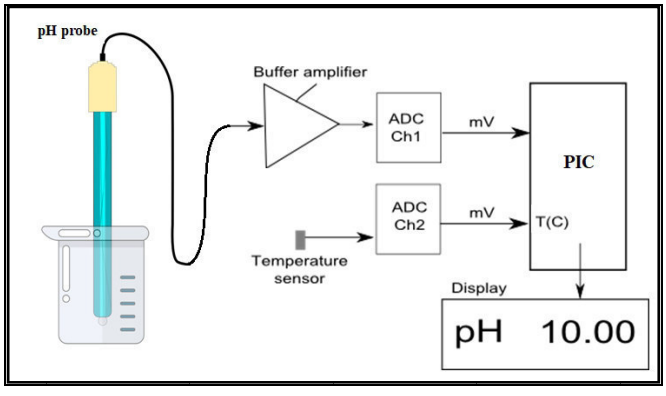


Fig. 2. Simplified block diagram of the digital pH mete

The block diagram in Fig. 2 shows the high-impedance pH electrode measuring potential difference based on

hydrogen ion activity. This signal goes through a buffer amplifier and then an ADC (Ch1) to the microcontroller. A temperature sensor is connected to a separate ADC (Ch2) for temperature compensation. The microcontroller calculates the pH using the digitized voltage and temperature, applying filtering algorithms. The calculated pH value is displayed on an output module.

B. pH probe

The E201-C BNC pH probe is an economical choice for a reliable pH electrode. Its main parameters are as follows :

TABLE I. PH PROBE PARAMETERS

Parameter	***Range***
Measuring Range	0 to 14 pH
Measuring temperature	0 to 80 C°
Zero point	7 +/- 0.5 PH (25C°)
Internal resistance	250M Ω(25C°)
Percent theoretical slope	(pts) >= 98.5
Response time	< 2minute
Noise	< 0.5 mVEquations

IV. ELECTRICAL CIRCUIT

The pH meter is basically a simple gain/offset circuit with high input impedance. Its purpose is to convert the voltage range of a typical pH probe (between -0.414 at pH 14 and 0.414V at pH 0). The very high input impedance of the pH meter is critical because pH probe has high output impedance, so we must use an even higher input impedance stage for the pH meter. The different stages of the pH meter are:

A. Input buffer stage

The circuit input is the BNC connector R1(1) . R1 limits the input current. 4.7MΩ is quite small compared to the impedance of the probe itself which is over 100MΩ. The input buffer amplifier U4 which is the **CA3240E op-Amp** provides the highest input impedance for our pH meter. It's a simple follower with unitary gain (buffer) and it just reproduces the input voltage to its output.

Following the Buffer we find a low pass RC filter (C1 and R2) which prevents higher frequencies to get into our circuit. Its cut-off frequency is set at about 94Hz (1/RC). Anything higher than that will be attenuated. The output of this stage is between -0.414 at pH 14 and 0.414V at pH 0.

B. Gain Stage circuit

P1 and the resistors R4 and R3 form the gain circuit. Using op-amp formula, the gain can be adjusted between (R4+R3+P1)/(P1+R3) and (R4+R3+P1)/(R3), which in this case turns out to be between 3.6 and 8. Why this range? A pH probe typically returns between -0.414 and 0.414V, and we'd like to have +/-2V at the output of this stage. Thus we need a gain of 5.

C. Offset stage

This op-amp changes the previous stage's +/-2V output range into a 0 to 4V range. This stage also needs to invert the signal, since with 2V input , pH is 0 and with -2V, pH 14. The op-amp is used in a inverting summing amplifier configuration. The pH signal coming through R5 will see a negative gain of -R8/R5. we add the offset which is obtained by -5V going through R7, R6 and P2. The gain for this branch is: -R8/(R7+P2+R6). The extremes for P2 (0 or 10KΩ) give an adjustment range of 1.6V to 2.5V which will be added to the pH signal. Note that the ideal offset value of 2V sits nicely in the middle of this interval. The resistors R9 and R10 are just for voltage division in case we need to change the range , which depends on the ESP32 internal ADC limitation of 3.3V input. The output of this stage is between 0V at pH 0 and 4V at pH 14 .

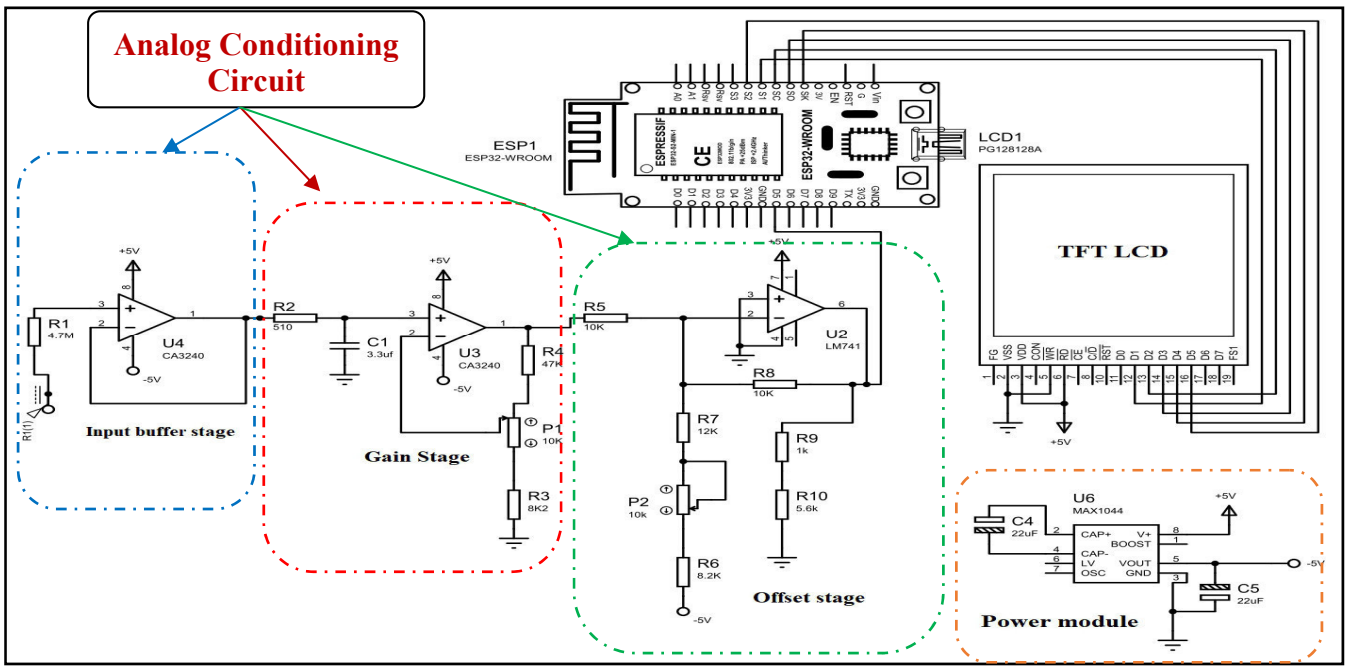


Fig. 3. Electrical circuit of the digital pH mete

D. ESP32 microcontroller [11]

The ESP32 is a low-cost, low-power microcontroller from Espressif Systems, featuring a dual-core 32-bit processor, Wi-Fi and Bluetooth connectivity, and a rich set of peripherals. Like the ADCs, capable of measuring signals from sensors and external circuitry.One ADC channel is used to digitize the conditioned pH signal, while the other channel reads the temperature sensor for compensation. The ESP32's processing power and memory resources allow for efficient execution of filtering algorithms, pH calculation using the Nernst equation, and data processing for display.

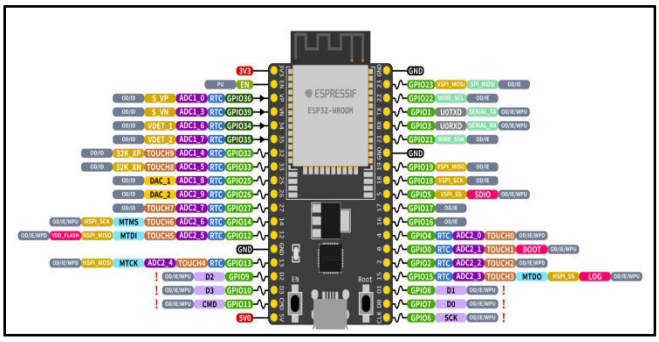


Fig. 4. ESP32 microcontroller pin-out

E. Power module [12]

This module provides regulated power supply voltages of 5V and 3.3V, commonly required for powering various electronic components and circuits. The module typically includes a USB connector or a barrel jack for connecting an external power source, such as a USB cable or a wall adapter. It has an on-board voltage regulator that converts the input voltage (usually 5V or higher) to the desired output voltages of 5V and 3.3V.

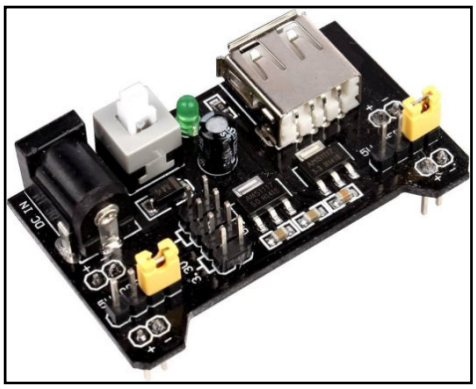


Fig. 5. 5/3.3V Power module

F. TFT display

TFT (Thin-Film Transistor) LCD display with a 65K color depth and a resolution of 128x128 pixels was used for the digital pH meter. This type of display offers several advantages over traditional character-based LCD displays like the 16x2 LCD:

- High resolution and color depth: The 128x128 resolution and 65K colors provide a more visually appealing and informative display, allowing for graphical representations, icons, and clearer text rendering.
- Larger display area: The 1.44-inch display size offers a larger viewing area compared to the compact 16x2 LCD, enabling the display of more information simultaneously.
- Graphical user interface (GUI): The TFT display supports the creation of graphical user interfaces, which can enhance the user experience and make the pH meter more intuitive to operate.
- Better readability: The high resolution and color depth of the TFT display ensure better readability, especially in varying lighting conditions, making it suitable for both indoor and outdoor applications.

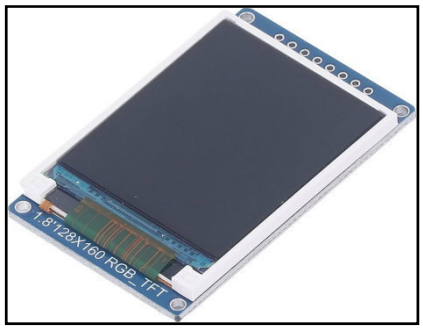


Fig. 6. 1.44 Inch TFT LCD 65K Color 128x128 Display

G. MAX1044 Voltage converter

The MAX1044 is a versatile CMOS voltage converter, primarily used for inverting DC voltage to generate a negative voltage.The MAX1044 operates with input voltages ranging from 1.5V to 10V and can provide a negative output voltage close to the magnitude of the input voltage. It features low quiescent current, low power consumption, and can operate at frequencies up to 10 kHz, making it suitable for battery-powered and low-power applications.

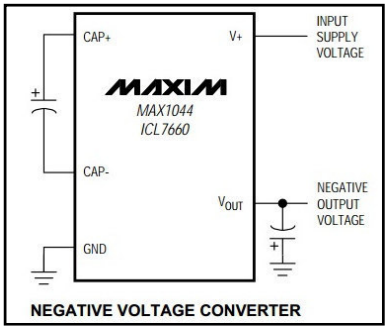


Fig. 7. MAX1044 Internal Circuit

H. Display stage

The display stage of the pH meter circuit involves the precise calculation and visualization ofpH values. Initially, the output voltage of the circuit, ranging from 0 to 3.3V, is connected to theADC of the ESP32 microcontroller. The ESP32 then processes this analog voltage, convertingit to digital data. Utilizing the appropriate formulas, such as the Nernst equation, the ESP32calculates the corresponding pH value. This calculated pH value is subsequently displayed on aTFT display interfaced with the ESP32, providing a clear and accurate digital readout of the pHmeasurement.

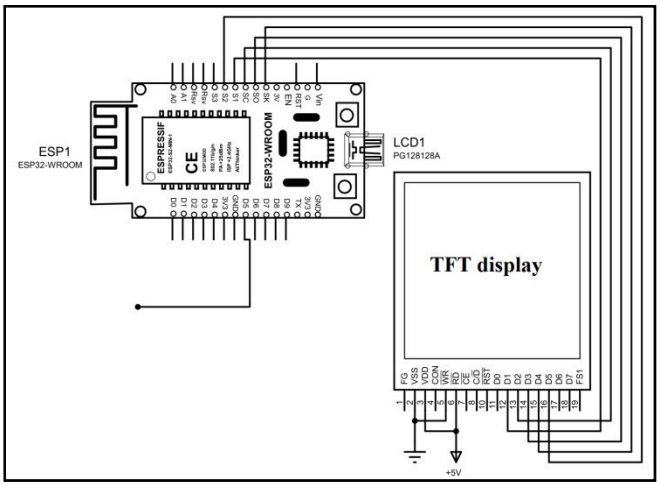


Fig. 8. Display stage

V. CIRCUIT IMPLEMENTATION AND TESTS

A. Implementation

First, the pH probe is placed securely to ensure stability, with its leads easily accessible for connection. The pH probe output is connected to the input of the buffer amplifier, whose output is then wired to the gain stage, and subsequently, the offset stage. The final output from the analog conditioning circuit is connected to one of the ADC pins of the ESP32. The TFT display module is connected to the appropriate GPIO pins on the ESP32 for data and control signals. All connections are double-checked for accuracy, ensuring there are no loose wires, and verifying that the power supply connections are correct.

B. Microcontroller Programming

Programming the ESP32 microcontroller involves writing code to read the ADC values, apply filtering algorithms, and calculate the pH value. The Arduino IDE is used to program the ESP32, with the board package installed and the ESP32 connected to the computer via USB. Necessary libraries for ADC reading and TFT display control are included. The code initializes the ADC pin and sets up the TFT display. The analogRead() function reads the voltage from the ADC pin, and a filtering algorithm, such as a moving average filter, is implemented to smooth out the readings. The Nernst equation is applied to convert the filtered voltage reading to a pH value, adjusted for calibration constants and temperature compensation if needed. The TFT library functions are used to display the calculated pH value on the screen. The code is uploaded to the ESP32, with serial output monitored to verify ADC readings and calculated pH values, ensuring they are as expected, and the pH value is displayed on the TFT screen.

C. Circuit Calibration rules

First we short circuit the input, this is essentially equivalent to having a perfect electrode in a perfect pH7 solution in a perfect world. we Measure the output of the gain stage and confirm that it remains at zero when the gain is modified with P1. Still with the input shorted, set the offset with P2 so that the output of offset stage is 2V. This setting was little difficult as we don't have very accurate material. but we can just set it around 2V,The question is simple: what should be shown by the display for the voltage found after offset stage ? Since 2V is pH 7 the relation is pH=V*7/2.

Given the voltage that we have after offset stage we can thus obtain what we want to see on the display. For example, if the voltage is 2.053V then the value that should be shown on the display is 7.186. Note that we will choose an output value of 3 to 3.5V to obtain a greater display value and thus a greater accuracy in this setting. Now we can just read the value displayed by the TFT.

D. Calibration with known-pH

The pH-meter must now be calibrated with known-pH solutions before it can be used. This should be done before each use. we have a neutral pH solution (pH=7) and a alkali of pH= 9. First we put the pH probe in the pH 7 solution and adjust the offset (P2) to reach pH 7 on the LCD display.

The gain potentiometer (P1) should not have any effect on the reading since any gain applied to a zero voltage will yield a zero output after the gain stage, knowing that a pH=7 corresponds to a probe signal of 0V, If we see a (significant) variation when changing P1 then it means that the pH of our solution is not exactly 7. Note also that the response of the pH meter strongly depends on the probe and the calibration should be performed with stable values only.

After calibrating the offset, we put the probe in the pH=9 solution (gently we must wipe the probe to avoid contamination of the solutions). we won't touch the offset potentiometer P2 anymore, instead we set the gain (P1) so that the display reads pH=9.

The calibration is finished! now we can just use some solutions with known pH values to test the functioning of the pH meter after the calibration process.

If we noticed a gain-variation with P1 during the offset calibration (done with P2) then we will have to repeat the calibration several times; it means that our pH 7 solution was not really pH 7.

This iterative process can diverge if we are not close to the solution. In fact, we could actually use a pH 4 and pH 9 solutions for the calibration, but this would obviously force an iterative calibration process and this is often difficult.

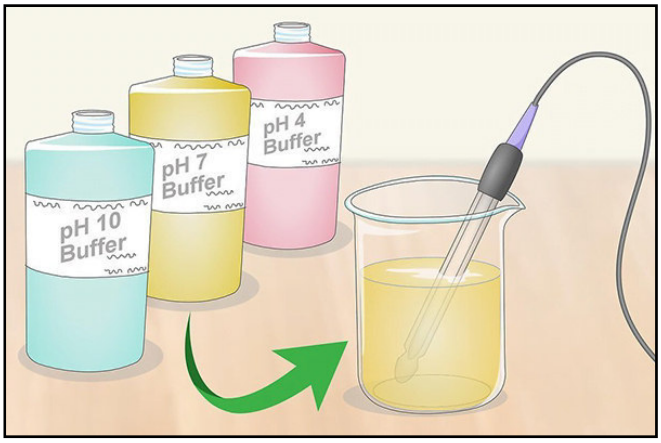


Fig. 9. The Calibration Process

E. Troubleshooting

During implementation, various issues may arise. Common problems include no reading or incorrect reading, which can be addressed by checking all connections on the breadboard to ensure they are secure, verifying the pH

probe's functionality, and ensuring the buffer amplifier provides the expected high input impedance.

To test the circuit without pH electrode we just replaced it with a dc voltage of the ranges -0.414V and 0.414V. Table II collect the 3 input voltages as well as the associate pH values recorded on the screen of the TFT display.

TABLE II. TESTS

Input voltage	*The output pH value*
220 mV	3.40
0 mV	7.13
-220 mV	11.14

It's worth noting that, in our case, the final pH value is not based on a single input measurement. Instead, we collect a specified number of voltage readings (approximately 5000) and calculate their average to obtain a more accurate final pH value.

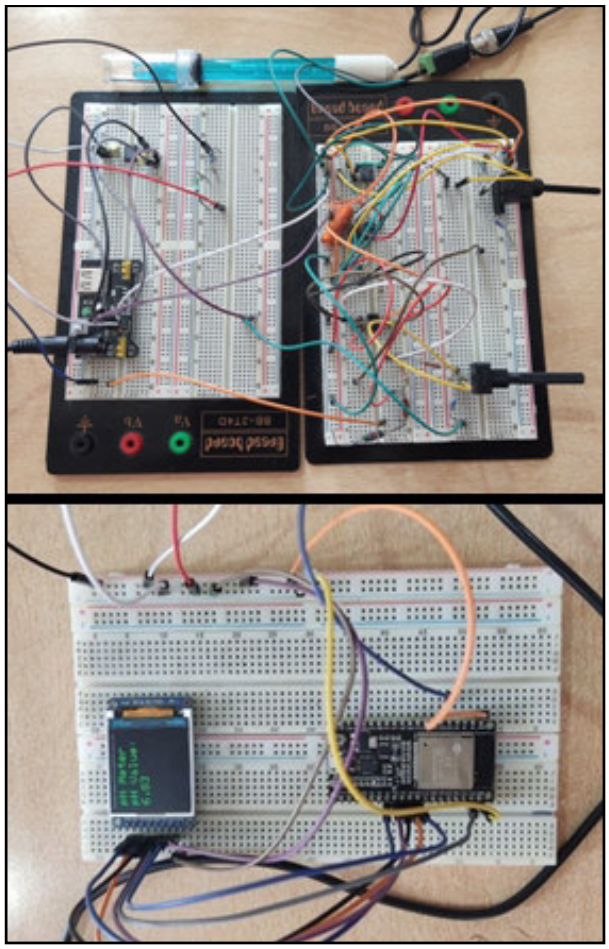


Fig. 10. Circuit implementation

After initial testing and troubleshooting, final adjustments are made. The filtering algorithm is enhanced to improve performance, and the display update routine is optimized to ensure smooth and accurate pH readout, with careful attention to refresh rates and data update intervals.

VI. Conclusion

The digital pH meter has become an indispensable tool in various fields due to its critical role in monitoring and controlling pH levels. We designed and implemented a digital pH meter, achieving highly satisfactory results. The device accurately measures the potential difference from a high impedance glass electrode, conditions the signal, and uses an ESP32 microcontroller to convert and display pH values on a TFT screen. Through meticulous design, using rigorous testing, the system demonstrated high reliability and precision in various pH environments. Although the current design lacks temperature compensation, which is essential for accurate pH measurement across different temperatures, this device successfully applied advanced concepts from analog and digital electronics, reinforcing our theoretical knowledge with practical application. The integration of a microcontroller highlights the modern approach in circuit design. Future enhancements could include adding temperature compensation to improve accuracy and extending the device's functionality to meet broader application requirements. This design not only showcased our ability to create and implement a complex digital system but also provided a solid foundation for further innovations in digital pH measurement.

References

[1] https://www.spectrascientific.com/blogs/ph-guides/ph-measurement-methods-ways-to-measure-ph (Accessed on June 2024)

[2] M.F. Sancho-Madriz , "preservation of food," Encyclopedia of Food Sciences and Nutrition (Second Edition), pp. 4766-4772, 2003.

[3] J.A. Tantray, S. Mansoor, R. F. Choh Wani and N. Un Nissa, "Chapter 3 - pH meter: Its use and calibration," A Laboratory Manual for Students and Researchers, pp. 9-10 , 2023.

[4] M. F. Camões, "A Century of pH Measurements," Chemistry International -- Newsmagazine for IUPAC, vol. 32, no. 2, pp. 3-7, 2010.

[5] J. N. Brönsted, "Einige Bemerkungen über den Begriff der Säuren und Basen, " [Some observations about the concept of acids and bases], Recueil des Travaux Chimiques des Pays-Bas. vol.42 (8), pp.718–728, 1923.

[6] J.John, B. C. Ashton and L. Geary, "The Effects of Temperature on pH Measurement," a reagecon technical paper – Technical Services Department.

[7] https://laboratorycognizance.com/2023/07/temperature-influence-on-ph/ (Accessed on June 2024)

[8] https://www.chemicals.co.uk/blog/what-is-neutralisation-in-chemistry (Accessed on June 2024)

[9] https://digital.sciencehistory.org/works/44558f07m (Accessed on June 2024)

[10] https://www.chemistrylearner.com/nernst-equation.html (Accessed on June 2024)

[11] https://www.espressif.com/en/products/socs/esp32 (Accessed on June 2024)

[12] https://sound-au.com/project95.htm (Accessed on June 2024)

Real-Time PCB Defect Detection on Android Devices Using YOLOv9-tiny: A Mobile Solution for Industrial Quality Control

Mohamed Achit
Departement Of Electrical Engineerig
Laboratory of Materials and Sustainable Development.
University of bouira
bouira, algeria
m.achit@univ-bouira.dz

Nacera Yassa
Departement Of Electrical Engineerig
Laboratory of Materials and Sustainable Development.
University of bouira
bouira, algeria
n.yassa@univ-bouira.dz

Ahcene Bouzida
Departement Of Electrical Engineerig
Laboratory of Materials and Sustainable Development.
University of bouira
bouira, algeria
a.bouzida@univ-bouira.dz

Ali Mazari
Departement Of Mecanical Engineerig
Laboratory of The Applied And Automation And Industial Diagnostic
University Of Djelfa
Djelfa, Algeria
ali.mazari@univ-djelfa.dz

Okba Fergani
Departement Of Electrical Engineerig
Laboratory of Identification, Command, Control and Communication
university of biskra
Biskra, Algeria
okba.fergani@univ-biskra.dz

Billel Naceri
Departement Of Electrical Engineerig
Laboratory of Identification, Command, Control and Communication
university of biskra
Biskra, Algeria
billel.naceri@univ-biskra.dz

***Abstract*— The detection of defects in printed circuit boards (PCBs) is critical in ensuring the reliability and quality of electronic products. Traditional inspection methods, including manual and electrical testing, often fall short in accuracy, efficiency, and adaptability to complex, multi-layered PCBs. In this study, we present a novel approach that leverages deep learning-based object detection, specifically using the optimized YOLOv9-tiny (YOLOv9t) model, to identify a wide range of PCB defects in real time. The model was trained on the Mixed Defect Detection in PCB (MDD_PCB) dataset, achieving a mean Average Precision (mAP) of 98.1% at 50% Intersection over Union (IoU) threshold, demonstrating exceptional accuracy. The result is a lightweight, mobile-compatible Android application, *PCB Inspector*, which provides an accessible, accurate, and real-time defect detection solution using only a standard smartphone camera. Performance evaluations highlight the model's reliability and suitability for industrial applications. This application represents a step forward in mobile-based PCB inspection, offering a practical, cost-effective tool for quality assurance in manufacturing environments.**

Keywords—Solar cell defects, object detection, RT-DETR, Electroluminescence (EL) imaging, Photovoltaic systems, Deep learning

I. INTRODUCTION

Printed circuit boards (PCBs) are vital components in electronic devices, providing both connectivity and structural support. They are essential across various electronic applications, extending well beyond basic models. The global PCB market, valued at $82 billion in 2022, is expected to grow to around $140.73 billion by 2032 [1]. However, as PCBs become smaller, detecting flaws presents significant challenges, making accurate defect detection during manufacturing crucial to enhancing product quality and reducing operational costs.

PCB defect detection methods include manual visual inspection, electrical testing, and optical inspection [2]. Manual inspection is limited by low accuracy and efficiency, while electrical testing requires complex circuitry and costly equipment and is less effective for multi-layered PCBs, sometimes causing secondary damage. Automated Optical Inspection (AOI) systems use cameras and image processing to identify issues like missing components and soldering defects in PCBs, providing a non-contact, machine vision-based method that is more accurate and faster than other approaches [3], [4]. Despite these advantages, AOI systems are sensitive to environmental conditions, slow, and less effective for complex defects [5].

In recent years, deep learning, specifically convolutional neural networks (CNNs), has become prominent in computer vision-based object detection. Detection methods in deep learning are divided into two primary approaches: two-stage and one-stage methods. Two-stage methods first generate candidate boxes for potential objects, then classify them using CNNs, with popular algorithms like RCNN and Faster R-CNN. Despite its accuracy, Faster R-CNN [6] has limited processing speed due to the region proposal network (RPN), which hinders its real-time application.

One-stage methods, such as SSD [7] and YOLO [8], bypass candidate box sampling by directly transforming object localization into a regression problem, achieving faster and more accurate results. Although SSD can struggle with detecting small objects due to low-resolution features, YOLO models, particularly YOLOv5 [9], are preferred for object detection because of their balanced speed, accuracy, and robustness[10, 11].

In the domain of PCB defect detection, Ancha et al. introduced the *Mixed Defect Detection in PCB* (MDD_PCB) dataset to enhance model training by including realistic combinations of defect types within single images. This dataset was evaluated using YOLO models, achieving a mean Average Precision (mAP) of 95%, along with high precision and recall metrics. Their work demonstrated the efficacy of

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

using YOLO models on embedded systems like the Jetson Nano, achieving efficient, real-time PCB defect detection, which is critical for practical applications in manufacturing environments [12]

To develop effective PCB defect detection models, comprehensive and diverse datasets are essential. However, existing datasets often focus on specific defect types and lack complete labeling. This research makes the following contributions:

- **Optimization of the YOLOv9 model for PCB defect detection**: Trained on the "Mixed Defect Detection in PCB" (MDD_PCB) dataset [12], the YOLOv9 [13] model achieved improved accuracy over previous models in detecting a variety of PCB defects.

- **Real-time implementation on Android**: The optimized YOLOv9 model was developed into a mobile application, allowing efficient, real-time defect detection using a standard Android phone camera, increasing accessibility for on-site use without specialized equipment.

This work demonstrates a practical, mobile-compatible framework for PCB defect detection, addressing key limitations in current detection methods and offering a reliable tool for industrial applications.

II. METHODOLOGY

A. *YOLOv9 Algorithm*

T The YOLOv9 algorithm integrates advanced features designed to minimize information loss during training while enhancing both detection efficiency and accuracy. In this study, we specifically used YOLOv9-tiny (YOLOv9t) due to its compact size, making it ideal for deployment on Android mobile applications where memory and processing power are limited.

One of the standout features in YOLOv9t is Programmable Gradient Information (PGI), which plays a crucial role in preserving critical data during the feedforward process. PGI also enhances the gradient flow, leading to improved model performance. As shown in Fig. 1, the PGI architecture incorporates bottleneck and reversible branches to efficiently manage information flow, reducing data loss. This architecture, applied to tasks such as defect detection in photovoltaic cells, significantly improves the model's precision and reliability.

Another key advancement in YOLOv9t is the Generalized Efficient Layer Aggregation Network (GELAN), illustrated in Fig. 2, which optimizes parameter usage through structured layer aggregation. This approach enables YOLOv9t to exceed the efficiency of earlier methods, such as depth-wise convolutions, by enhancing learning scalability. The GELAN architecture further supports YOLOv9t's adaptability across various hardware setups, including Android devices, where limited computational resources make efficient parameter utilization essential.

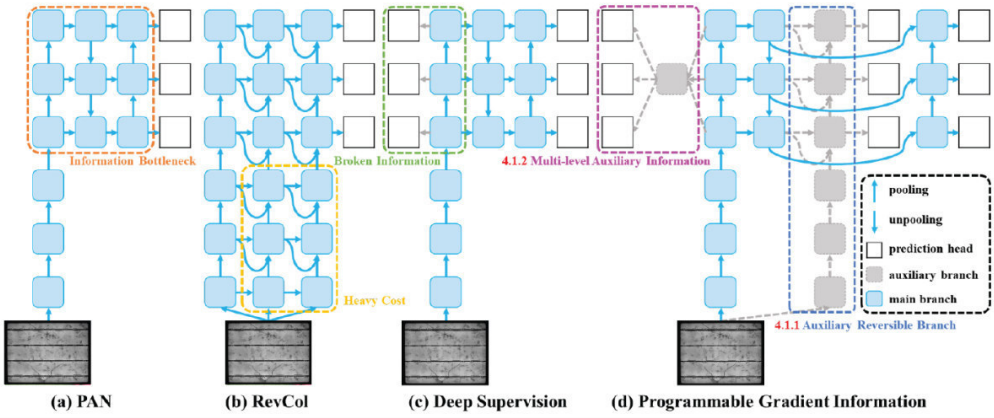


Fig. 1. YOLOv9 PGI (Programmable Gradient Information) Architecture for Enhanced Gradient Flow and Information Retention [14].

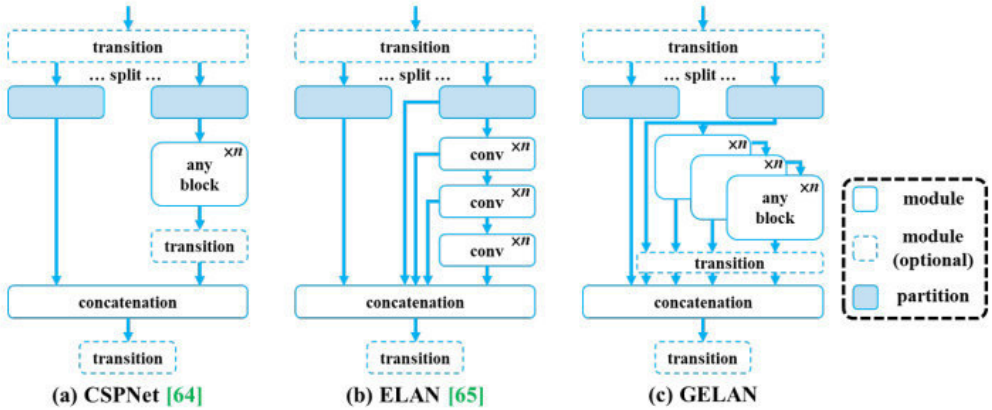


Fig. 2. GELAN (Generalized Efficient Layer Aggregation Network) Architecture within YOLOv9 for Optimized Parameter Utilization [14].

B. PCB dataset

The dataset used in this study is the "Mixed Defect Detection in PCB" (MDD_PCB) dataset, specifically crafted to address limitations found in other PCB defect datasets by providing a realistic variety of defect scenarios. The MDD_PCB dataset was designed to enhance model performance by including multiple types of defects in each image, allowing models to learn from diverse and representative examples of PCB faults commonly encountered in manufacturing.

The MDD_PCB dataset comprises 1,741 images, each with a resolution of 640 × 640 pixels, annotated with a total of 3,704 defects. These defects are distributed across six common PCB defect types, as summarized in Table 1 below. Each image includes two to three defects, facilitating a comprehensive training experience for deep learning models by exposing them to various defect combinations in a single samples.

TABLE I. DISTRIBUTION OF PCB DEFECTS IN THE MDD_PCB DATASET

Dataset (PCB Defect Type)	Number of Images	Number of Defects
Missing Hole	600	600
Mouse Bite	674	674
Open Circuit	847	847
Short Circuit	483	483
Spur	442	442
Spurious Copper	658	658
Total	1741	3704

C. Dataset Preprocessing and Augmentation

To prepare the MDD_PCB dataset for effective training, both preprocessing and augmentation techniques were applied to enhance model robustness and generalization.

1) Preprocessing

Each image in the dataset underwent the following preprocessing steps:

- **Auto-orientation**: The pixel data of each image was adjusted to ensure correct orientation, with EXIF-orientation metadata stripped from the images.
- **Resizing**: Each image was resized to a fixed resolution of 640 × 640 pixels to standardize the input dimensions for the model, using a stretch approach to maintain uniform sizing.

2) Data Augmentation

To further improve model performance by providing diverse training samples, a set of augmentation techniques was applied to create three augmented versions of each original image, resulting in a total of 6,964 images (1,741 original images + 5,223 augmented images). The following transformations were randomly applied to each image:

- **Horizontal Flip**: Each image had a 50% probability of being flipped horizontally.
- **Vertical Flip**: Each image had a 50% probability of being flipped vertically.
- **90-Degree Rotations**: Each image was randomly rotated with an equal probability of one of the following orientations: no rotation, 90° clockwise, 90° counterclockwise, or 180° (upside-down).
- **Salt and Pepper Noise**: Random noise was added to 5% of the pixels in each image to simulate image imperfections and enhance the model's resilience to noise.

These augmentation techniques expanded the dataset significantly and provided the YOLOv9 model with a wider variety of examples, improving its ability to detect PCB defects across various conditions and orientations in real-world scenarios.

D. Evaluation Metrics

To evaluate the YOLOv9t model's performance, we used key metrics, namely Precision, Recall, and mean Average Precision (mAP). These metrics are essential in object detection tasks, offering insights into both the accuracy of predictions and the model's ability to recognize relevant instances across various classes.

Precision quantifies the proportion of correct positive predictions among all positive predictions made by the model. It reflects the model's capability to avoid false positives:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

where TP represents true positives, and FP denotes false positives. A high precision score indicates that most detected objects are indeed relevant.

Recall evaluates the model's ability to identify all relevant instances by measuring the proportion of actual positives that were correctly detected:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

where FN refers to false negatives. A high recall score implies that the model can detect a majority of the relevant objects.

Mean Average Precision (mAP) provides a comprehensive measure of the model's performance across all classes by averaging precision over different recall thresholds. It is calculated as the area under the Precision-Recall curve:

$$\text{mAP} = \frac{\sum_{i=1}^{N} APi}{N} \quad (3)$$

where APi is the Average Precision for the i-th class, and N represents the total number of classes. This metric provides a holistic view of the model's accuracy in detecting objects across all categories.

E. Experimental Conditions

All experiments in this study were conducted on the Kaggle platform, which provides cloud-based Jupyter Notebooks (Kernels) with free access to high-performance GPUs, including the Tesla T4 with 15,095 MiB of memory. The software environment was configured with Python 3.10.13 and PyTorch 2.1.2, leveraging the GPU's parallel processing power for efficient computation.

F. Training configuration and hyperparameters

In this study, the YOLOv9t model was trained using the following configuration and key hyperparameters:

- **Batch size**: 32
- **Epochs**: 200
- **Image size**: 640
- **Optimizer**: auto
- **Momentum**: 0.937
- **Weight decay**: 0.0005

The hyperparameters were meticulously chosen to maximize the model's performance throughout training and validation. The batch size and number of epochs were selected to maintain a balance between computational efficiency and accuracy, while the learning rate and momentum were adjusted to promote stable and effective convergence.

III. RESULTS AND DISCUSSION

The performance of the YOLOv9t model on the MDD_PCB dataset was evaluated using key object detection metrics, including Precision, Recall, mean Average Precision at 50% Intersection over Union (mAP@50), and mean Average Precision across IoU thresholds (mAP@50-95). These metrics provide a comprehensive view of the model's detection accuracy and reliability in identifying PCB defects.

A. Evaluation Metrics

1) Precision and Recall:

- The Precision and Recall metrics showed consistent improvement across the 200 training epochs. As seen in the precision and recall plots (Fig. 3, top-left and top-right), the model achieved high levels of precision and recall, indicating its strong ability to correctly identify relevant defects while minimizing false positives.

- Precision stabilized above 0.9, demonstrating the model's effectiveness in detecting true positives with minimal error, which is crucial for practical applications in PCB defect detection.

- Recall also stabilized around 0.9, reflecting the model's ability to capture almost all relevant instances, ensuring comprehensive defect identification.

2) Mean Average Precision (mAP):

- The model's mAP@50 and mAP@50-95 metrics indicate its high accuracy across varying IoU thresholds. The mAP@50 metric, which reflects precision at a single IoU threshold, reached close to 1.0, indicating near-perfect detection performance in identifying PCB defects.

- The mAP@50-95 metric (Fig. 3, bottom-left and bottom-right), which averages precision across multiple IoU thresholds, reached around 0.6, showing that the model maintains strong performance across diverse detection scenarios, even when stricter IoU thresholds are applied.

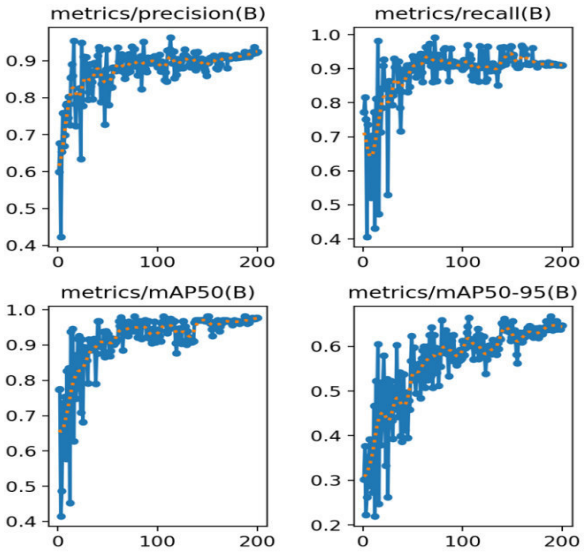


Fig. 3. YOLOv9t Model Performance Metrics on the MDD_PCB Dataset.

B. Training and Validation Loss

1) Box Loss:

The box loss plots (Fig. 4, top-left and bottom-left) demonstrate a steady decrease over the course of training, both in the training and validation phases. This reduction in box loss indicates that the model consistently improved in accurately predicting bounding box coordinates for detected defects. By the end of training, box loss was minimized, reflecting the model's ability to closely match ground truth locations.

2) Classification Loss:

The classification loss for both training and validation (Fig. 4, top-right and bottom-right) also decreased substantially throughout training. This metric's steady decline shows that the model became increasingly effective in classifying PCB defects into their respective categories with high accuracy.

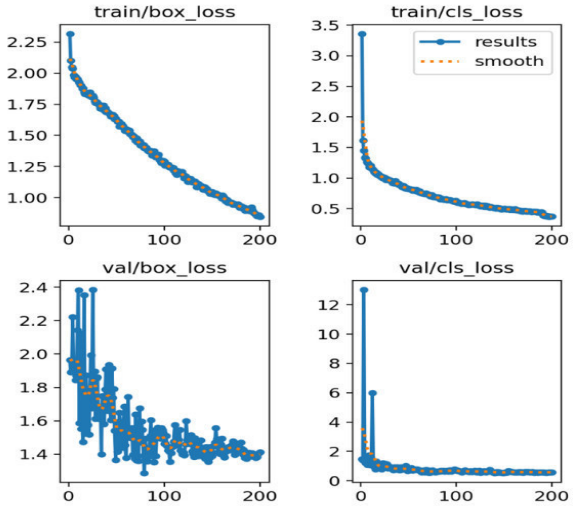


Fig. 4. Training and Validation Loss Curves for YOLOv9t Model.

C. *Visual Representation of YOLOv9t Outputs*

Figure 5 shows the output of the YOLOv9t model applied to the MDD_PCB dataset, highlighting its ability to accurately detect multiple types of PCB defects within single images. Each defect, such as short circuit, mouse bite, spur, spurious copper, open circuit, and missing hole, is labeled with a bounding box and a confidence score, indicating the model's certainty in its predictions.

This visual illustrates the effectiveness of YOLOv9t in distinguishing between various defect types, with high confidence scores providing assurance in detection reliability. The model's ability to detect and label multiple defects within each PCB sample demonstrates its potential for real-time, comprehensive quality inspection in manufacturing environments, where rapid and accurate defect identification is essential.

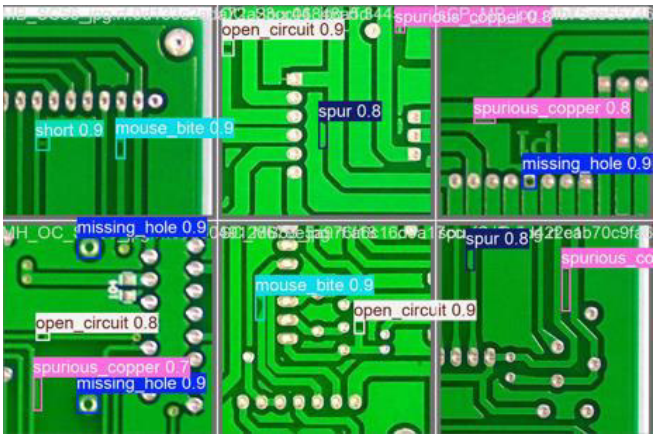


Fig. 5. Detection of PCB Defects Using YOLOv9t Model on the MDD_PCB Dataset.

D. *Comparison of YOLOv8n, YOLOv10n, and YOLOv9t*

This section compares the YOLOv8n, YOLOv10n, and YOLOv9t models on the MDD_PCB dataset. We focus on Precision, Recall, mAP@50, mAP@50-95, and inference speed.

TABLE II. PERFORMANCE COMPARISON OF YOLOV8N, YOLOV10N, AND YOLOV9T MODELS ON THE MDD_PCB DATASET.

Model	Layers	Params	GFLOPs	Precision	Recall	mAP@50	mAP@50-95	Speed (img/s)
YOLOv8n	238	9.4M	21.3	0.867	0.984	0.967	0.662	8.66
YOLOv10n	402	8.07M	24.8	0.887	0.881	0.926	0.669	10.52
YOLOv9t	917	7.29M	27.4	0.92	0.912	0.981	0.667	9.41

- **YOLOv9t** is the most suitable model, achieving the highest precision (0.92) and mAP@50 (0.981), making it ideal for applications requiring accurate defect detection.
- **YOLOv10n** offers the fastest inference speed (10.52 img/s) with competitive accuracy, suitable for real-time applications with moderate precision needs.
- **YOLOv8n** excels in recall (0.984), capturing more instances but with lower precision, which may lead to more false positives.

In summary, YOLOv9t stands out as the optimal model due to its balance of high precision and overall detection accuracy, making it well-suited for PCB defect detection tasks.

E. *Development of an Android Application for PCB Defect Detection*

In the pursuit of practical, accessible solutions for real-time quality inspection in PCB manufacturing, we developed an Android-based application named PCB Inspector. This application utilizes the YOLOv9-tiny (YOLOv9t) model, optimized for mobile deployment. With an interface designed for ease of use, the application can identify common PCB defects, including missing holes, mouse bites, open circuits, short circuits, spurious copper, and spurs, using the phone's camera to scan and detect defects within milliseconds.

Application Features and Performance

As shown in Figure 6, the PCB Inspector app can accurately detect various PCB defect types. Each defect is labeled with a bounding box and identified by its defect type, making it easier for operators to quickly recognize and address quality issues. This composite figure illustrates the application's performance in detecting defects in real-time, even under varied lighting and orientation conditions.

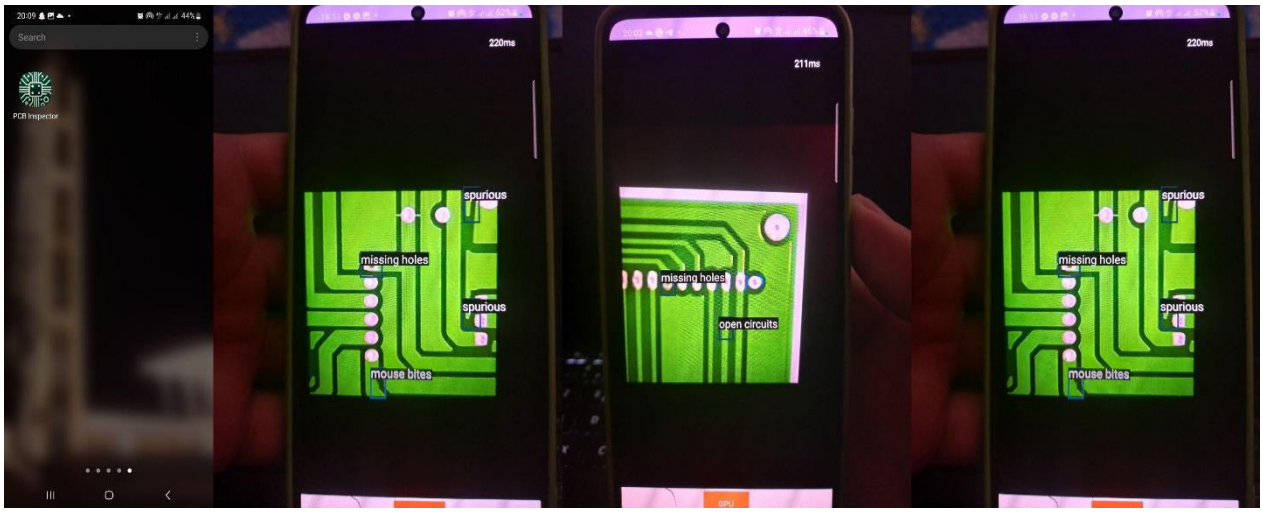


Fig. 6. Visual representation of the PCB Inspector app detecting multiple PCB defects.

Conclusion

The PCB Inspector app, powered by the YOLOv9-tiny model, offers a practical, mobile-based solution for real-time PCB defect detection, achieving an impressive mean Average Precision (mAP) of 98.1% on the MDD_PCB dataset. This lightweight application provides manufacturers with an accessible, cost-effective tool for accurate defect identification without specialized equipment. Its high precision and recall underscore its potential for quality control in various manufacturing environments. Future enhancements may include cloud integration for data storage and expanded defect detection capabilities, further positioning PCB Inspector as a versatile and efficient tool for industrial quality assurance.

REFERENCES

1. Marques, A.C., J.-M. Cabrera, and C. de Fraga Malfatti, Printed circuit boards: a review on the perspective of sustainability. Journal of environmental management, 2013. 131: p. 298-306.
2. Deng, L., Research on PCB surface assembly defect detection method based on machine vision. Wuhan University of technology, 2019.
3. Yun, Z., L. Zhi-gang, and Z. Yu-qiang, Research progress and prospect of machine vision technology. Journal of graphics, 2020. 41(6): p. 871.
4. Khalid, N.K., Z. Ibrahim, and M.S.Z. Abidin, An algorithm to group defects on printed circuit board for automated visual inspection. Int. J. Simul. Syst. Sci. Technol, 2008. 9(2): p. 1-10.
5. Ling, Q. and N.A.M. Isa, Printed circuit board defect detection methods based on image processing, machine learning and deep learning: A survey. IEEE Access, 2023. 11: p. 15921-15944.
6. Ren, S., et al., Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence, 2016. 39(6): p. 1137-1149.
7. Liu, W., et al. Ssd: Single shot multibox detector. in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. 2016. Springer.
8. Redmon, J. You only look once: Unified, real-time object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
9. Parambil, M.M.A., et al., Navigating the YOLO landscape: A comparative study of Object detection models for Emotion Recognition. IEEE Access, 2024.
10. Dewi, C., et al., Deep convolutional neural network for enhancing traffic sign recognition developed on Yolo V4. Multimedia Tools and Applications, 2022. 81(26): p. 37821-37845.
11. Zhao, J., et al., Detection of passenger flow on and off buses based on video images and YOLO algorithm. Multimedia Tools and Applications, 2022. 81(4): p. 4669-4692.
12. Ancha, V.K., et al., Utilizing YOLO Models for Real-World Scenarios: Assessing Novel Mixed Defect Detection Dataset in PCBs. IEEE Access, 2024.
13. Wang, C.-Y., I.-H. Yeh, and H.-Y. Mark Liao. Yolov9: Learning what you want to learn using programmable gradient information. in European Conference on Computer Vision. 2025. Springer.

Design and characterisation of dipole photoconductive antenna for Terahertz multiband applications

Hamza Zeraoula
Laboratoire Systèmes Lasers
Ecole Militaire Polytechnique
Bordj El-Bahri, Algiers, Algeria
d_zeraoula.hamza@emp.mdn.dz

Chellali Slimani
Ecole Supérieure de la Defense Aerienne du Territoire
Algiers, Algeria
slimanichelali@gmail.com

HadjAbdelkader Benzater
Ecole Supérieure de la Defense Aerienne du Territoire
Algiers, Algeria
BENZATER.hadj@gmail.com

***Abstract*— Photoconductive antennas are considered as fundamental components in the terahertz technology, offering wideband capabilities essential for diverse applications including imaging, spectroscopy, security, and communications. PCAs operate based on unique principles that combine the generation and radiation of terahertz waves. This dual functionality necessitates a distinct approach to their design and performance evaluation. In this study, we focus on the design and characterization of dipole PCAs tailored for broadband and multiband applications. We will explore the theoretical foundations of PCA operation, emphasizing how variations in semiconductor substrate thickness influence performance metrics such as impedance matching, bandwidth, directivity, and gain. By fixing key design parameters while varying others, we aim to provide insights into optimizing PCA configurations for enhanced operational efficiency across multiple frequency bands. The findings contribute to optimizing PCA designs for enhanced performance in broadband applications.**

Keywords—Photoconductive antenna, Terahertz, Dipole antenna, impedance matching.

I. INTRODUCTION

Terahertz (THz) frequency band is lying between microwave and optical regions, from 0.3 THz to 10 THz, and considered as a part of sub-millimeter waves [1]. Recently, there has been an increase in emergent research on THz applications, such as communications [2], bioimaging [3], spectroscopy, sensing, and security. Some of the important motivations for the increasing interest in the THz frequency band are the attractive and unique characterizing features: THz radiations are highly penetrating through covering materials, and do not damage the scanned items under THz radiations, furthermore, the High spectral resolution [4].

Among the innovative components developed for THz applications are photoconductive antennas (PCAs), which leverage the unique properties of semiconductor materials to generate and detect THz radiation efficiently. Unlike conventional antennas that operate primarily within microwave or radio frequency ranges, PCAs can achieve broad bandwidths due to their photoconductive nature. THz photoconductive antennas (PCAs) are critical components for THz systems to radiate and detect waves, and their performances directly influence the quality of the entire system. Thus, THz PCAs have more performance requirements [5]. The first PCA that was used to generate electromagnetic transients on a pico-second time scale was the Auston switch [6]. The PCA consists of a DC-biased metal dipole antenna deposited on a photoconductive substrate, the exciting optical pulse is focused on the antenna gap, and propagates into the photoconductor, which begins to generate photo carriers. These latter are accelerated in the DC bias field, producing a transient photocurrent, which drives the dipole antenna, and then, a THz frequency pulse is radiated [7], PCA composition is depicted in Fig. 1. The biggest challenge for the wide use of THz devices is improving PCA performance, through better antenna design models. Extensive research, both experimental and theoretical, has focused on optimizing PCA in THz devices.

The design and characterization of dipole photoconductive antennas (PCA) for broadband and multiband applications have attracted significant interest due to their superior bandwidth and versatility. In our work, we present a comprehensive study on the design and characterization of dipole PCAs, focusing on their theoretical modeling, impedance matching, and radiation performance. By fixing PCA antenna parameters related to the exciting femtosecond laser, semiconductor parameters, dipole dimensions, and varying the thickness of the low-temperature gallium arsenide (LT-GaAs) substrate, we investigate the influence of the substrate thickness on antenna bandwidth, and gain. The results include detailed analyses of impedance matching through S11 and voltage standing wave ratio (VSWR) parameters, as well as evaluations of antenna gain across multiple frequency bands.

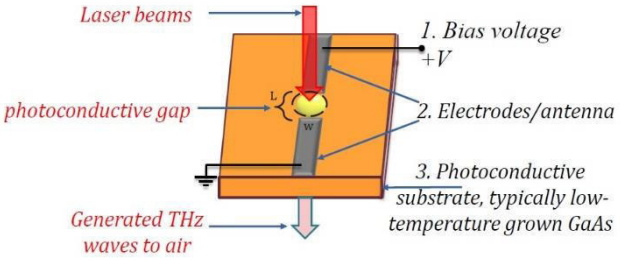


Fig. 1. Dipole photoconductive antenna composition

II. DIPOLE PCA THEORETICAL MODELING

Distinction in photoconductive antenna gap size imposes different theories to explain the radiation mechanism. For the small gap PCA, the equivalent circuit model (ECM) can describe well the obtained experimental results, which explains mainly the saturation phenomenon due to the screening effect induced by both radiation screening and space charge (polarization) screening [8-10]. In the ECM principle, the photoconductive dipole antenna is represented by lumped

elements: a combination of voltage or current sources with time-varying resistance and antenna impedance [11-14], as depicted in Fig. 2. ECM approach allows the use of the conventional antenna theory and concepts. In addition, it provides the ability of the matching study to improve the matching efficiency. Furthermore, the calculation of the generated photocurrent permits to be used as input for full wave simulation.

The required formula to calculate the considered parameters are illustrated in the following:

Based on the continuity equation, the time-dependent carrier density is expressed by (1):

$$\frac{dn(t)}{dt} = -\frac{n(t)}{\tau_c} + g(t) \quad (1)$$

where $n(t)$ is the carrier density, τ_c is the carrier lifetime, and $g(t)$ is the photo-carrier generation rate, by solving (1), we get [11]:

$$n(t) = P_0 \left(1 - exp\left(-\frac{2r^2}{w_0^2}\right)\right) \frac{\eta}{\sqrt{2\pi} w_0^2 dh\upsilon} T_l \, exp\left(\frac{T_l^2}{8\tau_c^2} - \frac{t}{\tau_c}\right) \left(erf\left(\frac{\sqrt{2}t}{T_l} - \frac{\sqrt{2}T_l}{2\tau_c}\right) + 1\right) \quad (2)$$

with P_0 is the laser peak power, W_0 is the laser's beam waist radius, T_l is the laser pulse duration, d is the depth of the excitation region, h being the Planck constant, υ is laser frequency, and η is the quantum efficiency defined by:

$$\eta = (1 - R)(1 - e^{-\alpha.d}) \quad (3)$$

with R representing the power reflection coefficient, and α *is* the optical absorption coefficient.

The conductivity formula is given by:

$$\sigma(t) = e.\mu_e.n(t) \quad (4)$$

where μ_e represents the electron mobility.

Equation (5) provides conductance within the gap during laser excitation for dipole PCA:

$$G(t) = \sigma(t).d.\frac{W_g}{L_g} \quad (5)$$

with W_g and L_g are the width and the length of the gap, respectively.

According to ECM in [11], the total current flowing through the gap for the dipole PCA is given by (6).

$$I_{total}(t) = G(t)V_c(t) + I_{cap}(t) + \frac{V_{bias}}{R_{dark}} \quad (6)$$

with $V_c(t)$ is the resultant voltage on the gap, R_{dark} being the dark resistance. While $I_{cap}(t)$ determines the capacitor current, this is caused by carriers staying near the positive electrode in Δt time step [11], and calculated using (7).

$$I_{cap}(t) = \frac{\pi e \mu_e^2 \Delta t W_g n(t) V_c^2(t)}{4L_g^2} \quad (7)$$

The total generated current flowing through antenna electrodes yields a radiation voltage, specified by (8)

$$V_{rad}(t) = Z_a.I_{total}(t) \quad (8)$$

with Z_a is the antenna impedance. Moreover, the remaining carriers within the gap generate a screening voltage, as outlined by (9).

$$\frac{dV_{sc}(t)}{dt} = -\frac{V_{sc}(t)}{\tau_r} + \frac{\sigma(t)}{\zeta\varepsilon} V_c(t) \quad (9)$$

where V_{sc} is the screening voltage, τ_r is the carrier recombination time, ζ is the screening factor, and ε is the medium permittivity.

Applying Kirchhoff's voltage law, we find:

$$V_c(t) = V_{bias} - V_{sc}(t) - V_{rad}(t) \quad (10)$$

where V_{bias} is the externally applied bias.

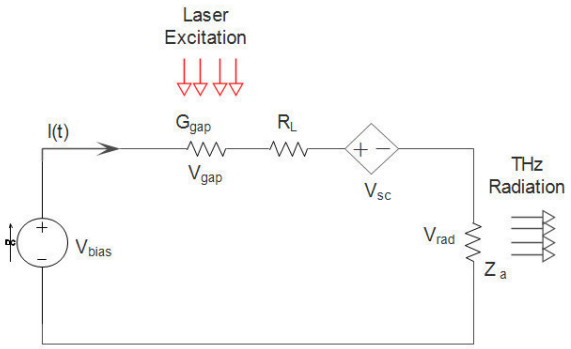


Fig. 2. *Schematic of the small gap PCA equivalent circuit model [14]*

III. Antenna Design and Optimization

The principle of the designed dipole PCA with specific parameters, is the use of LT-GaAs as substrate, and with its high relative permittivity, dielectric properties influence radiation patterns and wave propagation, effectively guiding electromagnetic waves through the substrate. By systematically varying the substrate thickness and analyzing the resulting parameters, we optimize antenna structure according our requirements.

In the proposed approach, we take profit initially of the equivalent circuit modeling to evaluate the conductance within the gap in terms of different parameters (5). Table I indicates the considered parameters related to the exciting laser source and the semiconductor substrate. For electron mobility computation, we have considered two models: the Caughey–Thomas mobility model [15], and the Arora mobility model [16]. Then, we use the obtained results to evaluate THz radiations performance using full-wave simulation tool HFSS (High Frequency Structure Simulator) in terms of S11, antenna gain, and radiation pattern.

The proposed approach steps are outlined as follows:

1. Use analytical models to derive initial parameters

Calculate the conductance within the gap, which is considered as source conductance in the full-wave simulation.

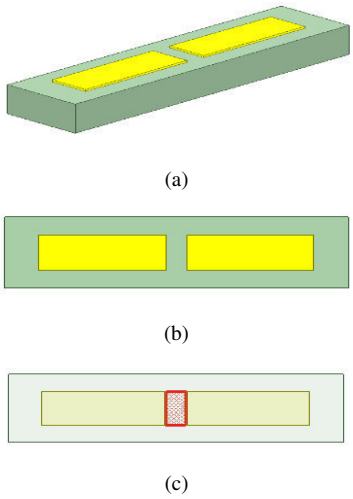
2. **Use Full-wave simulator to evaluate the design**

a) Design the antenna structure using HFSS, as shown in Fig. 3a.

b) Select the exposed surface within the gap to the laser excitation as a lumped port to excite the antenna structure, as illustrated in Fig. 3c

c) Use the conductance results from the analytical model to set the source impedance in the simulation.

3. **Vary semiconductor thickness (use optimetrics option in software)**
4. **Adjust design parameters to optimize performance, and reiterate simulations**



(a) (b) (c)

Fig. 3. *Dipole PCA Design under HFSS*

(a) Dipole PCA structure, (b) Top view of the Dipole PCA, (c) Excitation of Dipole PCA

TABLE I. DIPOLE PCA SIMULATION PARAMETERS

Parameter	Value
Laser wavelength (λ)	800 *nm*
Laser repetition rate (f_{rep})	80 *MHz*
Laser pulse duration (T_l)	100 *fs*
Reflection coefficient for Lt-GaAs (R)	0.318
Optical absorption coefficient (α)	6000 cm^{-1}
Carrier trapping time (τ_c)	1 *ps*
Carrier recombination time (τ_r)	100 *ps*
Screening factor (ζ)	900
Bias voltage (V_b)	30 *V*
Gap length (L)	5 μm
Gap width (W)	10 μm
Antenna total length	65 um
Depth of excitation region (d)	$1\mu m$

IV. RESULTS AND CHARACTERIZATION ANALYSIS

To analyze the results of the proposed approach, we start by Simulating the performance of a planar half-wave dipole antenna without substrate. Simulation results shown in Fig. 4 approve that the resonance frequency and the radiation pattern meet well the theoretical parameters.

In the next step, we simulate the dipole PCA performance with substrate thickness varying from 10 to 200 um, with a step of 10 um. In Fig. 5, to make sure that the display is clear, our presentation includes S11 results from the dipole PCA simulation with substrate thicknesses 10, 70, 130, and 190. It is apparent that the dipole PCA with a substrate thickness of 130 um provides the most effective performance compromise of the operating bands number and 10 dB bandwidth for each region.

Figs. 6 depict the analysis and the characterization of the designed dipole PCA with substrate thickness of 130 um. In Fig. 6a, we show designed antenna impedance for the frequency band from 300 GHz to 2 THz, while S11 parameter is displayed in Fig. 6b. We can see that the designed dipole PCA has 6 operating bands, and Fig. 6d shows the limits of each 10 dB bandwidth. Furthermore, Table II presents the 10 dB bandwidth for each operating band.

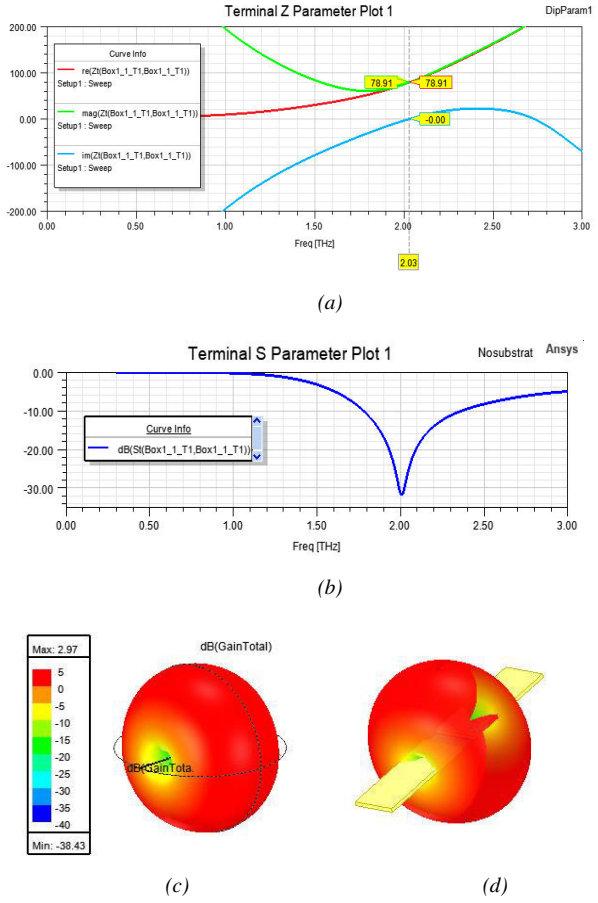


Fig. 4. *Planar half-wave dipole antenna (no substrate) (a) Impedance parameters, (b) S11 parameter, (c) Diagramm pattern, and (d) Diagramm pattern on antenna structure*

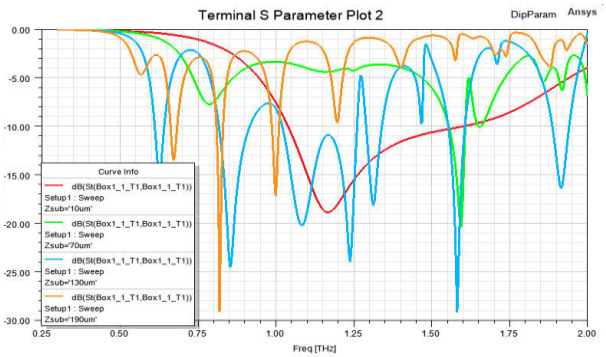
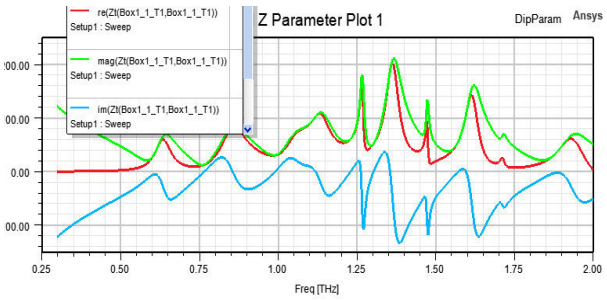
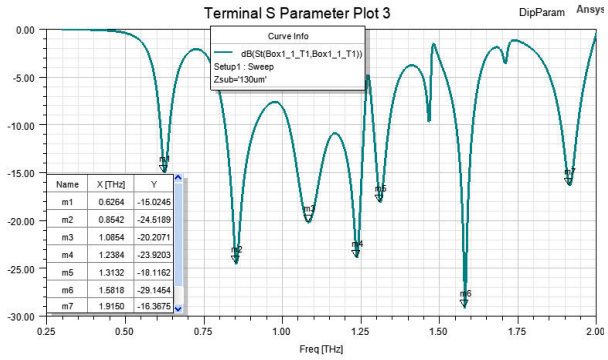


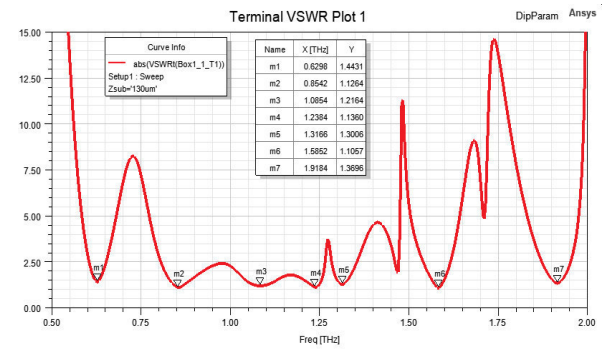
Fig. 5. *S11 Parameter for Dipole PCA with different substrate thickness values*



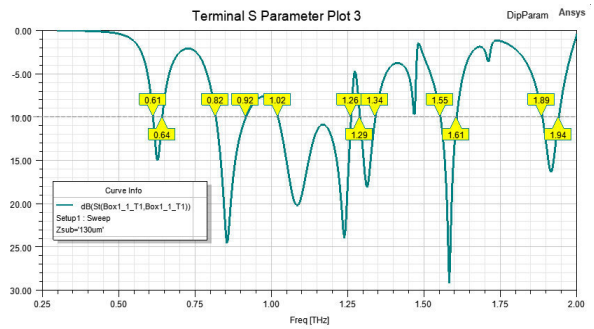
(a)



(b)



(c)



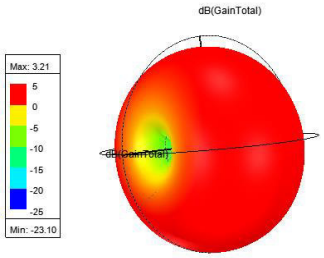
(d)

Fig. 6. *Parameters of the designed Dipole PCA with a Substrate Thickness of 130 um*
(a) Antenna impedance, (b) S11 parameter,
(c) VSWR Parameter, and (d)10 dB Bandwidth limitations

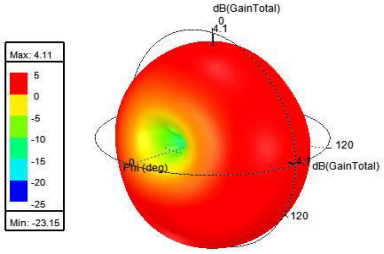
TABLE II. DESIGNED DIPOLE PCA OPERATING BANDS

Band number	Band limits	10 dB Bandwidth (GHz)
1.	610-640	30
2.	820-920	100
3.	1020-1260	240
4.	1290-1340	50
5.	1550-1610	60
6.	1890-1940	50

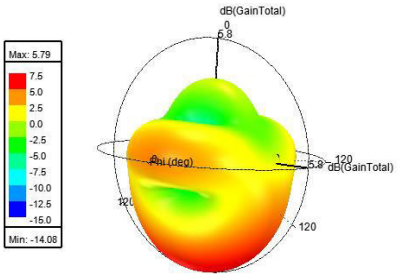
In the last part, we evaluate the radiation pattern for the resonant frequencies of the operating bands to characterize antenna's directivity, and gain. In Fig. 7a, and 7b, we notice that the radiation pattern is slightly focused toward the substrate, because the resonant frequency is close to the effective frequency that equals 750 GHz. On the other hand, we can notice clearly the influence of the substrate on the radiation pattern in Fig. 7c, 7d, and 7e, where antenna gain attends 9.36 dB for the frequency 1200 GHz. Then, more the frequency increases, antenna gain reduces, because of the radiation in other directions as illustrated in Fig. 7F, and 7g.



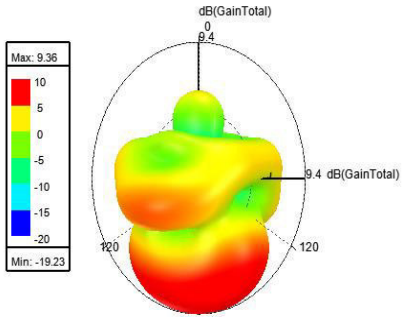
(a)



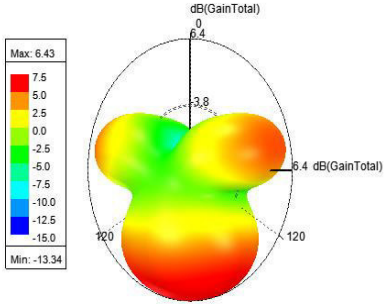
(b)



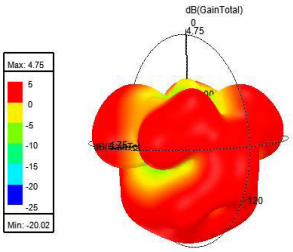
(c)



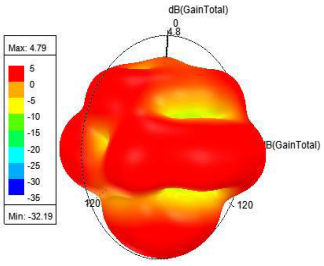
(d)



(e)



(f)



(g)

Fig. 7. *Diagramm patterns of the resonante frequencies for each operating band,*
(a) 625 GHz, (b)860 GHz, (c) 1080 GHz, (d) 1200 GHz, (e) 1310 GHz, (f) 1580 GHz, and (g) 1915 GHz

V. CONCLUSION

In conclusion, this work presents a comprehensive framework for the design and characterization of dipole photoconductive antennas tailored for broadband and multiband applications. By exploring the theoretical modeling, simulation analysis, and performance evaluation, we demonstrated how variations in semiconductor substrate thickness impact key parameters such as impedance matching, bandwidth, and radiation pattern. Our analysis includes detailed evaluations of impedance matching, band characteristics, and radiation performance, highlighting the potential of PCAs in advanced applications such as imaging, spectroscopy, security, and communications. The proposed design framework enhances our understanding of PCA behavior, and also provides a foundation for optimizing antenna performance based on specific application requirements, thereby advancing the capabilities of photoconductive antenna technology.

REFERENCES

[1] H. Chen, H. Sarieddeen, T. Ballal, H. Wymeersch, M.S. Alouini, and T.Y. Al-Naffouri, "A Tutorial on Terahertz-Band Localization for 6G Communication Systems," IEEE Communications Surveys & Tutorials 24(3), 1780–1815 (2022)

[2] M. Rahaman, et al "THz communication technology in India present and future,". Defense Sci. J. *69*(5), 510 – 516 (2019)

[3] B. Appasani, "A hybrid terahertz metamaterial sensor using a hexagonal ring resonator with bio-medical applications," Plasmonics. 17(2), 519–524 (2022)

[4] V. Anitha, A. Beohar, A. Nella, "THz imaging technology trends and a wide variety of applications: a detailed survey," Plasmonics 1, 1– 43 (2023)

[5] Y. He, Y. Chen, L. Zhang, S.W. Wong, and Z.N. Chen, "An overview of terahertz antennas," China Communications 17 (7), 124–165 (2020)

[6] D.H. Auston, K.P. Cheung, and P.R Smith, "Picosecond photoconducting Hertzian dipoles," Appl. Phys. Lett. 45(3), 284–286 (1984)

[7] N.M. Burford, and M.O. El-Shenawee, "Review of terahertz photoconductive antenna technology," Opt. Engineering 56(1), 010901 (2017)

[8] G.C. Loata, "Investigation of low-temperature-grown GaAs photoconductive antennae for continuous-wave and pulsed terahertz generation," Ph.D. Thesis, Department of Physics, Goethe-University, Frankfurt, Germany (2007).

[9] K. Moon, H. Han, and I. Park, "Terahertz folded half-wavelength dipole antenna for high output power", Proceeding of the IEEE International Topical Meeting on Microwave Photonics (MWP 2005), Seoul, Korea, 1– 4 (2005).

[10] H. Zeraoula, D. Benyahia, M. Lazoul, C. Slimani, and Y.Hadji, "Design and performance analysis of terahertz photoconductive antenna," International Telecommunications Conference (ITC-Egypt), pp. 152–157. IEEE, Cairo, Egypt (2024).

[11] J. Prajapati, M. Bharadwaj, A. Chatterjee, and R. Bhattacharjee, "Circuit modeling and performance analysis of Photoconductive Antenna," Optics Communications. **394**, 69-79, (2017).

[12] N. Khiabani, 'Modelling, design and characterization of Terahertz photoconductive antenna', PhD thesis, Department of Electrical and Electronics Engineering, University of Liverpool, (2013)

[13] E. N. F. Boby, J. Prajapati, V. Rathinasamy, S. Mukherjee, and S. Mondal, "Parametric Investigation of Interdigitated Photoconductive Antenna for Efficient Terahertz Applications," Arab. J Sci. Eng. **47**, 3597–3609 (2022).

[14] H. Zeraoula, D. Benyahia, and M. Lazoul, 'Proposing a refined and simplified equivalent circuit model for precise THz photoconductive antenna analysis,' Opt Quant Electron. **56**, 1055 (2024).

[15] D. M. Caughey, and R. E. Thomas, "Carrier mobilities in silicon empirically related to doping and field," Proceedings of the IEEE, **55(12)**, 2192-2193(1967).

[16] N. D. Arora, J. R. Hauser, and D. J. Roulston, "Electron and hole mobilities in silicon as a function of concentration and temperature," *IEEE Transactions on Electron Devices*, 29(2), 292-295 (1982).

Ada-Boost for Delineation of Raining Cloud over Mediterranean Sea.

BENSAFI Noureddine [1], LAZRI Mourad [2] and AMEUR Soltane [3]

[1] Laboratoire LAMPA, Université UMMTO, Tizi Ouzou, Algeria
MCB, Université UMBB, Boumerdès, Algeria
n.bensafi@univ-boumerdes.dz
[2] Laboratoire LAMPA, Université UMMTO, Tizi Ouzou, Algeria
m_lazri@yahoo.fr
[3] Laboratoire LAMPA, Université UMMTO, Tizi Ouzou, Algeria
s_ameur@yahoo.fr

Abstract. In this paper, we presented the application of the classifier combination technique to classify cloud systems, namely stratiform and convective systems. The data used come from the SEVIRI (Spinning Enhanced Visible and Infra-Red Imager) radiometric imager on board the MSG (Meteosat Second Generation) satellite. The methods were trained and evaluated with satellite and radar data of the rain event of October 10, 2018 in the south of France provided by Météo-France. This allowed us to evaluate and compare the performances for cloud identification of method frequently used in combination of classifiers

Keywords: Machine Learning, Combining Classifiers, Boosting, Ada-Boost, Rainfall Estimation, Delineation Raining Cloud, SEVIRI, MSG, Radar.

1 Introduction

The techniques based on the supervised learning methods proposed, for solving the problem of identifying cloud systems and estimating precipitation [2, 3, 11, 12], exploit the information from the multi-spectral channels of the SEVIRI (Spinning Enhanced Visible and Infra-Red Imager) radiometer imager. Some of these methods have shown a performance that surpasses the others to predict one or more classes. The idea of making these different methods cooperate in order to improve the precision of the classification appeared with the development of new called hybrid techniques.

Two approaches are used for multi-classifier cooperation. The first approach is to merge information from different classifiers for decision making, considering each classifier as an independent source of information and exploiting the characteristic performance of these sources to favor the most reliable classifier to predict the class membership. To do this, classifier fusion relies on three main strategies that are the voting principle [10, 16], possibility theory [7, 17] and belief theory [1, 4]. Each of these strategies is adapted for the type of information processed or the relevant decision criterion for classification. The second approach is to use strategies or mechanisms for classifiers using different subset of data for each classifier or using different classifiers for a set of training data. These strategies are at the origin of several popular multi-classifier algorithms such as Boosting [9] and Bagging [5].

It is the second approach that we will focus on this study. To do this, this paper we will detail the principle of these different mechanisms of combining classifiers used in Section 2, and by subjecting our database to tests that will be analyzed in Section 3. Finally, we draw some conclusions in Section 4.

2 Combining Classifiers

Combining classifiers consists of combining the prediction of a set of independent classifiers. What has been at the origin of a very active field of research in supervised learning, by exploiting the local precision of the different learning methods to form a set of classifiers, which improves the global precision for the same application [6]. Several methods have been implemented by combining several classifiers, in this section we will expose some methods, which have shown their effectiveness for the data processed in our research study.

2.1 Bagging

The principle of Bagging is "Bootstrap aggregation", it is a resampling of the training dataset, by randomly generating other subsets of data from the original training set, while keeping the number of samples it contains, which means the appearance of the same sample several times and the disappearance of several others in the same subset. Indeed, a larger classifier prediction error for the absent or less represented classes in the subsets. However, the combination of classifiers used for the subsets gives higher prediction accuracy than a single classifier and a training dataset. The advantage of bagging is the reduction of the variance without influencing the accuracy of the

prediction. Also, the study conducted by the effectiveness of Bagging is more significant for learning methods whose prediction decision is significantly different when there is an insignificant change in the learning set [5].

2.2 Boosting

Boosting consists of a series of classifiers of different learning methods whose prediction is weak. Its principle is that the samples misclassified by the first classifier are chosen more often unlike the correctly predicted samples, with the aim of constituting the training data set submitted to the next classifier, this step is repeated for all of the classifier until the end of the series. Methods based on the Boosting principle have been developed, the Arcing method and the Ada-Boost method [9]. The Arcing method [5], the frequency of samples misclassified by a classifier is the selection criterion in the training dataset submitted to the next classifier; this step is repeated until the end of the series of classifiers. The Ada-Boost method [8] uses a series of classifiers, which exploits the probabilities of the set of samples at the output of a classifier, and then weight the prediction error of each sample, the new set retains the completeness of the samples in the training set.

2.3 Stacking

The mechanism at the base of Stacking consists of two levels, a Meta level where a classifier is found that allows the combination of the classification of a set of lower level classifiers, which aims to improve the prediction of all the classifiers compared to those obtained by these same classifiers individually. The selection algorithm is based on the principle of cross-validation, a base level classifier having collected the best vote for the prediction of a class among the set of classifiers submitted to the different subsets of the training set, will be chosen by the meta level classifier to predict this same class [16].

3 Data and Application

Three types of climate for which their precipitation rates are important. Namely, the oceanic climate found in Western Europe bordered by the Atlantic Ocean to the west, with annual precipitation varying between 900 and 1800 mm in the form of drizzle and fog. In addition, the Mediterranean climate, with annual precipitation varying between 250 and 1150 mm, Also, the transitional climate of the mountain ranges of the Tell Atlas and the Saharan Atlas, which separates the region bordered by the Mediterranean Sea from the desert from the Sahara to the south, with annual precipitation decreasing to less than 100 mm south of the Saharan Atlas. The exclusion of the arid desert climate from our study is due to very rare precipitation in the form of downpours and of little interest for our study, with annual quantities which often do not exceed 50 mm. The rainy season lasts from October to March. A band is defined with a latitude of 30° to 50° North and a longitude of 6° West to 12° East as shown in following figure.

3.1 Data

The data used in this study are satellite and radar images, which represent a rainy and stormy event in the Mediterranean and going back to the south of France from October 10, 2018 from 08h GMT to 18h GMT. In addition, precipitation intensities exceeding 30 mm/h and cumulative rainfall more than 250 mm in certain regions, which is an indication of the presence of an intense convective system, caused by an anti-cyclone in the center of Europe, a depression in the North Atlantic Ocean and upwelling of a warm air masses from Africa.

The satellite data are images acquired by the SEVIRI radiometer imager (Spinning Enhanced Visible and Infra-Red Imager) on board the MSG4 (Meteosat Second Generation 4) satellite. These images are captured every 15 minutes in 12 spectral channels and spatial resolution of 3 km above the equator. Also, the images were processed in real time in order to correct the radiometric and geometric effects induced during the capture and geo-located, then the information is localized on the Earth, calibrated and the radiances linearized so that they are exploitable [8].

These data exploited in order to provide parameters for the identification of the characteristics of clouds and the types of precipitation of each pixel of the images captured, the parameters we will discuss in this section.

Two types of clouds are the source of significant precipitation in the region of our study, stratiform and convective clouds, which can be detected by exploiting the parameters of the cloud as detailed in the study [3]. Relationship which connects them to the data provided by SEVIRI channels and summarized by the following table:

Table 1. Relationship between cloud parameters and SEVIRI data.

Parameters	Channel	Relationship
Cloud Top Temperature	$IR_{10.8}$	$AT = \frac{\sum_{i=1}^{N} T_{IR10.8}(i)}{N}$
Vertical Extension of the Cloud	$\Delta T_{IR10.8-IR12.0}$	$H = \frac{T_{max} - T_{min}}{6.5} - dH$
Altitude of the Cloud	$\Delta T_{WV6.2-IR10.8}$ $\Delta T_{WV7.3-IR12.0}$	Very negative value => Low altitude Very low value => High altitude
Cloud Water Path	During the day $\Delta T_{VIS0.6-NIR1.6}$ During the night $\Delta T_{IR3.9-IR10.8}$	$CWP = \frac{2}{3} * \rho * \tau * r_e$
Cloud Phase	$\Delta T_{IR8.7-IR10.8}$	Low and negative value => Low altitude

With:

- N is the number of pixels that form the cloud.
- dH is the minimum height of the cloud or the rain column which is 650 m.
- ρ is the density of the water contained in the cloud.
- (τ) is the optical thickness and the effective particle radius (r_e) of the cloud.
- (r_e) is the effective particle radius of the cloud.

Radar data are images captured by the Météo-France radar network located in the south of France. These images are captured every 5 minutes with a spatial resolution of 1 km, processing has been applied to eliminate noise caused by the terrain and neighboring radars [13]. The information is the reflectivity, which is the measurement of the energy reflected by the precipitation in (dBZ) and it gives us explicit information about the water content of the cloud.

For a spatial and temporal correspondence of the data manipulated, we carried out the co-localization and the synchronization of the satellite and radar data.

Table 1 illustrates the relationship for calculating the precipitation rate R (mm/h) from the reflectivity measured by the radar Z (dBZ) [12].

Table 2. Reflectivity and precipitation rate relation

Type	Relationship	Corr. Coeff.
Strat. Precipitation $Z \leq 38\ dBZ$	$Z = 248.28 \times R^{1.58}$.	0.79
Conv. Precipitation $Z \geq 42\ dBZ$	$Z = 412.35 \times R^{1.49}$.	0.84

With

$$Z(dBZ) = 10\log(Z) \quad (1)$$

3.1 Application

The input variables of our models are calculated from the data of multispectral channels of the SEVIRI radiometer of the Meteosat Second Generation satellite. Which contains characteristic parameters of the clouds deduced from the SEVIRI data, which correspond to the ground truth data, which are the radar data.

The model allows assigning a class to a novel sample using the samples from the training set. For this, we present as input the variables that characterize this novel samples; the class thus assigned by the model represents one of the three classes non-precipitant, stratiform precipitant and convective precipitant.

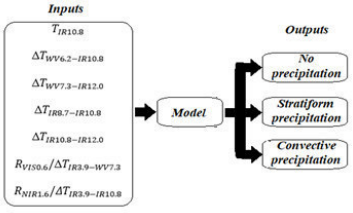


Figure 1 Structure of the Bagging and Ada-Boost model that combines spectral parameters from SEVIRI images

4 Results and Discussion

To evaluate the performance of Ada-Boost method for cloud identification were tested to the SEVIRI and radar data samples, these datasets being taken simultaneously and respecting the spatial correspondence between them. We chose a database belonging to the stormy event of October 10, 2018. This database contains 3,187,904 samples, which correspond to eight scenes captured between 12 h UTC and 14 h UTC, which were divided in two, 80% for training and 20% for testing of the two ensemble classifiers, Bagging and Ada-Boost that we chose for our application because they showed good performance in other application areas.

In order to compare the Bagging and Ada-Boost methods with other methods, the two stratiform and convective classes are in a single class called precipitant.

The appropriate statistical parameters used to quantify the estimation errors of the methods with radar measurements calculated from the contingency table summarized in Table 4.

Table 3. Contingency table of estimates and measures.

Identified by WkNN	Observed by the radar		
	Precipitating	Non-precipitating	Total
Precipitating	a	b	$a+b$
Non-precipitating	c	d	$c+d$
Total	$a+c$	$b+d$	T

Where a, b, c and d are contingency table values and $T = a + b + c + d$

- The rate of correctly identified events is calculated with the Probability Of Detection (*POD).*

$$POD = \frac{a}{a+b} \quad \textbf{(2)}$$

- The rate of number of pixels misidentified by the WKNN method is calculated with The Probability Of False Detection (*POFD*).

$$POFD = \frac{b}{b+d} \quad \textbf{(3)}$$

- The rate of events estimated when they were not events is calculated with the False Alarm Ratio (*FAR).*

$$FAR = \frac{b}{a+b} \quad \textbf{(4)}$$

- The difference between the estimate and the measurements is calculated with the Frequency BIAS index (*Bias*):

$$Bias = \frac{a+b}{a+c} \quad \textbf{(5)}$$

- The estimated correctly diagnosed event rate is calculated the Critical Success Index (*CSI*).

$$CSI = \frac{a}{a+b+c} \quad \textbf{(6)}$$

- The rate of correct estimates is calculated with the Percentage of Corrects (*PC*).

$$PC = \frac{a+d}{T} \quad \textbf{(7)}$$

The comparison between the Bagging and Ada-Boost methods and *WkNN* method the *ECST* technique is summarized in Table 5.

Table 4. The statistical results of the verification for *Ada-Boost* and *Bagging*.

	POD	POFD	FAR	Bias	CSI	PC
ECST	0.64	0.06	0.29	0.85	0.58	0.93
WkNN	0.68	0.05	0.26	0.89	0.61	0.95
Ada-Boost	0.56	0.1	0.35	1.21	0.55	0.86
Optimal values	1	0	0	1	1	1

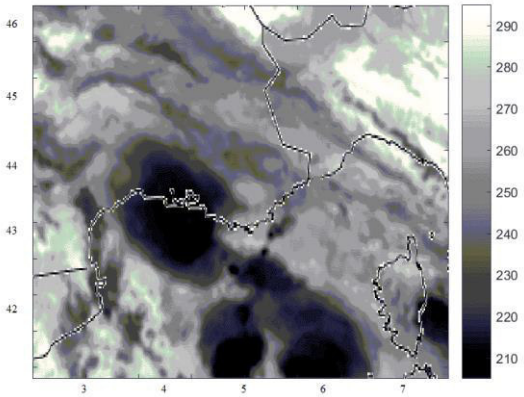


Figure 2: $T_{IR10.8}$ image scene from October 10, 2018 at 12:40 UTC

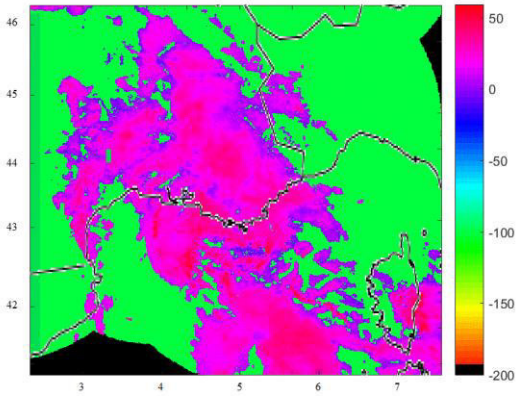


Figure 3: The rain area delineated by the Météo-France radar network in dBZ

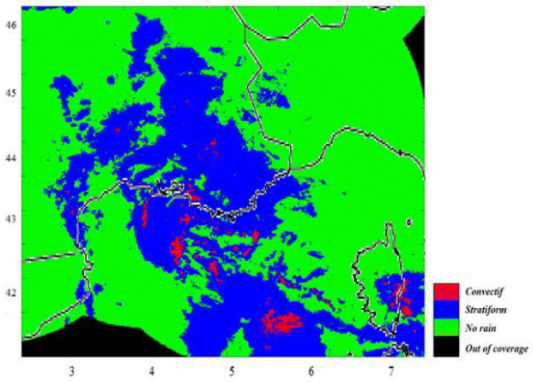


Figure 4: The rain area delineated by the Météo-France radar network in 4 levels

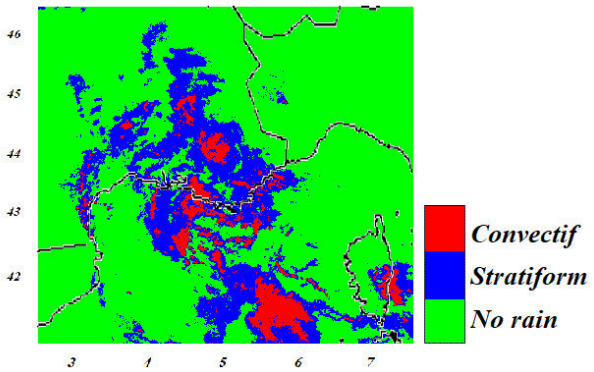


Figure 5 The rain area delineated by the Ada Boost method in South East of France

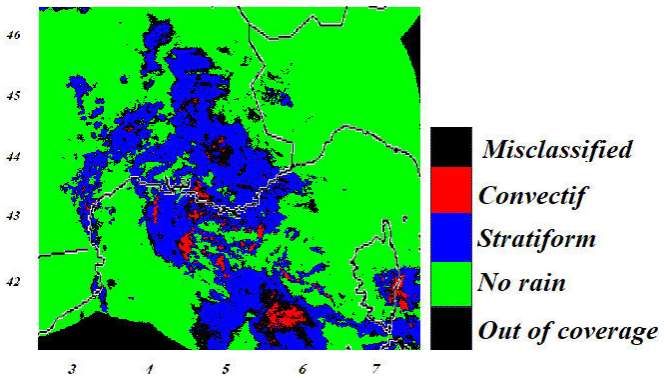


Figure 6 Comparison between Ada Boost method and the Météo-France radar network delineation

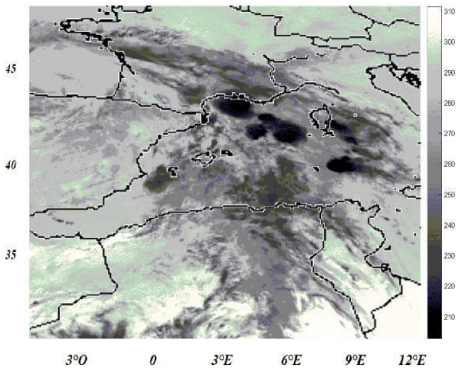


Figure 7 $T_{IR10.8}$ image scene from October 10, 2018 at 12:40 UTC over North of Algeria and Mediterranean sea

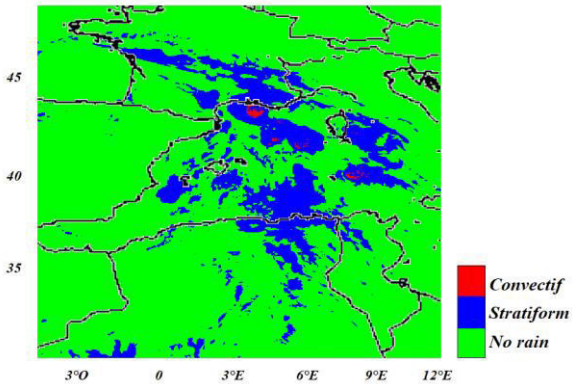


Figure 8 The rain area delineated by the Bagging method over North of Algeria and Mediterranean sea

5 Conclusion

In light of the results obtained, we deduce that the Ada-Boost classifiers have a good prediction; this is valid for the four predefined classes. When the number of samples of the same class weakly represented in the database, the prediction rate is good for the Ada-Boost method. The Ada-Boost method is more suitable for a heterogeneous database. As a perspective to this study, we recommend to use a homogeneous database for an application to Ada-Boost classifiers.

References

1. Appriou, A. (2002), Discrimination multi signal par la théorie de l'évidence, In décision et Reconnaissance des formes en signal, Hermes Science Publication, pp 219-258, 2002.
2. Bensafi, N., Lazri, M., Ameur, S., 2019. Novel WkNN-based technique to improve instantaneous rainfall estimation over the north of Algeria using the multispectral MSG SEVIRI imagery. Journal of Atmospheric and Solar-Terrestrial Physics 183, 110–119.
3. Bensafi, N., Attaf, Y., Lazri, M., Ameur, S. (2023). Delineation of Raining Cloud Using a WkNN from Multispectral Data of SEVIRI Radiometer. In: Laouar, M.R., Balas, V.E., Lejdel, B., Eom, S., Boudia, M.A. (eds) 12th International Conference on Information Systems and Advanced Technologies "ICISAT 2022". ICISAT 2022. Lecture Notes in Networks and Systems, vol 624. Springer, Cham. https://doi.org/10.1007/978-3-031-25344-7_21
4. Bloch, I. (2003), Fusion d'informations en traitement du signal et des images, Lavoisier (eds), Hermes Science Publication, 2003
5. Breiman, L.: Bagging Predictors. Machine Learning 24(3), 123–140 (1996)
6. Dietterich, T. G. (1997). Machine-learning research: Four current directions. AI Magazine, 18:4, 97–136.
7. Dubois, D. et Prade, H. (1988), Possibility Theory, Plenum Press, New York, 1988.
8. https://www.eumetsat.int
9. Freund, Y., Schapire, R.E.: Experiments with a New Boosting Algorithm. In: proceedings of ICML'96, pp. 148–156 (1996)
10. Lam, L. et Suen, C.Y. (1997), Application of Majority Voting to Pattern Recognition: An Analysis of Its Behavior and Performance, IEEE Transactions on Systems, Man Cybernetics 1997, 27(5), pp 553-568.
11. Lazri, M., Ameur, S., 2018. Combination of support vector machine, artificial neural network and random forest for improving the classification of convective and stratiform rain using spectral features of SEVIRI data. Atmospheric Research 203, 118–129.
12. Lazri, Mourad, Zohra Ameur, Soltane Ameur, Yacine Mohia, Jean Michel Brucker, et Jacques Testud. « Rainfall Estimation over a Mediterranean Region Using a Method Based on Various Spectral Parameters of SEVIRI-MSG ». Advances in Space Research 52, no 8 (octobre 2013): 1450-66.
13. Ouallouche, F., Lazri, M., Ameur, S., 2018. Improvement of rainfall estimation from MSG data using Random Forests classification and regression. Atmospheric Research 211, 62–72.
14. Tabary, P., Augros, C., Champeaux, J.-L., Chèze, J.-L., Faure, D., Idziorek, D., Lorandel, R., Urban, B., Vogt, V., 2013. Le réseau et les produits radars de Météo-France. Météorologie 8, 15.
15. Ting, K.M., Witten, I.H.: Stacking Bagged and Dagged Models. In: Fourteenth international Conference on Machine Learning, San Francisco, CA, pp. 367–375 (1997)
16. Wolpert, D. (1992). Stacked generalization. Neural Networks, 5:2, 241–260
17. Xu, B.V., Krzyzak, A. et Suen, C.Y. (1992), Methods of Combining Multiple Classifiers and Their Application to Handwriting Recognition, IEEE Transactions on Systems, Man Cybernetics 1992, 22(3), pp 418-435.
18. Zadeh, L.A. (1978), Fuzzy Sets as a Basis For a Theory of Possibility, Fuzzy Sets and Systems 1978, 1, pp 3-28.

A Novel Approach Based on Geometric Mean Algorithm for Optimum Multiple Capacitor and DG Allocation in Radial Energy Distribution

Imene Djedidi
Electrical Engineering departement
University of Biskra, Biskra, Algeria
Biskra BP 145, 07000 Biskra, Algeria
imene.djedidi@univ-biskra.dz

Ahmed Salhi
Electrical Engineering departement
University of Biskra, Biskra, Algeria
Biskra BP 145, 07000 Biskra, Algeria
a_salhi_m@yahoo.fr

Anes Bouhanik
Electrical Engineering departement
University of Biskra, Biskra, Algeria
Biskra BP 145, 07000 Biskra, Algeria
bouhanikanes1@gmail.com

Djemai Naimi
Electrical Engineering departement
University of Biskra, Biskra, Algeria
Biskra BP 145, 07000 Biskra, Algeria
d.naimi@univ-biskra.dz

***Abstract*—This research paper proposes a new efficient optimization approach based on the Geometric Mean Algorithm (GMO) to solve an optimal the size and location problem of multiple of capacitor banks (CBs) and distributed generations (DGs) in a radial distribution system (ONRSACD). The integrating capacitor banks (CBs) and distributed generations (DGs) at distribution sub-stations enhances energy quality and cuts power costs under equal and unequal constraints. Validation and performance assessment of the proposed methodology are conducted on radial IEEE 33-node network. In addition, the effectiveness of the Geometric Mean Optimizer Algorithm (GMO) is compared to that of a well-established algorithm, including Particle Swarm Optimization (PSO), to identify the top tool with respect to convergence characteristics and the best quality solution. Our results demonstrate that (GMO) gives the top solutions at the right time for the computation.**

Keywords— Distribution System Radial, GMO, Optimal Allocation, Power Loss, CB, DG, Voltage Profile.

I. INTRODUCTION

The distribution network grid is an important part of the power grid, and is considered to be one of the most critical infrastructures for a country's industrial and economic development, as it is responsible for supplying electrical energy to customers. Indeed, even the slightest power failure can result in a failure to meet electricity needs, with undesirable economic and social consequences. The need for reliable electrical networks operating to contractual standards is therefore a technical challenge and an increasingly important issue [1].

In recent times, increasing growth in load demand and financial pressures to modernize the electricity distribution system have been increasing. The system faces challenges such as power loss, voltage deviations, unreliability, and voltage instability. In the wake of environmental and energy crises, there is also a sense of responsibility to adopt distributed renewable energy resources for power generation.

In order to overcome these problems, the installation of distributed generation can be considered as an alternative solution. Distributed generation is nothing more than small-scale power generation close to load centres. The installation of distributed generation also makes it possible to accommodate new load areas when the grid is expanded. Today, renewable energy sources such as wind and solar are increasingly being used because of their environmental friendliness. Based on energy production, DG can be classified as follows [2]:

Type 1: injects only real energy into the system.

Type 2: injects a mixture of active and reactive power.

Type 3: injects reactive power only.

Type 4: injects active power but consumes reactive power.

The proposed approach leverages the geometric mean optimizer algorithm (GMO) [15] to maximize techno-economic benefits and satisfy critical operational constraints to achieve a robust solution. The primary objective is to achieve a good compromise between mutually conflicting costs: substation energy costs and device investment costs. In pursuit of a realistic network behavior model, the method considers the hourly variations of diverse load types and DGs output. To address the inherent multi-objective nature of the problem, the proposed approach seamlessly integrates the fuzzy logic tool. Each objective function is converted into the fuzzy domain using its respective membership function, and upper and lower bounds are determined through a carefully devised strategy [3].

Many studies with different solution approaches have been done for this optimization problem. These approaches are classified in [4-5] into (1) analytical methods such as those proposed in [6-7], (2) metaheuristic methods such as genetic algorithm (GA) [8], artificial bee colony (ABC) [9], artificial ecosystem-based optimization (AEO) [10], Harris hawk optimization (HHO) [11], and moth flame optimization (MFO) [12]. The hybrid approach, which combines analytical and meta-theoretical tools to address the issue of optimal

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

location and sizing of generators, is proposed in reference [13].

An innovative multi-objective optimization methodology aimed at addressing the complex issue of determining the locations and sizes of capacitor banks (CBs) and distributed generations (DGs) that are optimal at the lowest cost of energy operating at distribution radial systems is presented in this article. This application is achieved at radial IEEE 33-node network. For each system test, a list of sensitive nodes, so the most likely ones, is first generated based on their voltage profile, and then each algorithm is applied to determine the optimal solution. Finally, the results obtained are compared with each other to identify the most effective tool together with the respect of the constraints of equality and inequality.

Rest of the paper is organized as follows: part II explains the formulation of the problem, part III gives an overview for the algorithm proposed, simulation result is presents in section IV as well as the conclusions and some perspectives for further work, in part V.

II. PROBLEM FORMULATION

A. Objectve Function

For load flow calculation, backward/forward [14] sweep method is used. The objective function is to minimize the real power losses and the cost in the system by various constraints in the distribution network sweep method by optimizing the GMO, which is described below:

$$min\ f = min\ TP_{LOSS} = \ min \sum_{i=1}^{n} R_i \cdot |I_i|^2 \quad (1)$$

Where, TP_{loss} is total active power loss in branch,

n_{br} The total number of branches,

R The resistance of i-th branch in the network,

I_i The current magnitude of i-th branch.

Operating cost are calculated using the following equations [6,9]:

$$TOC = C_{TP_{LOSS}} + C_{DG} \ + C_{CB} \quad (2)$$

$$C_{TP_{LOSS}} = TP_{LOSS} \times \left(K_P + K_e + L_{sf} \times 8760\right) \$ \quad (3)$$

$$C_{DG} = \sum K_{DG_P} \times P_{DG} + \sum K_{DG_Q} \times Q_{DG} \quad (4)$$

$$C_{CB} = C_{cb} \times \sum_{k=1}^{ncb} Q_{cb,k} \quad (5)$$

$$L_{sf} = K \times L_f + (1 - K) \times L_f^2 \quad (6)$$

Where, K_p is annual demand cost of power loss (\$/kW),

K_e Annual cost of energy loss (\$/kWh),

L_{sf} The loss factor,

$Q_{cb,K}$ The reactive power of the capacitor bank k ,

C_{cb} is the purchase cost (\$/KVAR).

B. Equality Constraints

The constraints on equality are defined by power balance equations such as the following:

$$P_{sub} + \sum_{k=1}^{nDG} Q_{DG} = \sum_{i=1}^{nbr} P_{loss,i} + \sum_{j=1}^{Nb} P_{load,j} \quad (7)$$

$$Q_{sub} + \sum_{k=1}^{nDG} Q_{DG} = \sum_{i=1}^{nbr} Q_{loss,i} + \sum_{j=1}^{Nb} Q_{load,j} \quad (8)$$

Where P_{sub} and Q_{sub} are real and reactive substation power respectively, $P_{load,j}$ and $Q_{load,j}$ are real and reactive load demands at node j respectively, N_b number of nodes.

C. Inequality Constraints

- *Voltage limits of the node*

$$V_{min} \leq |V_i| \leq V_{max} \quad ; \quad i = 1,2, \dots, n_{node} \quad (9)$$

Where, $V_{min} = 0.95\ (pu)$, and V_{max}=1.05 (pu).

- Line capacity limits

$$I_{n,i} \leq I_{max,i} \quad ; \quad i = 1, \dots, nbr \quad (10)$$

Where, $I_{Line\ (i)}$ is the current flow through network branches, I_{max} is maximum allowed branch current.

- Size of DG

$$P_{DG}^{min} \leq |P_{DGi}| \leq P_{DG}^{max} \quad (11)$$

$$Q_{DG}^{min} \leq |Q_{DGi}| \leq Q_{DG}^{max} \quad (12)$$

- Position of DG

$$2 \leq DG_{node} \leq n_{node} \quad (13)$$

- Reactive power limits

$$Q_{cb,min} \leq Q_{cb,\ k} \leq Q_{cb,max} \quad (14)$$

- Intégration of DG, CB

$$C_{DG} + \ C_{CB} \ \leq \sum_{j=1}^{Nb} Q_{load,j} \quad (15)$$

n_{node} is the number of nodes, DG_{node} is the node number of the DG installation, V_i the node voltage.

III. GEOMETRIC MEAN OPTIMIZER (GMO)

In 2023, Seyed Ali Mirjalili et al. introduced a new metaheuristic algorithm called Geometric Mean Optimizer (GMO), which mimics the unique properties of the geometric mean operator in mathematics according to [15], The solution optimization mechanism used in the GMO framework is shown in Fig.1. During the optimization process, the magnitude of the vector φ diminishes and gradually decreases. This decrease, particularly in the value ranges of φ, is intended to increase the exploration capabilities of the GMO in the initial iterations. Simultaneously, it substantially improves the emphasis on exploitation during the final iterations. This approach is strategically developed to guarantee a well-

balanced transition between exploration and exploitation, thereby enabling a more efficient optimization process.

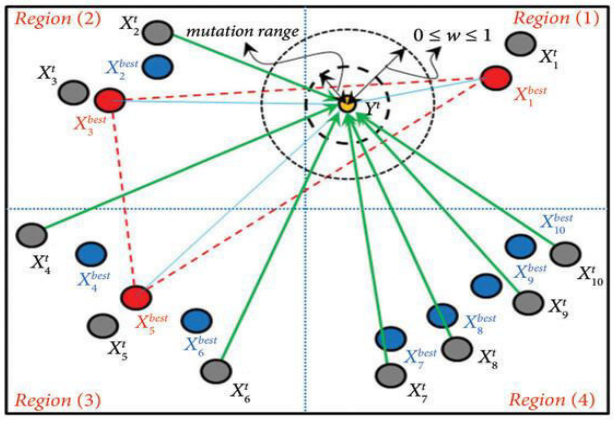


Fig.1. Solution updating concept of GMO

The equations mathematical that support the GMO model are as follows [15]:

$$f(x) = \frac{1}{1 + e^{a(x-c)}} \; ; a < 0 \qquad (16)$$

Subject to

$$c = \mu \quad and \quad a = \frac{-4}{\sigma\sqrt{e'}} \qquad (17)$$

Where, a and c are parameters of the sigmoidal MF, with μ and σ representing the mean and standard deviation of the x values, respectively. Meanwhile, e denotes Napier's constant.

Substituting x with the value of the objective function calculated for a personal best-so-far search agent enables the calculation of the fuzzy MF value for this agent as follows:

$$MF_j^t = \frac{1}{1 + \exp[-(\frac{4}{\sigma^t}\sqrt{e}) \times (Z_{best,j}^t - \mu^t)}; \quad j = 1,2,..,N, \qquad (18)$$

Where, t represents the current iteration number and N denotes the total population size. $Z_{best,j}^t$ and MF_j^t signify, respectively, the objective function value and the MF value of the j[th] best-known solution position at the t[th] iteration.

Utilizing a specialized measure termed the dual-fitness index (DFI), the GMO can evaluate both the performance metric and solution diversity within the search space simultaneously. The definition of the DFI associated with the ith solution is as follows:

$$DFI_i^t = MF_1^t \times ... \times MF_{i+1}^t \times .. \times MF_N^t = \prod_{j}^{N} MF_j^t \quad j \neq i \qquad (19)$$

To guarantee that all personal best-so-far agents contribute to the creation of a unique global guide agent for each individual agent, a weighted average of all personal best-so-far solutions is established. The weights employed in this calculation correspond to the DFI values assigned to individual solutions. Equ (20) outlines this relationship as follows:

$$\sum_{j=1}^{N} DFI_j^t \times X_j^{best}; \quad Y_i^t = \frac{j \neq i}{\sum_{j=1}^{N} DFI_j^t + \varepsilon} \qquad (20)$$

Where, Y_i^t denotes the position vector of the unique global guide agent derived for solution i during iteration t, X_j^{best} signifies the best-so-far position vector of the j[th] solution, and ε functions as a coefficient to mitigate occurrences of singularity.

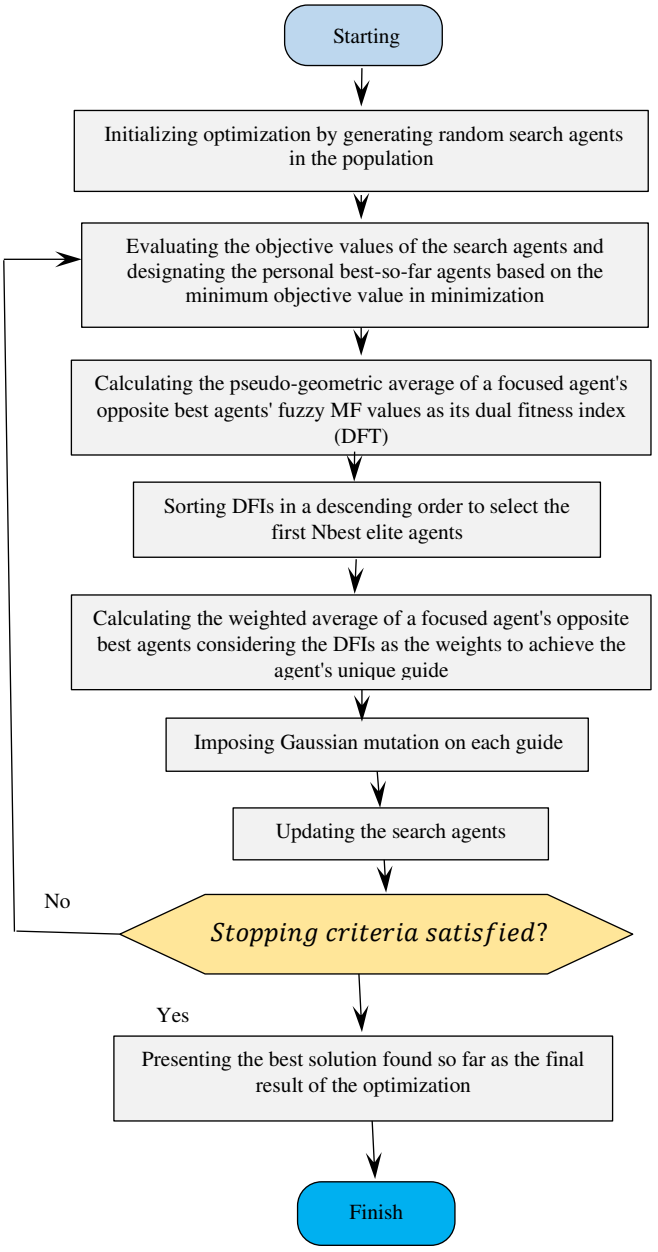


Fig.2. Steps of GMO algorithm

To optimize the efficiency of the exploration of GMO and reduce the computational load of executing the algorithm, personal best-so-far solutions are prioritized according to their DFI values, from highest to lowest. The top N_{best}

personal best-so-far solutions are identified as elite solutions. Determining the number of N_{best} involves a method where it linearly decreases across iterations, starting with a value equal to the total population size at the initial iteration and reducing to 2 by the final iteration. Consequently, on (20) is adapted as follows:

$$Y_i^t = \frac{\sum_{j \in N_{best}\, j \neq i} DFI_j^t \times X_j^{best}}{\sum_{j \in N_{best}} DFI_j^t + \varepsilon} \quad (21)$$

GMO, the Gaussian mutation strategy is integrated to increase the diversity among guide agents. This mutation strategy, when applied to the guide agents within this framework, is expressed through the following equation:

$$Y_{i,mut}^t = Y_i^t + \omega \times r_n \times (Std_{max}^t - Std^t) \quad (22)$$

Where, $Y_{i,mut}^t$ denotes the mutated version of Y_i^t w represents a parameter designed to reduce the magnitude of mutation over the course of iterations, r_n is a random vector, Std^t is the vector indicating the standard deviation as calculated across the best-so-far solutions, and Std_{max}^t refers to a vector that contains the maximum standard deviation values, each aligned with the dimensions of the personal best-so-far solutions.

The updating equations for each solution can be formulated as follows:

$$X_i^{t+1} = X_i^t + V_i^{t+1} \quad (23)$$

Subject to

$$X_i = (x_{i1}, x_{i2}, \ldots\ldots\ldots, x_{iD}),$$

$$V_i = (v_{i1}, v_{i2}, \ldots\ldots\ldots\ldots, v_{iD}), \quad (24)$$

$$V_i^{t+1} = \omega \times V_i^t + \varphi \times (Y_{i,mut}^t - X_i^t) \quad (25)$$

$$\omega = 1 - \frac{t}{t_{max}} \quad (26)$$

$$\varphi = 1 + (2 \times r_1 - 1) \times \omega \quad (27)$$

Where, t_{max} denotes the maximum iteration ; V_i^t and V_i^{t+1} represent the speed vector of the solution i at iterations t and (t + 1), respectively; φ is a vector of adjustment parameters; and r_1 is a random number generated with the interval [0, 1]. Finally, the optimal solution will be entered. The overall GMO process is shown in Fig.2.

IV. Tests and Results

GMO was implemented via MATLAB programming language to solve the shunt capacitor bank (CB) and distributed generation (DG) allocation problem to determine the optimal size and placement. The robustness of the above optimisation tools is verified using the IEEE-33 node test system distribution in radial presented in Fig.3.

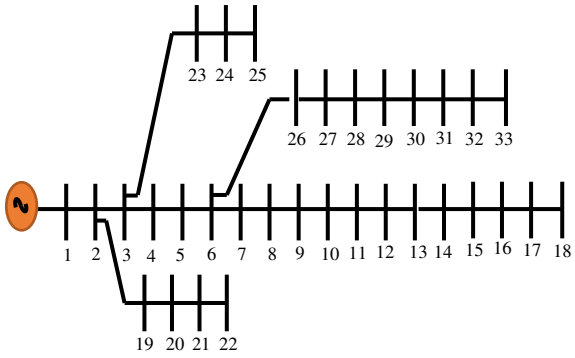


Fig. 3. the IEEE-33 node test in distribution system radial

In this study, the following values are applied:

- $V_{min} = 0.95\ (p.u), V_{max} = 1.05\ (p.u)$,
- $V_{sub} = 12.66\ (KV)$,
- Size of DG: 10 < Size $_{DG}$ <500 (KW),
- Size of CB: 10 < Size $_{CB}$ <100 (KVAR),
- Max number of iterations is chosen to T=1000,
- The population size is set to be equal to 50.

TABLE I. Comparison Table For Optimal alLocations Of Capacitor Banks and Distributed Generations In 33-Node System

Method	Node	P_{DG} KW	Q_{CB} KVAR	P_{Loss} kW	$V_{magnitude}$ (p.u)
Uncompensated	-	-	-	210.99	0.9038
GMO [15]	14 30 24	500 500 500	100 100 100	70.0652	0.9603
PSO [16]	13 30 24	473.2261 499.9660 491.5410	83.6651 95.6226 97.4151	72.8164	0.9595

A. *Discussion of obtained results*

A comparative analysis of the Geometric Mean Optimizer (GMO) and Particle Swarm Optimization (PSO) algorithms was performed to optimize the concurrent allocation of Distributed Generations (DGs) and Capacitor Banks (CBs) with the aim of minimizing power losses. With a population size of 50 and a maximum of 1000 iterations, the results demonstrate that (GMO) consistently surpassed (PSO) in attaining a superior solution. Specifically, (GMO) recorded total power losses of 70.0652 (kW) and a minimum voltage magnitude of 0.9603 (p.u) of (base values). whereas (PSO) incurred greater losses of 72.8164 (kW) and a somewhat lower minimum voltage of 0.9595 (p.u) of (base values). Furthermore, the convergence curve of (GMO) exhibited markedly enhanced efficiency and expedited advancement towards an ideal solution relative to (PSO). These findings highlight the strength and effectiveness of the (GMO) algorithm in tackling intricate optimization challenges in power distribution networks."

These important results were achieved in the first iterations, as shown in Fig.5. In Fig.4, we can see that the improvement in the voltage profile is accompanied by an improvement in the highlighted voltage profile

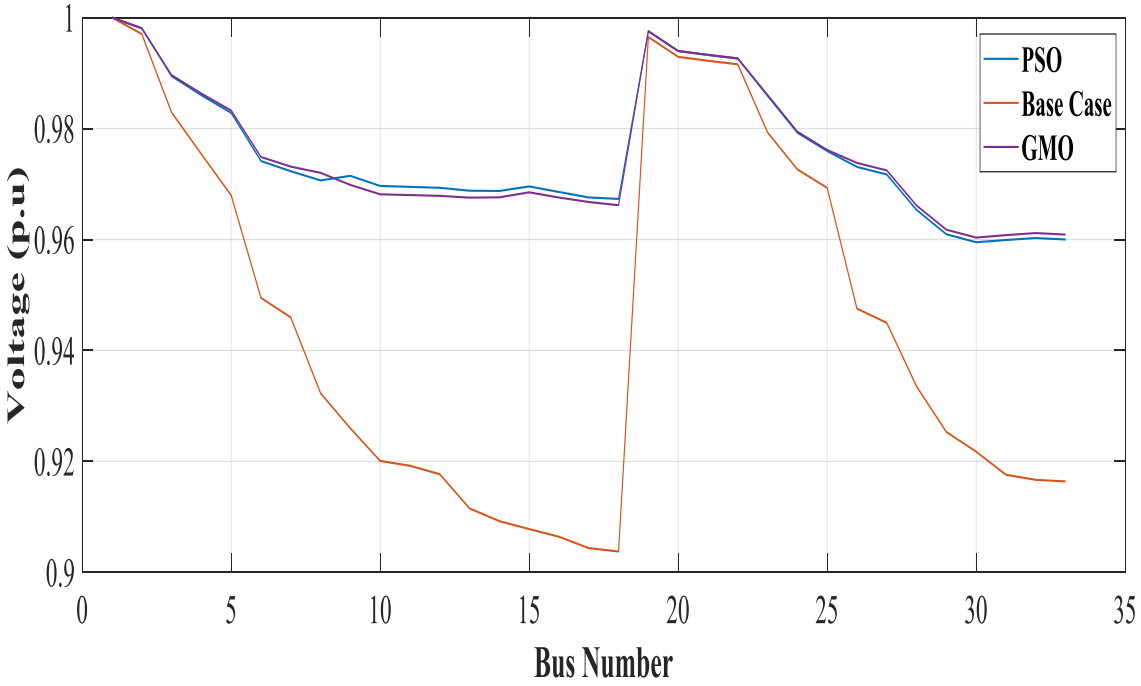


Fig.4. Voltage Profile in IEEE-33 node test

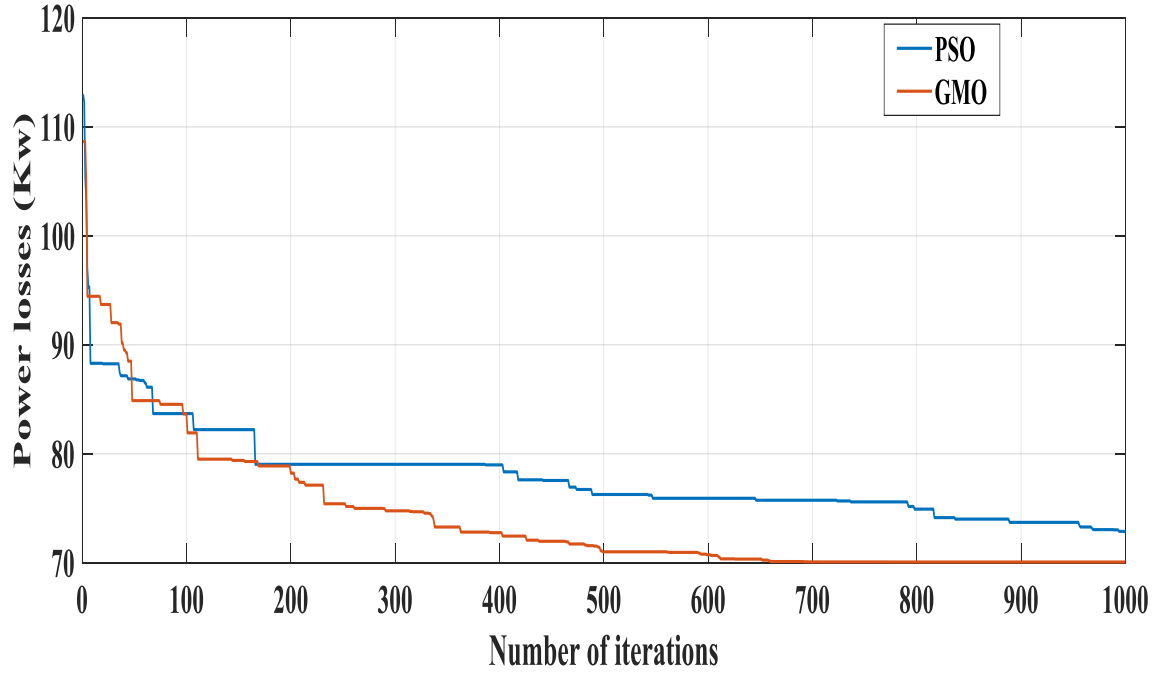


Fig.5. Convergence characteristics of the optimization algorithms for IEEE-33 node test system

V. CONCLUSION

This work proposes to optimise the Geometric Mean (GMO), using a recently developed efficient meta-heuristic optimisation technique, to solve the allocation of CBs and generators from the selected node in a radial distribution system based on objective of minimizing the total power loss and voltage deviation. Different test systems from 33 IEEE nodes are used to verify the efficiency and superiority of the GMO technique, as well as the accuracy and relevance in solving these problems.

We compare the obtained results with the PSO approach to verify the effectiveness of the GMO algorithm, and the results show that this technique is easily implemented and gives good solutions (power loss, tension profile, operating cost and convergence capability), however, the GMO provides the best ones compared to PSO algorithm.

For future work, we intend to apply these techniques for problem of optimal network reconfiguration simultaneously with the allocation of capacitor banks (CBs) and distributed generations (DGs) in a radial distribution system, or the issue of allocating CBs and DGs in different load demands can be studied. All these processes can be confirmed in real time runs.

REFERENCES

[1] Djedidi, Imene. "Optimisation Des Performances Du Réseau Électrique de Distribution Par Des Méthodes Évolutionnaires". Diss. Université Mohamed Khider Biskra, (2023).

[2] D.Q. Hung, N. Mithulananthan, R.C. Bansal, Analytical expressions for DG allocation in primary distribution networks, IEEE Trans. Energy Convers. (2010) 814–820.

[3] Bouhanik, Anes, et al. "An efficient hybrid multi-population algorithm (HMPA) for enhancing techno-economic benefits." Soft Computing (2024): 1-33.
https://doi.org/10.1007/s00500-024-09807-8

[4] H.N. Ng, M.M.A. Salama and A.Y. Chikhani, "Classification of capacitor allocation techniques, " IEEE Trans. Power Deliv., vol.15, 2000, pp. 387–392.
https://doi.org/10.1109/61.847278

[5] S. Kalambe, and G. Agnihotri, "Loss minimization techniques used in distribution network: Bibliographical survey," Renew. Sustain. Energy Rev.2014, pp. 184-200.
https://doi.org /10.1016/j.rser.2013.08.075

[6] S. Nawaz, M.P. Sharma and A. Cupta, "Optimal allocation of capacitor bank in radial distribution system using analytical approach," IJECE, vol. 7, 2017, pp. 748–753.
http://doi.org/10.11591/ijece.v7i2

[7] Arief, Antamil, and M. B. Nappu, "An Analytical Method for Optimal Capacitors Placement from the Inversed Reduced Jacobian Matrix, " Energy Procedia, vol.100, 2016, pp. 307–310.
https://doi.org/10.1016/j.egypro.2016.10.182.

[8] M. A. Mehanna, and M. H. Abdullah, "Practical Capacitor Bank Location Optimization Based On Genetic Algorithm," Int. J. Eng. Adv. Tech.Studies, vol.3, 2015, pp.23-31. ISSN 2053-5791 (online).

[9] A.A. El-Fergany and A. Y. Abdelaziz, "Artificial Bee Colony Algorithm to Allocate Fixed and Switched Static Shunt Capacitors in Radial Distribution Networks," Electr. Power Compon. Syst., vol.42, 2014, pp. 427– 438.
https://doi.org/10.1080/15325008.2013.856965

[10] Zhao, W.G., Wang, L.Y., Zhang, Z.X. (2019). Artificial ecosystem-based optimization: A novel nature-inspired meta-heuristic algorithm. Neural Comput & Applic., 32: 9383-9425. https://doi.org/10.1007/s00521-019-04452-x

[11] HEIDARI, Ali Asghar, MIRJALILI, Seyedali, FARIS, Hossam, et al. Harris hawks optimization: Algorithm and applications. Future generation computer systems, (2019), vol. 97, p. 849-872.
https://doi.org/10.1016/j.future.2019.02.028

[12] Mirjalili, S. Moth-flame optimization algorithm: a novel nature-inspired heuristics paradigm. Knowl Based Syst **89**, 228–249 (1989).
https://doi.org/10.1016/j.knosys.2015.07.006

[13] Mohamed, A. A., Kamel, S., Selim, A., Khurshaid, T., & Rhee, S. B. (2021). Developing a hybrid approach based on analytical and metaheuristic optimization algorithms for the optimization of renewable dg allocation considering various types of loads. Sustainability(Switzerland),13(8).
https://doi.org/10.3390/su13084447

[14] P. Parmar, and C. Patel, "Optimal palcement of capacitor using backward forward sweep method," Springer Nature Singapore Pte Ltd., 2018. pp. 255– 262.
https://doi.org/10.1007/978-981-10-6977-2_23

[15] REZAEI, Farshad, SAFAVI, Hamid R., ABD ELAZIZ, Mohamed, et al. GMO: geometric mean optimizer for solving engineering problems. Soft Computing, 2023, vol. 27, no 15, p. 10571-10606.
https://doi.org/10.21203/rs.3.rs-2052464/v1

[16] Marini, Federico, and Beata Walczak. "Particle swarm optimization (PSO). A tutorial." Chemometrics and Intelligent Laboratory Systems 149 (2015): 153-165.

Selecting an Appropriate Deep Learning Algorithm for Predictive Maintenance

1st Djahida Belayadi
National Higher school of Cyber security (ENSCS)
National Higher School of Advanced Technologies (ENSTA)
Algiers, Algeria
0000-0003-2392-8146

2nd Asma Djaghlouf
National Higher school of Cyber security (ENSCS)
Algiers, Algeria
asma_djaghlouf@enscs.edu.dz

3rd Hadjer Laribi
Department of Industrial Engineering & Maintenance
National Higher School of Advanced Technologies (ENSTA)
Algiers, Algeria
o_meliani@ensta.edu.dz

4th Ouezna Meliani
Department of Industrial Engineering & Maintenance
National Higher School of Advanced Technologies (ENSTA)
Algiers, Algeria
h_laribi@ensta.edu.dz

***Abstract*—Presently, the realm of research is witnessing a burgeoning interest in the utilization of Deep Learning (DL), a branch of machine learning (ML) and artificial intelligence (AI), within the domain of the Industrial Internet of Things. The prediction of failures of industrial equipment before their occurrence, referred to as Predictive Maintenance (PdM), is also a very trendy topic, primarily driven by its potential for cost-savings and numerous advantages over other maintenance types. DL techniques have the potential to provide significant benefices in the context of PdM applications, mainly with highly complex, non-linear, and unlabeled data in real-world scenarios. However, there are a variety of DL algorithms available, and each each possessing distinct characteristics and can be used in different cases. Therefore, choosing the best DL algorithm to resolve PdM issues is not not trivial. This paper aims to improve the performance of PdM by selecting the most suitable DL algorithms. By conducting a comparative study, three deep learning algorithms were chosen, using the criteria commonly utilized in research articles.**

***Index Terms*—Industry 4.0, Deep Learning, Neural Networks, Predictive Maintenance, Algorithm Selection**

I. Introduction

The industry has increasingly focused on artificial intelligence and machine learning methods, because of their ability to design automatic models that handle the massive amount of data currently collected, which is growing exponentially. The research trend of machine learning has switched towards more intricate models, notably deep learning, due to their enhanced accuracy in handling extensive data sets. The advancement of computing power, particularly with the evolution of GPUs, has played a significant role in the development of these methods. As a result, these models have made remarkable progress in various domains such as intrusion detection systems, computer vision, and language processing, consistently achieving state-of-the-art results. [1].

Today, maintenance is a strategic factor to ensure high productivity of industrial systems; however, due to economic pressures, companies are compelled to reduce maintenance's costs, leading to severe implications for long-term reliability. Consequently, it has become crucial for companies to devise suitable maintenance strategies that guarantee efficient operation of production plants in terms of both quality and availability. As a result, the definition and practices of maintenance have undergone significant transformations over time, largely due to substantial contributions from research endeavors [2].

Currently, there is a transition towards the fourth revolution known as Industry 4.0, which is characterized by the integration of cyber-physical systems and the industrial Internet of Things. It involves the utilization of software, sensors, and intelligent control units to enhance industrial processes and meet their requirements. These advanced techniques facilitate automated predictive maintenance functions by analyzing vast volumes of process and associated data through condition monitoring (CM) [3]. Predictive maintenance (PdM) initially conceived in the late 1940s [4], and has become a promising approach, providing solutions for the remaining life of equipment through the prediction of data collected by various sensors on equipment [5]. It has reached critical significance for industries due to the growing intricacy of interactions among various production activities within expansive manufacturing ecosystems. Over time, DL methods have been proposed for PdM. These methods, such as AutoEncoders, Recurrent Neural Networks, Convolutional Neural Networks, and Long Short-Term Memory, aim to predict equipment failures by automatically learning from historical data and estimating the probability of future failures [6]. In this context, numerous DL algorithms are available. Therefore, the research question that guides this study is: What are the best deep learning algorithms to resolve predictive maintenance issues?

With this aim, the article is structured into five sections, starting with an introduction in Section **??**, Section II provides a detailed description of predictive maintenance, including its approaches and the various monitoring techniques employed

Identify applicable funding agency here. If none, delete this.

in PdM. Following that, Section III represents the principle of deep learning, at the end of this section, we cited deep learning algorithms with brief definitions and illustrations. Then Section IV focuses on the application of DL in PdM, exploring different research experiences in this field. In Section V, a comparison of DL algorithms is conducted using commonly utilized criteria found in research articles, aiming to identify the most effective algorithms for solving PdM problems. Finally, Section VI concludes the research work.

II. Predictive Maintenance

Industrial Maintenance was defined by the French Association for Standardization (AFNOR) in 1994 (standard NFX60-010) as "All actions allowing to maintain or restore a property in a state specified or capable of providing a specified service", this definition was replaced in 2001 (NF EN 13306 X 60-319) with "the combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function" [11]. The activities induced today are of two types: corrective or preventive. Corrective maintenance is scheduled following a failure. However, these activities do not incorporate a preventive aspect. Based on the adage "prevention is better than cure", preventive maintenance is based on predetermined plan implemented proactively [12].

Predictive maintenance is the most recent kind of maintenance that has gotten the attention of researchers and industries. It is defined according to standard NF EN 13306 X 60- 319: start-ups and specialized companies as "conditional maintenance carried out by following extrapolated forecasts from the analysis and evaluation of significant degradation parameters", it means that one can anticipate and detect the failure before it occurs, thanks to the continuous monitoring [14].

III. Deep Learning

Deep learning, a branch of machine learning, employs artificial neural networks to learn from data. These neural networks are specifically crafted to emulate the behavior of human brain biological neurons. They operate by processing data through multiple layers and employ machine learning algorithms to iteratively fine-tune the weights and biases of each layer, ultimately attaining precise predictions or outputs [24].

A. Definitions

1) Biological neurons: Biological neurons are specialized cells present in the nervous system of living organisms, including humans. Their primary role is to transmit information between cells by utilizing electrical and chemical signals [41]. A biological neuron is composed of a cell body, dendrites, an axon, and synapses. The cell body contains the nucleus and organelles necessary for cell function. Dendrites are short, branch-like extensions that receive input from other neurons or sensory cells. The axon, on the other hand, is a long, slender extension that carries signals away from the cell body to other neurons or to target cells, such as muscles or glands. Lastly, synapses are the junctions between the axon of one neuron and the dendrites or cell body of another neuron [41].
When a biological neuron is activated by an input signal, such as a neuron transmitter released by a nearby neuron, it produces an electrical impulse that propagates along its axon. This impulse can then activate other neurons or target cells at synapses [41].

2) Artificial Neural Networks ANNs: ANNs are a mathematical representation that draws inspiration from the arrangement and operation of neurons in the human brain. They are composed of interconnected nodes (neurons) that are organized in layers, with inputs and outputs [26]. In ANNs, neurons receive input signals from other neurons and use a non-linear mathematical function to compute their output. This output is then transmit to other neurons in the network through weighted connections. During training , the weights of these connections are adjusted to reduce the error between the network's predicted output and the actual output [25].

B. Deep Learning techniques

Various architectures have been devised for deep learning, each with its unique strengths and weaknesses, tailored to specific tasks. The following are the prevalent deep learning architectures:

1) Feedforward Neural Networks (FNNs): The first, most common, and simplest architecture is characterized by a series of stacked neurons arranged in layers. Each layer's neurons are connected to all the neurons in the subsequent layer, transmitting their output as input. However, there are no connections between neurons in previous layers or within the same layer. The architecture consists of an input layer, hidden layers, and an output layer as shown in figure (1)). To train the neural network, it is supplied with pairs of input features and target features, which are used to establish the relationship between them. The network achieves this by minimizing the error it produces and mapping the input data to the desired output [1].

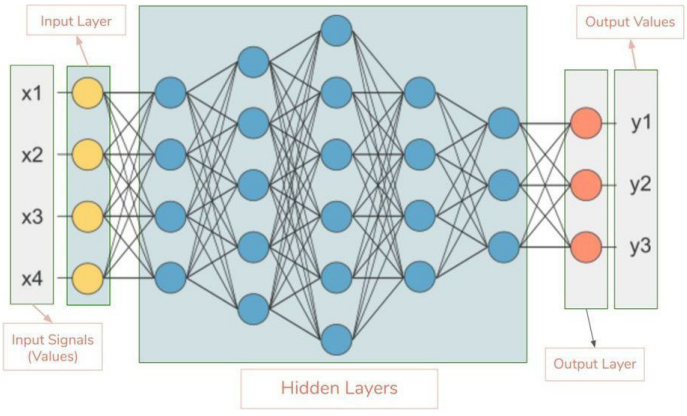


Fig. 1. Feedforward Neural Networks architecture [42]

2) Convolutional Neural Networks (CNNs): CNN is a type of feedforward network that employs convolutional filters to maintain the connectivity between neurons. It draws inspiration from the visual cortex of animals and finds applications in various domains such as image and signal recognition, recommendation systems, and natural language processing (NLP). As shown in figure (2), CNN is composed of the convolution layers which are generally linear and are followed by the application of a function activation, they perform convolution calculations to extract features important. Subsequently, pooling layers are used to reduce the dimensionality of the data. Finally, the classification layers (fully connected) are used to predict the class the output [1].

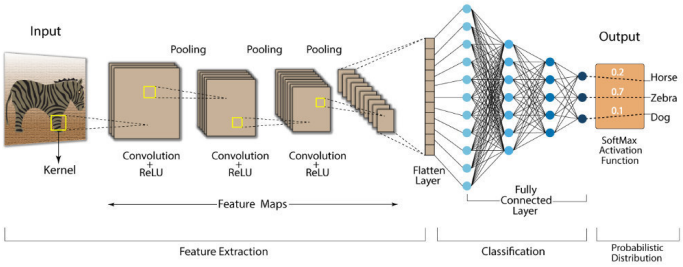


Fig. 2. Convolutional Neural Networks architecture [38]

3) Recurrent Neural Networks (RNNs): RNNs Are a type of neural network that can process sequential data, such as time series or natural language [32]. Contrast to FNNs, RNNs incorporate a feedback loop that allows them to process sequential data by taking into account previous inputs, as is represented in figure (3). As a result, they are well-suited for tasks speech recognition, and language translation [33]. Nevertheless, RNNs come up against the problem of disappearance of the gradient to learn to memorize past events.

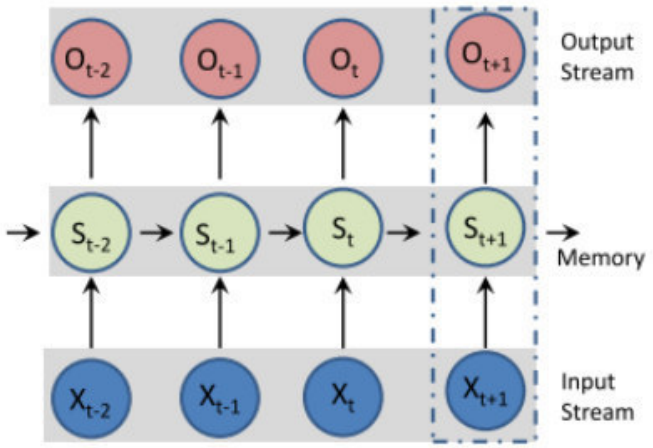


Fig. 3. Recurrent Neural Networks architecture [38]

4) Long Short-Term Memory (LSTM) Networks: LSTMs are derivatives of RNN [48]. They can learn and memorize addictions over a long period. LSTMs thus retain the information stored over the long term. They are especially useful for predicting time series, they remember previous entries. In addition to this use case, LSTMs are also used to compose musical notes and recognize voice [47] [1].

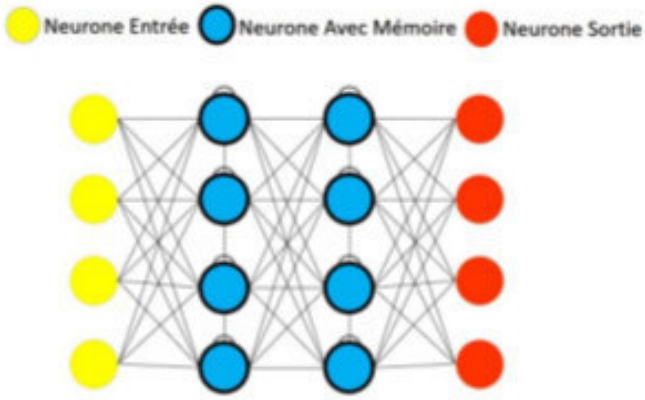


Fig. 4. Long Short-Term Memory architecture [40]

5) AutoEncoders (AE): AE is based on the concept of singular value decomposition to capture the essential nonlinear features of the input data within a reduced dimensional space. As illustrated in figure (5), it consists of two parts: an encoder that maps input data to the encoded, latent space, and the decoder, which projects latent space data to the reconstructed space that has the same dimension as input data. The network is trained to minimize the reconstruction error, quantified as the discrepancy between the input and output [1] [34].

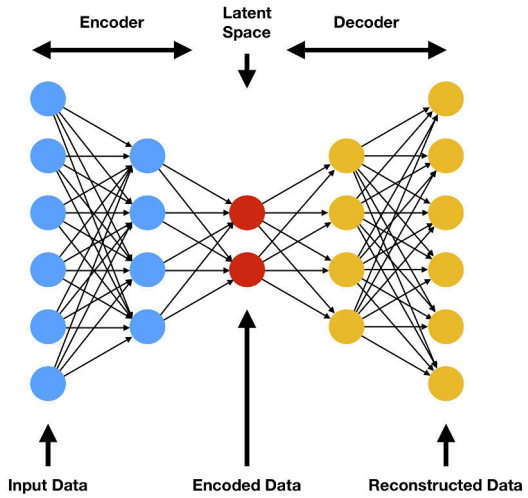


Fig. 5. Autoencoders architecture [38]

6) Generative Adversarial Networks (GANs): The GAN is a model as shown in the figure (6), which includes two neural networks: the generator is responsible for producing synthetic data by taking random noise as input, while the discriminator is responsible for distinguishing between real and synthetic

data by examining the two types [36]. In particular, these networks are trained simultaneously and they compete in an adversarial game [37] [35].

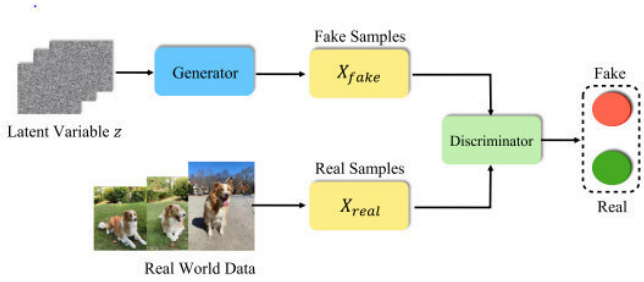


Fig. 6. Generative Adversarial Networks architecture [35]

7) Boltzmann machines restraints (RBM): RBMs are a type of ANNs algorithm characterized by a two-layer structure (figure(7)) comprising a hidden layer and a visible layer. In RBMs, only the hidden and visible layers are connected; however, there is no correlation between the neurons of the same layer. The visible layer receives input data, while the hidden layer extracts relevant features. Moreover, in RBMs, the hidden nodes are self-determining based on certain conditions and do not depend on each other. The weights and biases of the visible and hidden layers are updated iteratively to enable the visible layer to approximate the input data [39] [1] [48].

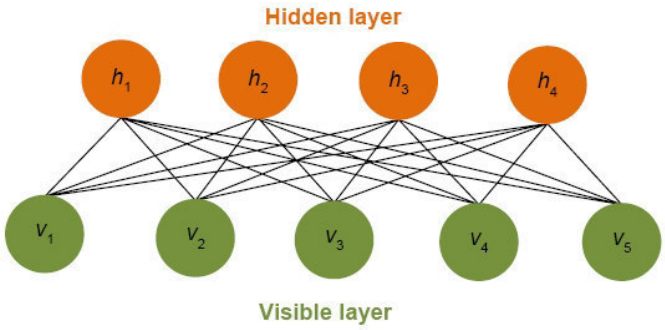


Fig. 7. Boltzmann machines restraints architecture [39]

8) Deep Belief Networks (DBNs): A DBN can be viewed as a collection of RBMs, combined in a way that each RBM's hidden layer is connected to the visible layer of the subsequent RBM. In DBNs, there are undirected connections only between the top two layers, while directed connections exist towards the lower layers. The initialization of a DBN is achieved using a layer-by-layer greedy learning approach that employs unsupervised learning. Subsequently, the network is fine-tuned based on the desired output [38].

IV. Deep Learning for Predictive Maintenance

As systems require greater levels of reliability, availability, maintainability, and safety, traditional maintenance strategies are losing their effectiveness and becoming outdated. Furthermore, the emergence of Industry 4.0 has created more

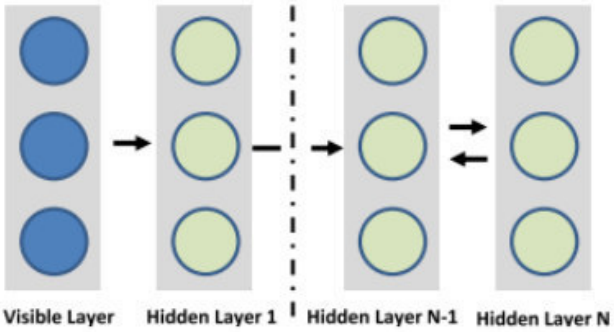


Fig. 8. Deep Belief Networks architecture [38]

convenient avenues for implementing predictive maintenance (PdM) on a broader scale [43].

An instance would be the application of intelligent sensors that offer a dependable approach to real-time system monitoring. By leveraging this data, managers can optimize maintenance operations and minimize machine downtimes, ultimately enhancing production flow [43].

A. Deployment of predictive maintenance

The suggested predictive maintenance deployment scheme takes advantage of a broad spectrum of emergent technologies, such as IoT, Machine learning (ML), and Deep learning (DL), and involves a series of sequential actions, namely sensors installation, data analysis and processing, data visualization, and decision making, as depicted in the accompanying figure 9.

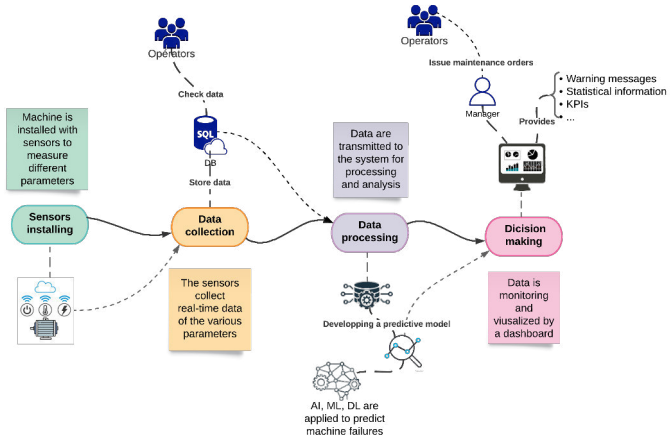


Fig. 9. Deployment of predictive maintenance [8]

To initiate predictive maintenance, real-time data from various parameters including temperature, pressure, and vibration are collected through sensors installed on the equipment. These data are then checked, stored in databases, and subjected to data cleaning, transformation, and reduction. The output of this process is utilized in implementing data

TABLE I
SUMMARY OF SELECTED ARTICLES IN THE LITERATURE ON DEEP LEARNING WITH PREDICTIVE MAINTENANCE

Title	Author(s)	Technique	Description
Forecasting Appliances Failures: A Machine-Learning Approach to Predictive Maintenance [49]	Fernandes et al	LSTM	develop an LTSM model for predicting severe faults in boilers of the HVAC system
A recurrent neural network based health indicator for remaining useful life prediction of bearings [50]	Guo et al	LSTM-based RNN	construct an LSTM-RNN-based health indicator to predict remaining useful life (RUL) of bearings.
Predicting remaining useful life using time series embeddings based on recurrent neural networks [51]	Gugulothu et al	RNN	create an embedded structure that captures machine degradation patterns for time to failure (TTF) estimation.
A bidirectional LSTM prognostics method under multiple operational conditions [52]	Huang et al	LSTM	developed a method for machine health prognostics under multiple operational conditions
A data-driven fault detection and diagnosis scheme for air handling units in building HVAC systems considering undefined states [53]	Yun et al	SAE	Conduct a data-driven based model approach for fault detection and diagnosis of FDD of an AHU
Motor Fault Detection and Feature Extraction Using RNN-Based Variational Autoencoder [54]	Huang & Chen	AE with LSTM	A motor experiment platform were designed to simulate 15 fault scenarios and to collect the time domain vibration signals at 3 different locations.
A Deep Learning Approach for Fault Diagnosis of Induction Motors in Manufacturing [55]	Shao et al	DBN	A deep learning model based on DBN for induction motor fault diagnosis.
Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving [56]	Guo et al	DBN	establish fault diagnosis model based on the DBN model
A Deep Learning Model for Predictive Maintenance in Cyber-Physical Production Systems Using LSTM Autoencoders [57]	Bampoula et al	AE LSTM	Assess the operational condition of production equipment, using a deep learning method for anomaly detection that is then mapped to different RUL values.
A semi-supervised approach to fault detection and diagnosis for building HVAC systems based on the modified generative adversarial network [58]	Cheng et al	semi-supervised FDD approach	Use a semi-supervised FDD approach for constructing an HVAC system using modified GAN.
A novel deep autoencoder feature learning method for rotating machinery fault diagnosis [59]	Shao et al	AE	Presents a methodology of AE optimization for rotating machinery fault diagnosis. The model was applied to fault diagnosis of gearbox and roller bearings.

analysis with ML/DL algorithms to create a predictive model. A dashboard is used to monitor and visualize the outcomes of the preceding stage. Finally, the manager assists the technicians during maintenance interventions by providing advice and orders [44] [45] [46].

B. Related research

Recently, the field of Predictive Maintenance (PdM) has witnessed a surge in the application of deep learning techniques, such as Auto-Encoders (AE), Convolutional Neural Networks (CNN), and Deep Belief Networks (DBN). These models have demonstrated remarkable proficiency in feature learning, fault classification, and fault prediction due to their multi-layer nonlinear transformations [44]. In the following table (table I), we present a literature review of some among various selected research that combines deep learning with predictive maintenance:

V. Comparative Study between Deep Learning algorithms

As outlined in figure 10, The key criteria utilized in assessing a DL algorithm, are:

1) **Accuracy**: represents the ratio of the number of correct predictions to the total number of samples in a data set [76] [15].
   Accuracy: Low [¡ 25%]* / Medium [25%-50%[** / High [50%-75%[*** / Very High [75%-100%]* ***
2) **Learning time**: is the time associated with training the dataset and varies depending on the size of the data and the algorithm we are using [75] [15].
   Learning time: Average*/ Good** / Very good*** / Excellent****
3) **Prediction time**: refers to the time it takes for an algorithm to generate a prediction or estimate from input data, and it varies depending on the size of the data and the algorithm used [75] [15].
   Prediction time: Average*/ Good** / Very good*** / Excellent****
   .
4) **Research works**: Consists of the number of articles that use one of the DL algorithms in the field of PdM.
   Number of papers: Very small [¡5]* /Small [5-8]**/ large]8-12]*** /Very large]12-16]****
5) **Parameterization**: are the values to be adjusted when configuring an algorithm [75] [15].
   Number of parameters: Large [¿8]* /Medium [4-8]**

TABLE II
COMPARISON BETWEEN DL ALGORITHMS ACCORDING TO CRITERIA

Criteria/Algorithm	CNN	RNN	LSTM	AE	GAN	FNN	RBM	DBN
Accuracy	****	****	****	***	***	***	***	****
Learning time	***	**	**	***	*	**	****	*
Parameterization	**	**	**	****	**	**	****	**
Memory size	***	***	****	**	**	**	**	**
Prediction time	**	****	****	***	***	***	*	**
Flexibility	****	****	****	**	**	**	****	**
Interpretation facilities	**	*	*	*	*	*	*	*
research works	***	***	****	***	**	*	**	**
Score	23	23	25	21	16	16	21	16

(**** represent the best and * the worst performance))

TABLE III
ALGORITHM SCORES IN DESCENDING ORDER

No	Algorithms	Score
1	**Long Short-Term Memory**	**25**
2	**Recurrent Neural Networks**	**23**
3	**Convolutional Neural Networks**	**23**
4	Autoencoders	21
5	Boltzmann machines rerestraint	21
6	Generative Adversarial Networks	16
7	Deep Belief Networks	16
8	Feedforward Neural Networks	16

/Small [2-4]*** /Very small [¡2]****

6) **Memory size**: is the capacity required to hold the parameters of the model, including the weights and biases of the various layers, the data and the variables [75] [15].
   Memory Size: Very Large* /Large** /Small*** /Very Small****
7) **Flexibility**: refers to the ability of a network to adapt to database models [75] [15].
   Flexibility: Low [¡25%]*/ Medium [25%-50%]** / High [50%-75%]*** / Very High [75%-100%]* ***
8) **Interpretation facilities**: refers to the ability to understand and explain how a model makes decisions or predictions [77] [15].
   Interpretation facilities: Difficult* /medium** /Easy *** /Very easy ****

The Mind Map in figure 10, summarizes the criteria for evaluating DL algorithms identified in the state-of-the-art articles.

Below is a comparative table between the different algorithms

After assessing the performance of the deep learning algorithms, the ratings are added up to determine which algorithms should be selected:

Based on the previous table, the LSTM algorithm claimed the first position, followed by the RNN and CNN algorithms.

The study's findings suggest that among the various algorithms considered for PdM, these three particularly excel and can be considered the top performers according to the selected criteria and papers.

VI. CONCLUSIONS

Incorporating automated data-driven predictive maintenance models can enable industrial companies that currently depend on corrective and periodic maintenance strategies to achieve cost optimization.

Predictive maintenance is an important application of ML and DL algorithms that can help organizations in minimizing system failures, enhancing predictability, and boosting system availability. With the rise in industrial data spaces globally, DL solutions have gained popularity for PdM. However, selecting the appropriate architecture for each use case can be challenging due to the numerous factors involved.

The primary aim of this article is to identify an appropriate DL algorithm for PdM, knowing that this selection requires careful consideration of the specific problem and the data at hand. Upon reviewing various DL algorithms, it becomes apparent that each algorithm possesses its own set of strengths and weaknesses with respect to precision, training duration, and comprehensibility.

The study's findings indicate that LSTM achieved the highest score for the set of criteria analyzed, followed by RNN and CNN. These three DL algorithms are regarded as the most effective for PdM. However, it should be emphasized that the performance of these algorithms is heavily influenced by the quality and quantity of training data. As a result, it is critical to establish a strong data collection and pre-processing plan to achieve the best results.

Future work will be dedicated to setting up an experimental study on the algorithms. This study will involve utilizing either real or simulated data-sets related to PdM. By focusing on the findings in terms of precision, recall, accuracy, or other appropriate performance measures, the most appropriate algorithm for PdM tasks will be selected.

REFERENCES

[1] Serradilla, O., Zugasti, E., Rodriguez, J., Zurutuza, U. (2022). Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. Applied Intelligence, 52(10), 10934-10964.

[2] Alaoui Belghiti, A. (2021). Prédiction de situations anormales par apprentissage automatique pour la maintenance prédictive: approches en transport optimal pour la détection d'anomalies (Doctoral dissertation, Université Paris-Saclay

[3] Dusko Lukac. 2016. The fourth ICT-based industrial revolution "industry 4.0" - HMI and the case of CAE/CAD innovation with EPLAN P8. In 2015 23rd Telecommunications Forum, TELFOR 2015. IEEE, 835–838. https://doi.org/10.1109/TELFOR.2015.7377595.

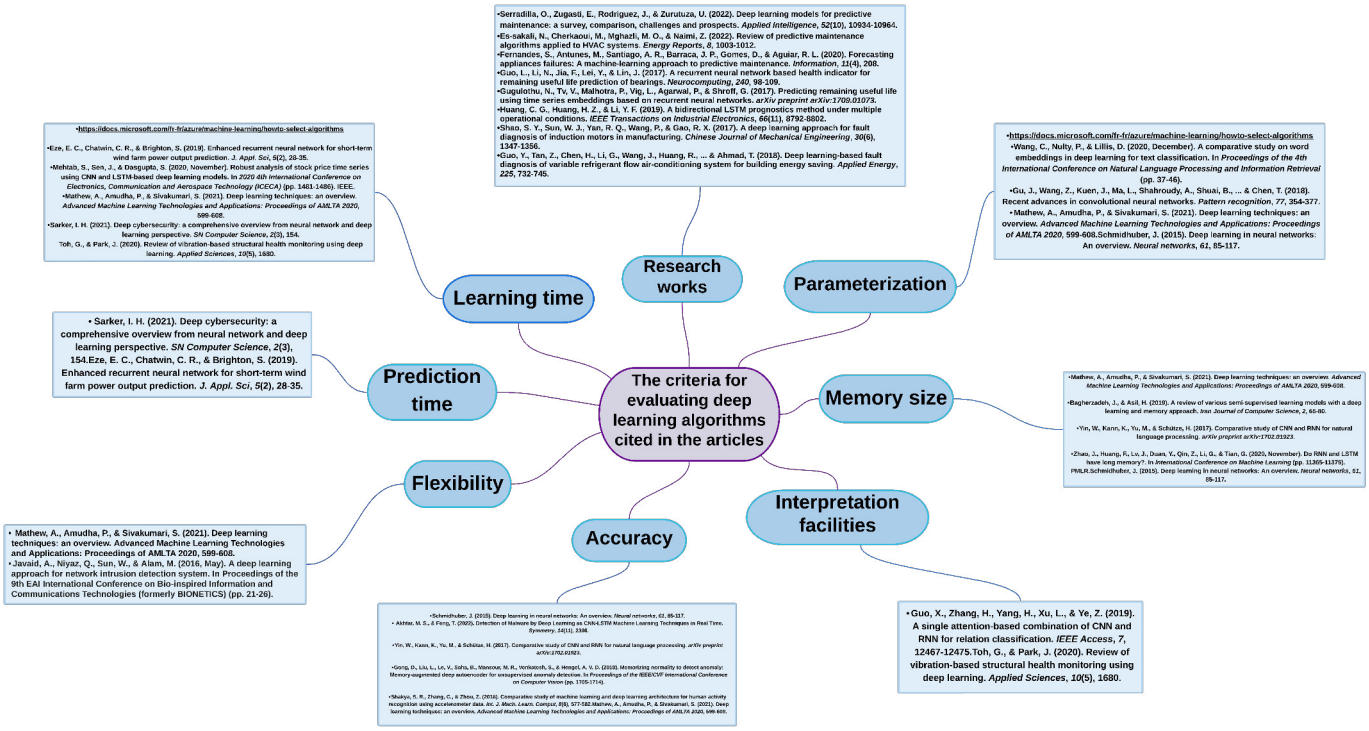


Fig. 10. Criteria for evaluating DL algorithms from the state of the art.

[4] Prajapati, A.; Bechtel, J.; Ganesan, S. Condition based maintenance: A survey. J. Qual. Maint. Eng. 2012, 18, 384–400.

[5] Tiddens, W.; Braaksma, J.; Tinga, T. Exploring predictive maintenance applications in industry. J. Qual. Maint. Eng. 2020, 28, 68–85.

[6] Davari, N., Veloso, B., Ribeiro, R. P., Pereira, P. M., Gama, J. (2021, October). Predictive maintenance based on anomaly detection using deep learning for air production unit in the railway industry. In 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 1-10). IEEE

[7] Kadir, B. A. (2020). Designing new ways of working in Industry 4.0: Aligning humans, technology, and organization in the transition to Industry 4.0.

[8] Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., Adda, M. (2022). On predictive maintenance in industry 4.0: Overview, models, and challenges. Applied Sciences, 12(16), 8081.

[9] Hermann, M., Pentek, T., Otto, B. (2015). Design principles for Industrie 4.0 scenarios: a literature review. Technische Universität Dortmund, Dortmund, 45.

[10] Maresova, P., Soukal, I., Svobodova, L., Hedvicakova, M., Javanmardi, E., Selamat, A., Krejcar, O. (2018). Consequences of industry 4.0 in business and economics. Economies, 6(3), 46.

[11] FD, X. (2002). 60-000, Norme Française. Fonction Maintenance. AFNOR.

[12] Mattioli, J., Robic, P. O., Reydellet, T. (2018, July). L'intelligence artificielle au service de la maintenance prévisionnelle. In 4ème conférence sur les Applications Pratiques de l'Intelligence Artificielle APIA2018.

[13] MAGHLOUT, H. (2010). Application de l'analyse en ondelettes discrète dans la maintenance prédictive des paliers à roulements.

[14] https://biblio.univ-annaba.dz/ingeniorat/wp-content/uploads/2019/10/Arab-Mustapha.pdf

[15] Ouadah, A., Zemmouchi-Ghomari, L., Salhi, N. (2022). Selecting an appropriate supervised machine learning algorithm for predictive maintenance. The International Journal of Advanced Manufacturing Technology, 119(7-8), 4277-4301.

[16] https://blog.infraspeak.com/condition-monitoring-techniques/

[17] BENCHETTOUH, S. E. (2019). Elaboration d'un système de prédiction des pannes et de planification des maintenances (Doctoral dissertation, UNIVERSITE MOHAMED BOUDIAF-M'SILA-FACULTE MATHEMATIQUES ET DE L'INFORMATIQUE).

[18] Djebili, O. (2013). Contribution à la maintenance prédictive par analyse vibratoire des composants mécaniques tournants. Application aux butées à billes soumises à la fatigue de contact de roulement (Doctoral dissertation, Reims).

[19] Es-sakali, N., Cherkaoui, M., Mghazli, M. O., Naimi, Z. (2022). Review of predictive maintenance algorithms applied to HVAC systems. Energy Reports, 8, 1003-1012.

[20] Cao, Q., Zanni-Merk, C., Samet, A., Reich, C., De Beuvron, F. D. B., Beckmann, A., Giannetti, C. (2022). KSPMI: a knowledge-based system for predictive maintenance in industry 4.0. Robotics and Computer-Integrated Manufacturing, 74, 102281.

[21] Jimenez, J. J. M., Schwartz, S., Vingerhoeds, R., Grabot, B., Salaün, M. (2020). Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. Journal of Manufacturing Systems, 56, 539-557.

[22] Sayyad, S., Kumar, S., Bongale, A., Bongale, A. M., Patil, S. (2021). Estimating remaining useful life in machines using artificial intelligence: A scoping review. Libr. Philos. Pract, 2021, 1-26.

[23] Gawde, S., Patil, S., Kumar, S., Kotecha, K. (2022). A scoping review on multi-fault diagnosis of industrial rotating machines using multi-sensor data fusion. Artificial Intelligence Review, 1-54.

[24] Chen, J., Ran, X. (2019). Deep learning with edge computing: A review. Proceedings of the IEEE, 107(8), 1655-1674.

[25] Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., ... Kiru, M. U. (2019). Comprehensive review of artificial neural network applications to pattern recognition. IEEE Access, 7, 158820-158846.

[26] Krenker, A., Bešter, J., Kos, A. (2011). Introduction to the artificial neural networks. Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech, 1-18.

[27] https://levity.ai/blog/difference-machine-learning-deep-learning

[28] https://nixustechnologies.com/deep-learning-vs-machine-learning/

[29] https://www.scs.org.sg/articles/machine-learning-vs-deep-learning

[30] https://www.apriorit.com/dev-blog/472-machine-learning-applications

[31] https://dev.to/bpb-online/machine-learning-vs-deep-learning-what-s-the-difference-2bm7

[32] Gheisari, S., Shariflou, S., Phu, J., Kennedy, P. J., Agar, A., Kalloniatis, M., Golzan, S. M. (2021). A combined convolutional and recurrent neural network for enhanced glaucoma detection. Scientific reports, 11(1), 1-11.
[33] Apaydin, H., Feizi, H., Sattari, M. T., Colak, M. S., Shamshirband, S., Chau, K. W. (2020). Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. Water, 12(5), 1500.
[34] Makhzani, A., Frey, B. (2013). K-sparse autoencoders. arXiv preprint arXiv:1312.5663.
[35] Cai, Z., Xiong, Z., Xu, H., Wang, P., Li, W., Pan, Y. (2021). Generative adversarial networks: A survey toward private and secure applications. ACM Computing Surveys (CSUR), 54(6), 1-38.
[36] Ye, H., Liang, L., Li, G. Y., Juang, B. H. (2020). Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels. IEEE Transactions on Wireless Communications, 19(5), 3133-3143.
[37] Zeng, Q., Ma, X., Cheng, B., Zhou, E., Pang, W. (2020). Gans-based data augmentation for citrus disease severity detection using deep learning. IEEE Access, 8, 172882-172891.
[38] Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., Yang, G. Z. (2016). Deep learning for health informatics. IEEE journal of biomedical and health informatics, 21(1), 4-21.
[39] En, T. Y., Ki, M. S., Hui, N. T., Jie, T. J., Vusoff, M. A. B. M. (2018, December). Predictive Maintenance of a Train System Using a Multilayer Perceptron Artificial Neural Network. In 2018 International Conference on Intelligent Rail Transportation (ICIRT) (pp. 1-5). IEEE.
[40] Bensiah, O. A., Berkane, M. (2020). La Proposition d'une nouvelle approche basée Deep learning pour la prédiction du cancer du sain.
[41] https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc
[42] https://afit-r.github.io/feedforward-DNN
[43] Nguyen, K. T., Medjaher, K. (2019). A new dynamic predictive maintenance framework using deep learning for failure prognostics. Reliability Engineering System Safety, 188, 251-262.
[44] Ran, Y., Zhou, X., Lin, P., Wen, Y., Deng, R. (2019). A survey of predictive maintenance: Systems, purposes and approaches. arXiv preprint arXiv:1912.07383
[45] Alves, F., Badikyan, H., Moreira, H. A., Azevedo, J., Moreira, P. M., Romero, L., Leitão, P. (2020, June). Deployment of a smart and predictive maintenance system in an industrial case study. In 2020 IEEE 29th International Symposium on Industrial Electronics (ISIE) (pp. 493-498). IEEE.
[46] Jimenez-Cortadi, A., Irigoien, I., Boto, F., Sierra, B., Rodriguez, G. (2019). Predictive maintenance on the machining process and machine tool. Applied Sciences, 10(1), 224.
[47] Rivas, A., Fraile, J. M., Chamoso, P., González-Briones, A., Sittón, I., Corchado, J. M. (2020). A predictive maintenance model using recurrent neural networks. In 14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019) Seville, Spain, May 13–15, 2019, Proceedings 14 (pp. 261-270). Springer International Publishing
[48] Khalil, R. A., Saeed, N., Masood, M., Fard, Y. M., Alouini, M. S., Al-Naffouri, T. Y. (2021). Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications. IEEE Internet of Things Journal, 8(14), 11016-11040.
[49] Fernandes, S., Antunes, M., Santiago, A. R., Barraca, J. P., Gomes, D., Aguiar, R. L. (2020). Forecasting appliances failures: A machine-learning approach to predictive maintenance. Information, 11(4), 208.
[50] Guo, L., Li, N., Jia, F., Lei, Y., Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing, 240, 98-109.
[51] Gugulothu, N., Tv, V., Malhotra, P., Vig, L., Agarwal, P., Shroff, G. (2017). Predicting remaining useful life using time series embeddings based on recurrent neural networks. arXiv preprint arXiv:1709.01073.
[52] Huang, C. G., Huang, H. Z., Li, Y. F. (2019). A bidirectional LSTM prognostics method under multiple operational conditions. IEEE Transactions on Industrial Electronics, 66(11), 8792-8802.
[53] yun W-S, Hong W-H, Seo H. A data-driven fault detection and diagnosis scheme for air handling units in building HVAC systemsconsidering undefined states. J Build Eng 2021;35:102111
[54] Huang, Y., Chen, C. H., Huang, C. J. (2019). Motor fault detection and feature extraction using RNN-based variational autoencoder. IEEE access, 7, 139086-139096.
[55] Shao, S. Y., Sun, W. J., Yan, R. Q., Wang, P., Gao, R. X. (2017). A deep learning approach for fault diagnosis of induction motors in manufacturing. Chinese Journal of Mechanical Engineering, 30(6), 1347-1356.
[56] Guo, Y., Tan, Z., Chen, H., Li, G., Wang, J., Huang, R., ... Ahmad, T. (2018). Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving. Applied Energy, 225, 732-745.
[57] Bampoula, X., Siaterlis, G., Nikolakis, N., Alexopoulos, K. (2021). A deep learning model for predictive maintenance in cyber-physical production systems using lstm autoencoders. Sensors, 21(3), 972.
[58] Li, B., Cheng, F., Cai, H., Zhang, X., Cai, W. (2021). A semi-supervised approach to fault detection and diagnosis for building HVAC systems based on the modified generative adversarial network. Energy and Buildings, 246, 111044.
[59] Shao, H., Jiang, H., Zhao, H., Wang, F. (2017). A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. Mechanical Systems and Signal Processing, 95, 187-204.
[60] Wang, C., Nulty, P., Lillis, D. (2020, December). A comparative study on word embeddings in deep learning for text classification. In Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval (pp. 37-46).
[61] Yin, W., Kann, K., Yu, M., Schütze, H. (2017). Comparative study of CNN and RNN for natural language processing. arXiv preprint arXiv:1702.01923
[62] Zhao, J., Huang, F., Lv, J., Duan, Y., Qin, Z., Li, G., Tian, G. (2020, November). Do RNN and LSTM have long memory?. In International Conference on Machine Learning (pp. 11365-11375). PMLR.
[63] Belkin, M., Ma, S., Mandal, S. (2018, July). To understand deep learning we need to understand kernel learning. In International Conference on Machine Learning (pp. 541-549). PMLR.
[64] Bagherzadeh, J., Asil, H. (2019). A review of various semi-supervised learning models with a deep learning and memory approach. Iran Journal of Computer Science, 2, 65-80.
[65] Mathew, A., Amudha, P., Sivakumari, S. (2021). Deep learning techniques: an overview. Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020, 599-608.
[66] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.
[67] Sarker, I. H. (2021). Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective. SN Computer Science, 2(3), 154.
[68] Eze, E. C., Chatwin, C. R., Brighton, S. (2019). Enhanced recurrent neural network for short-term wind farm power output prediction. J. Appl. Sci, 5(2), 28-35.
[69] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... Chen, T. (2018). Recent advances in convolutional neural networks. Pattern recognition, 77, 354-377.
[70] Mehtab, S., Sen, J., Dasgupta, S. (2020, November). Robust analysis of stock price time series using CNN and LSTM-based deep learning models. In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1481-1486). IEEE.
[71] Toh, G., Park, J. (2020). Review of vibration-based structural health monitoring using deep learning. Applied Sciences, 10(5), 1680.
[72] Akhtar, M. S., Feng, T. (2022). Detection of Malware by Deep Learning as CNN-LSTM Machine Learning Techniques in Real Time. Symmetry, 14(11), 2308
[73] Gong, D., Liu, L., Le, V., Saha, B., Mansour, M. R., Venkatesh, S., Hengel, A. V. D. (2019). Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1705-1714).
[74] Shakya, S. R., Zhang, C., Zhou, Z. (2018). Comparative study of machine learning and deep learning architecture for human activity recognition using accelerometer data. Int. J. Mach. Learn. Comput, 8(6), 577-582
[75] https://docs.microsoft.com/fr-fr/azure/machine-learning/howto-select-algorithms
[76] Kiangala, K. S., Wang, Z. (2020). An effective predictive maintenance framework for conveyor motors using dual time-series imaging and convolutional neural network in an industry 4.0 environment. Ieee Access, 8, 121033-121049.
[77] https://arxiv.org/abs/2103.10689

Integration of IoT for Effective Leak Detection in WDNs

1st Chaima CHABIRA
Department of Electronics
Laboratory of System Signal Analysis(LASS) ,Faculty of technology,University of M'sila.
M'sila, Algeria
chaima.chabira@univ-msila.dz

2nd Haddi BAKHTI
Department of Electronics
Laboratory of Genie Electric (LGE) ,Faculty of technology, University of M'sila.
M'sila, Algeria
elhadi.bakhti@univ-msila.dz

3rd Miloud BENTOUMI
Department of Electronics
Laboratory of System Signal Analysis(LASS) ,Faculty of technology, University of M'sila.
M'sila, Algeria
miloud.bentoumi@univ-msila.dz

4th Riyadh BENYETTOU
Department of Mechanical Engineering
Laboratory of Materials and Mechanics of Structures (LMMS), Faculty of technology, University of M'sila.
M'sila, Algeria.
riyadh.benyettou@univ-msila.dz

***Abstract*— In recent years, the convergence of the Internet of Things (IoT) has driven innovations in pressure sensor development, enhancing accuracy, responsiveness, and efficiency in data handling. IoT connects sensors in real-time, enabling them to transmit vast amounts of pressure data to centralized systems for analysis and decision-making. The development of IoT, the everincreasing number of internet users, and the advancement of data rates and services have produced enormous amounts of data that can be utilized to enhance service delivery. IoT provides opportunities for both gathering and applying this data. This paper explores relationship between IoT and sensors, emphasizing their roles in optimizing various sectors such as Autonomous Vehicles, Industrial Automation, Smart Homes, Healthcare and Medical, Smart City and Agriculture. By integrating advanced technologies, including Machine Learning (ML) algorithms, we also examine the relationship between pressure sensors and IoT in leak detection within Water Distribution Networks (WDNs).**

Keywords*— *Internet of Things (IoT), Machine Learning (ML), Pressure Sensors, WDNs.

I. INTRODUCTION

Internet of Things (IoT) is defined as a collection of devices that are equipped with special sensors and microchips; thereby, the devices are capable of collecting and sharing data with each other by employing the extensive communication network that they form. That situation creates further optimization opportunities for the application and increases the number of automated operations in the application. These consequences reduce costs and increase efficiency. To achieve these capabilities, IoT utilizes several technologies, such as advanced sensors and actuators, Bluetooth, cellular data, wifi, RFID, NFC, Ethernet, 5G, and cloud computing. IoT, which has been described as a promising innovation [1], has been evolving rapidly over the recent years. IoT technology makes use of many different smart objects that can communicate with and act on each other through a network infrastructure. The refore IoT can be described as a super complex, large scale ecosystem of heterogeneous elements, which serves as a bridge between the physical and cyber world [2]. Indeed, big data analysis tools and algorithms, which include artificial intelligence and ML, are also employed to properly handle the collected data. Besides, due to the interconnectedness that IoT causes, individuals are more vulnerable to cyber-attacks because that connectedness opens additional paths that unauthorized individuals can exploit; thus, advanced methods and protocols are utilized, namely data encryption, secure communication protocols, and robust authentication methods, to maintain the privacy and security of the users. In the future, IoT has the potential to become a cornerstone of the community if the challenges that it brings with itself are eliminated. Indeed, its impact will be felt strongly by agriculture, manufacturing, healthcare, and smart homes. Due to the fast growth and the important benefits associated with IoT, it has been implemented to many different areas of industries and technologies [3]. This paper is organized as follows: Section 2, Research Background, provides an overview of IoT's role in creating data-driven networks, and the use of ML algorithms to enhance IoT applications through predictive accuracy and automation. In section 5, this section explores how IoT technology is implemented to improve processes, enhance productivity, application of IoT in fields such as healthcare, agriculture, manufacturing, and smart cities. In section 6, this section highlights the role of advanced sensors, connectivity options (such as WiFi, Bluetooth, and 5G), cloud computing, edge computing, and data analytics. In section 7 we explore the critical role of pressure sensors within the IoT framework. This section examines how pressure sensors collect real-time data, enabling continuous monitoring and analysis across various applications, such as WDNs.

II. RESEARCH BACKGROUND

This enables large scale, long distance automation which makes people's life easier. Sensor data gathering and tag identification are two examples of how the IoT has made people's lives easier. The organic integration of agents, intelligent control, and the communication link between the two has altered people's present production and lifestyle. The physical environment can then be transformed into a completely controlled intelligent system [4]. Many applications can be considered with this technology such as smart parking, smart lighting, smart road and congestion control, waste management, etc [5]. In summary IoT is part of life from small scale technologies such as homes to large scale industries like health, city management, food production etc. and this involvement is producing enormous amounts of data that requires storage and processing so that they can be used to further develop mentioned automatic systems.

A. Internet of Things (IoT)

The The architecture of IoT systems is typically divided into six layers: Coding, Perception, Network, Middleware,

Business and Application layers, Fig 1 displays the typical structure of a system that utilizes IoT.

- The **Coding Layer** involves the development and implementation of software applications through programming languages, frameworks, and development tools, enabling interaction with hardware and other software components.
- The **Perception Layer** involves collecting data from the physical environment through sensors, RFID devices, and gateways.
- The **Network Layer** facilitates communication between devices using wired and wireless networks connected through the internet.
- The **Middleware Layer** serves as an intermediary that facilitates communication and data management between applications and databases, providing essential services such as messaging, authentication, and API management to ensure seamless integration and interaction among different systems.
- The **Application Layer** processes, analyzes, and applies the collected data in cloud computing, web services, and analytical services.
- The **Business Layer** focuses on the processing and management of business logic, rules, and operations, utilizing the data and services from the underlying layers to create value through business processes, customer interactions, and decision-making analytics in various domains.

III. Importance of IoT Today's Society

IoT enhances efficiency and innovation across agriculture, manufacturing, and healthcare. It automates irrigation, improves livestock management, enables autonomous machinery, optimizes supply chains, and transforms patient care with real-time monitoring and automation. IoT drives sustainability, productivity, and smarter decision-making in these fields, as shown in Fig 2.

IV. Application of IoT

In this section, a review of numerous IoT big data analytics solutions for different domains are given. The selected domains are Autonomous Vehicles, Industrial Automation, Smart Homes, Healthcare and Medical, Smart City, and Agriculture.

A. Autonomous Vehicles

The IoT is transforming the automotive industry by advancing autonomous vehicles, enhancing safety, efficiency, and passenger experiences. As technology evolves, IoT applications in transportation are set to expand, enabling a more connected and automated future.

B. Industrial Automation

IoT in industrial automation enhances efficiency, safety, and cost-effectiveness by enabling smart, connected systems. It drives Industry 4.0 with data-driven decision-making, automation, and real-time control in modern manufacturing processes.

C. Smart Homes

IoT-enabled smart home systems enhance convenience, energy efficiency, safety, and comfort by automating functions and connecting devices. They transform homes into intelligent spaces, allowing remote control and shaping the future of residential living [6, 7].

D. Healthcare and Medical

The application of IoT in healthcare is transforming the industry by enabling real-time monitoring, remote care, and efficient data management. Overall, IoT in healthcare improves patient care, enhances operational efficiency, and enables proactive health management [8].

E. Smart City

The IoT has become a transformative force in the development of smart cities, enabling various applications that enhance urban living, improve efficiency, and promote sustainability. By integrating these IoT applications, smart cities can improve quality of life, enhance sustainability, and create more efficient urban environments.

F. Agriculture

The IoT is revolutionizing agriculture by enhancing productivity, efficiency, and sustainability. By leveraging these IoT applications, agriculture can become more sustainable, efficient, and capable of meeting the growing demands of the global population [9].

V. Pressure sensors and IoT Relationship (Application)

In this section, a review of the relationship between pressure sensors and the IoT will be conducted. Pressure sensors are critical components in various IoT applications, as they enable real-time monitoring and data collection across multiple domains.

A. Recent Advances and Technologies

Pressure sensors have significantly advanced in terms of sensitivity, miniaturization, and connectivity. Modern pressure transducers can be easily integrated into IoT networks and provide accurate measurements across a variety of conditions. Technological innovations, such as wireless pressure sensors, enable the deployment of sensors in challenging locations, thereby promoting the development of smart environments.

B. Challenges and Hurdles in Integrating Pressure Sensors with IoT

Despite their potential, the implementation of pressure sensors in IoT systems faces several challenges. These include issues related to data accuracy, transducer calibration, and environmental factors that can affect readings. Additionally, security concerns arise from the vast amounts of big data transmitted, as unauthorized access can lead to compromised systems and data integrity.

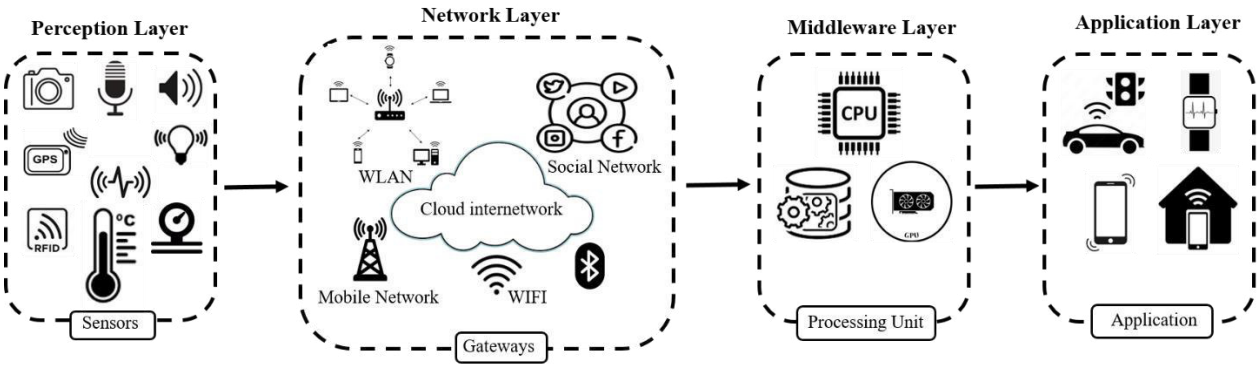


Fig. 1. Internet of Things IoT Architecture.

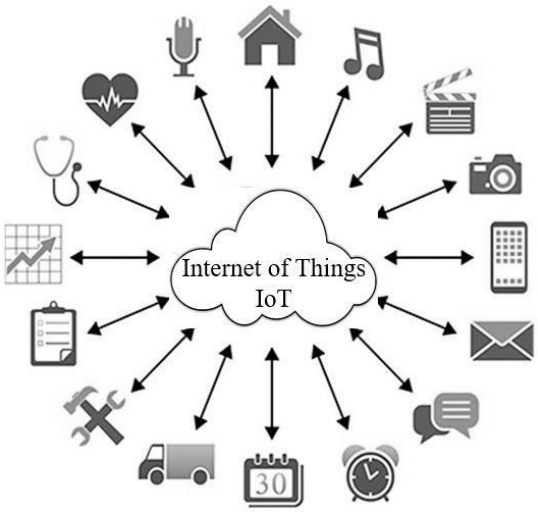


Fig. 2. The Role of IoT in Contemporary Society

C. *Key Applications of Pressure Sensors in IoT*

The efficient utilization of pressure sensors in IoT systems is demonstrated by several examples. Pressure sensors, for instance, are used in IoT applications for WDNs, playing a key role in leak detection, demand management, and pipeline health monitoring. They detect pressure fluctuations to identify leaks and prevent water loss, while also helping to optimize water distribution by tracking consumption patterns. These sensors improve pipeline maintenance by detecting pressure anomalies like surges or drops caused by blockages. They contribute to energy optimization by providing data to operate pumps more efficiently, reducing costs. Real-time monitoring through IoT ensures timely alerts and enhances decision-making. Additionally, pressure sensors support hydraulic modeling and prevent contamination by detecting backflows, ensuring better water quality and system sustainability.

D. *The Future of Pressure Sensors in IoT Systems*

The future of pressure sensors in IoT systems for WDNs is promising, with advancements in accuracy, sensitivity, and real-time data processing. These sensors will enhance predictive maintenance, enabling early issue detection and reducing downtime. The integration of AI and machine learning and IoT will improve data analysis for smarter decision-making, precise demand forecasting, and automated control systems. In the long run, pressure sensors will support the development of smart water grids, optimizing water usage, infrastructure management, leading to more sustainable and efficient water systems.

VI. Comparison with other works

The Table 1 summarizes other works and their performance in the use of IoT and sensor technologies across various fields, including medical, agriculture, and WDNs. Each study demonstrates how IoT and specific sensors enhance detection and monitoring. In the medical sector, sensor-equipped devices aid in real-time fall and heart attack detection, potentially saving lives through continuous monitoring. In agriculture, IoT-based systems help optimize crop management, increasing yield and reducing waste. Our study highlights IoT's impact in WDNs, where pressure sensors aid in early leak detection, ultimately conserving water and ensuring faster repairs.

TABLE I. APPLICATIONS OF IOT AND SENSOR TECHNOLOGIES IN VARIOUS SECTORS

Ref	Domain	Application	Sensors	Technology Used	Focus of the Review
[10]	Medical	• Fall Detection	• Accelerometer Sensors	Computer vision, radar	Overview of DL methods CNN LSTM Auto-encoder for fall detection
[11]		• Fall Detection	• Wearable sensors	threshold methods, traditional ML methods, DL methods,	A comprehensive review of the wearable sensors for fall detection systems
[12]		• Heart Attack Detection	• Heartbeat sensor. • Temperature sensor. • Pulse sensor.	Microcontroller, LCD display, IoT, Wi-Fi connectivity	The system aims to reduce the risk of heart attack fatalities through continuous health monitoring
[13]	Agriculture	• Precision Agriculture for enhancing crop yield	• Wireless Sensor Networks • Ground sensors • Crop monitoring sensors • Environmental sensors	Predictive analytics, WSN, IoT	The paper discusses the role of IoT and WSN in transforming agriculture by enabling precise crop monitoring, reducing resource waste, improving crop yield
Our Study	WDNs	• Leak detection in WDNs	• pressure sensors	Data Analytics and ML, IoT	The integration of pressure sensors with IoT in WDNs enhances leak detection by providing real-time monitoring, reducing water loss, and enabling quicker response times.

ACKNOWLEDGMENT

This study receives support from University research and training projects (PRFU) Project-N° 154/SDRF/2024 organized by the Algerian Ministry of Higher Education and Scientific and Research. The authors express gratitude to the general management of teachers and training and the management of doctoral training to support this work.

REFERENCES

[1] Cui Wei, WITHDRAWN: 5G-oriented IOT coverage enhancement and physical education resource management. 2020, Elsevier.

[2] Aman Ahi and Ajay Vikram Singh. Role of distributed ledger technology (dlt) to enhance resiliency in internet of things (iot) ecosystem. in 2019 Amity International Conference on Artificial Intelligence (AICAI). 2019. IEEE.

[3] Maria Stoyanova, Yannis Nikoloudakis, Spyridon Panagiotakis, Evangelos Pallis, and Evangelos K Markakis, A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues. IEEE Communications Surveys & Tutorials, 2020. 22(2): p. 1191-1221.

[4] Zhihan Lv, Ranran Lou, Jinhua Li, Amit Kumar Singh, and Houbing Song, Big data analytics for 6G-enabled massive internet of things. IEEE Internet of Things Journal, 2021. 8(7): p. 5350-5359.

[5] Laith Farhan, Rupak Kharel, Omprakash Kaiwartya, Marcela Quiroz-Castellanos, Ali Alissa, and Mohamed Abdulsalam. A concise review on Internet of Things (IoT)-problems, challenges and opportunities. in 2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP). 2018. IEEE.

[6] Ashish Tripathi, Anand Bhushan Pandey, Arun Kumar Singh, Pardeep Malik, Krishna Kant Singh, and Prem Chand Vashist, IoT for Smart Automation and Robot, in Emergence of Cyber Physical System and IoT in Smart Automation and Robotics: Computer Engineering in Automation. 2021, Springer. p. 189-201.

[7] Naser Hossein Motlagh, Siavash H Khajavi, Alireza Jaribion, and Jan Holmstrom. An IoT-based automation system for older homes: A use case for lighting system. in 2018 IEEE 11th conference on service-oriented computing and applications (SOCA). 2018. IEEE.

[8] Joel Jpc Rodrigues, Dante Borges De Rezende Segundo, Heres Arantes Junqueira, Murilo Henrique Sabino, Rafael Maciel Prince, Jalal Al-Muhtadi, and Victor Hugo C De Albuquerque, Enabling technologies for the internet of health things. Ieee Access, 2018. 6: p. 13129-13141.

[9] Dieisson Pivoto, Bradford Barham, Paulo Dabdab Waquil, Cristian Rogério Foguesatto, Vitor Francisco Dalla Corte, Debin Zhang, and Edson Talamini, Factors influencing the adoption of smart farming by Brazilian grain farmers. International Food and Agribusiness Management Review, 2019. 22(4): p. 571-588.

[10] Md Milon Islam, Omar Tayan, Md Repon Islam, Md Saiful Islam, Sheikh Nooruddin, Muhammad Nomani Kabir, and Md Rabiul Islam, Deep learning based systems developed for fall detection: a review. IEEE Access, 2020. 8: p. 166117-166137.

[11] Zhiyuan Jiang, Mohammed Aa Al-Qaness, Al-Alimi Dalal, Ahmed A Ewess, Mohamed Abd Elaziz, Abdelghani Dahou, and Ahmed M Helmi, Fall detection systems for internet of medical things based on wearable sensors: A review. IEEE Internet of Things Journal, 2024.

[12] Amit Kishor and Wilson Jeberson. Diagnosis of heart disease using internet of things and machine learning algorithms. in Proceedings of second international conference on computing, communications, and cyber-security: IC4S 2020. 2021. Springer.

[13] Pallabi Saha, Vikas Kumar, Samta Kathuria, Anita Gehlot, Vikrant Pachouri, and Angel Swastik Duggal. Precision Agriculture Using Internet of Things and Wireless Sensor Networks. in 2023 International Conference on Disruptive Technologies (ICDT). 2023. IEEE.

Leak Detection in WDNs Using Cepstrum Analysis of Pressure Signals

1st Chaima CHABIRA
Department of Electronics
Laboratory of System Signal Analysis(LASS) ,Faculty of technology,University of M'sila.
M'sila, Algeria
chaima.chabira@univ-msila.dz

2nd Haddi BAKHTI
Department of Electronics
Laboratory of Genie Electric (LGE) ,Faculty of technology, University of M'sila.
M'sila, Algeria
elhadi.bakhti@univ-msila.dz

3rd Miloud BENTOUMI
Department of Electronics
Laboratory of System Signal Analysis(LASS) ,Faculty of technology, University of M'sila.
M'sila, Algeria
miloud.bentoumi@univ-msila.dz

4th Riyadh BENYETTOU
Department of Mechanical Engineering
Laboratory of Materials and Mechanics of Structures (LMMS), Faculty of technology, University of M'sila.
M'sila, Algeria.
riyadh.benyettou@univ-msila.dz

***Abstract*— Detecting leaks in pipeline networks, particularly in water distribution systems, is a critical yet challenging task. Numerous techniques have been proposed for leak detection; however, none have proven entirely suitable for industrial applications due to limitations such as low efficiency or high costs. This study focuses on employing cepstrum analysis as a powerful signal processing technique to detect and locate leaks in pipelines based on pressure signal measurements. The proposed method involves preprocessing pressure signals to remove noise using orthogonal wavelets, followed by the application of cepstrum analysis to extract key features indicative of leaks. Earlier studies have demonstrated that this approach can effectively pinpoint the location of leaks and assess their severity in pipeline systems. Experiments were conducted on a controlled fluid-filled pipeline network, where pressure transducers recorded transient pressure drops triggered by simulated leaks. The analysis demonstrated the ability of cepstrum-based technique to detect accurately the location of leaks across varying pipeline distances. Pressure transducers were used to record pressure signals, with leakage initiated via a push-button mechanism acting on a solenoid valve. The transient pressure drop was captured during the recordings, which were conducted at varying pipeline distances from the leakage point.**

This work highlights the potential of cepstrum analysis in advancing leak detection methodologies, offering an effective tool for maintaining the integrity and efficiency of pipeline systems.

Keywords— Leak Detection, WDNs, Pressure Transducer, Cepstrum Analysis.

I. INTRODUCTION

Water is a precious and fundamental resource essential for sustaining life, supporting ecosystems, and driving economic and social development. Clean water is one of the most vital resources for humans to survive [1]. Access to clean, potable water is vital for public health, food production, and industrial processes, making its efficient management a top priority worldwide. WDNs serve as a critical infrastructure system for delivering potable water to millions of Population, that are difficult to manage and monitor due to their size and complexity [2] These networks are composed of extensive pipe systems, Pipe systems are the most cost-effective, efficient, and safest means of supplying potable water to the populace [3].

WDNs is a significant issue in water resource management, Significant amounts of treated water are lost, reducing the efficiency of water supply systems , and this loss is mainly due to pipe leakage[4]. Water leakage in distribution networks can lead to a range of negative consequences, including: Economic Impact and Environmental Damage, Damage to roads, buildings, and other facilities near the leak site, which lead to the invasion of external pollutant contamination and bacterial through leak holes [5]. Compromising water quality and posing risks to public health. Mitigating water leakage is critical for ensuring sustainable water management, reducing costs, and maintaining reliable water supply systems. An orifice on the wall of the pipe can cause several physical changes in the ground, and the pipe such as the pressure applied on the latter, the evolution of the soil moisture. Each of these phenomena can be used as a key symptom for monitoring the water distribution network. Therefore, the network operators began investigating ways to detect the presence of leaks and that as early as possible to intervene at the appropriate time [6] .

Leak detection in WDNs is the process of confirming the presence of leaks in a water distribution system. This can be done with specialized monitoring equipment such as modern or conventional leak detectors. These devices can detect a number of things such as changes in water pressure, vibration, sound input or flow, which can help indicate the presence of a leak. Network managers always look for harmless, fast and cheap leak detection systems. That is, seeking how to discover promptly a leak in a pipe is much accounted for when evaluating the water supply performance. Nowadays, leak detectors use different devices and techniques [7]. To detect and locate leaks with reasonable accuracy, various methods that rely on the analysis of measured signals are proposed in the open literature [8, 9].

The new work uses cepstrum techniques to analyze a range of pipe networks, both without and with leaks. This indicates that as the leak is applied, the analysed signal moves accordingly. The analysis reveals that the presence of a leak causes a noticeable shift in the signal. Additionally, it demonstrates that as the leak size increases, the amplitude of

the corresponding peak in the cepstrum also becomes more pronounced.

The objective of this research is to utilize signal processing through cepstrum analysis, exploring its application for accurate leak detection in WDNs.

II. Methodology

A. Experimental Setup for Pressure Signal Analysis

Leak detection in WDNs faces several key challenges. One of the main issues is the difficulty in identifying leaks early, particularly when they are small or hidden within underground pipes. Leaks are often subtle and can go undetected for long periods, causing significant water loss and infrastructure damage. In large and complex WDNs, traditional methods such as acoustic monitoring or manual inspections may be inefficient due to the extensive labor and time involved. Moreover, background noise and environmental factors can interfere with sensor readings, further complicating detection efforts. The limitations of conventional approaches highlight the need for more advanced techniques. As presented in Fig. 1.

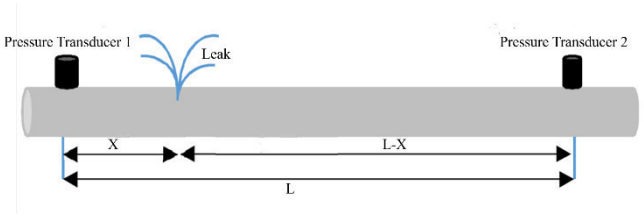


Fig. 1. The leak in a section of the pipe.

The proposed leak detection system uses pressure transducers to capture pressure signals. The setup consists of two transducers placed 14 meters from the pump, with one fixed 1.5 meters from the leak and the other movable along the pipe at 1.5-meter intervals, up to 78 meters. Leak detection is triggered by a push-button, and recordings are made for 20 seconds to capture the transient pressure drop. The pressure wave caused by the leak propagates in both directions and reaches the transducers at different times, allowing for leak detection based on the pressure signal changes.

B. Cepstrum Analysis Process

The cepstrum is a signal analysis technique used primarily in fields like audio processing, speech analysis, and vibration diagnostics. The term comes from reversing the first syllable of the word "spectrum." It provides a way to analyze the frequency content of a signal by transforming it into the "quefrency" domain, where periodicities in the spectrum become more apparent [10]. The principle of the procedure is shown in Fig 2.

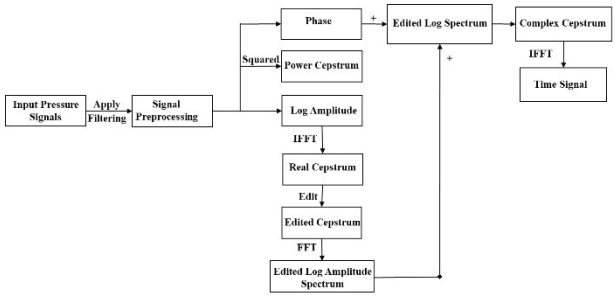


Fig. 2. Flowchart of Cepstrum Analysis for Leak Detection in Pipelines.

The diagram represents a comprehensive cepstrum analysis workflow for leak detection in WDNs using pressure signals. It processes the input signal through several stages, including filtering, phase extraction, cepstral analysis, and editing, to highlight periodic signals that correspond to leaks. The final output is a time-domain signal that helps detect and locate leaks in the network.

Fundamentally, the cepstrum is defined as the Fourier Transform of the logarithmic representation of a signal's Fourier Transform.

Mathematically: the cepstrum of signal=FT (Log (FT (the signal)))

Algorithmically: Signal → FFT → Log → IFFT → Cepstrum

where FFT indicates the Fast Fourier Transform

Cepstrum analysis can be categorized into two types: power cepstrum and complex cepstrum. The complex cepstrum is particularly notable for its ability to be reversed back to the original time-domain signal, making it a valuable tool for identifying localized singularities in a pressure time history. The complex cepstrum is defined as (1).

$$C(\tau) = F^{-1}(log \mid X(f) \mid) \quad (1)$$

C(τ) represents the cepstrum, where τ (quefrency) corresponds to time delays.

where X(f) is the complex spectrum of x(t). It can be represented in terms of the amplitude and phase at each frequency by (2).

$$X(f) = F\{x(t)\} \quad (2)$$

Take the logarithm of the magnitude of the frequency spectrum gives by (3)

$$log \mid X(f) \mid \quad (3)$$

The logarithmic operation enhances periodic components in the spectrum caused by reflections or leak-induced signals.

III. Result and Discussion

A. System

The system comprises pressure transducer signal readings from a laboratory-designed pipeline prototype intended to simulate leaks, thereby verifying the efficiency of the proposed method. Detection utilizing pressure transducers has emerged as a technique recently adopted for its superior performance relative to traditional methods. The document is divided into two parts:

Part 1: hydraulic circuit prototype

The hydraulic prototype circuit consists of a closed HDPE pipe with a diameter of Ø=40 mm, a length of 100 m, and a thickness of 2.4 mm. The latter is manufactured in 2022 by PLAST TUBE in SETIF as well as several elements such as the tank with 150 l, pump with (P=0.8Kw, flow rate = 100 l/min, and RPM = 2850), two very sensitive pressure transducers, model 811FMA with pressure range 0-100 PSIG, with accuracy of ±0.25% *FS* (Full Scale), and other accessories.

Part 2: Acquisition system

This part is an acquisition system comprising a conditioning card used to be able to exploit the 4-20 mA analog signals coming from the two transducers. Real-time control of the continuous system is achieved using a PC connected to a professional acquisition card, dSPACE DS1104. We take the sampling frequency equal to 1KHz. The signals from the pressure transmitters, which are in the range of 4-20mA, are relatively low and can be accurately modeled using the (4). The model of the transducer current function image of pressure is given by the following function taken from his datasheet:

$$I(mA) = 0{,}16 * P + 4mA \qquad (4)$$

The leak is simulated by a hole drilled in the pipeline connected to a solenoid valve (8 Watts, 0.7 Bar, 220V). The pressure image in the pipeline is measured by these pressure transmitters, which provide a current proportional to the pressure at a given point. This current is converted into a voltage by a conditioning card. The loop requires a 24 V power supply provided by a stabilized power supply. The two pressure signals are displayed on a PC using a professional dSPACE acquisition card. The displayed signals can be recorded to an Excel file with a CSV extension, and processed with programs developed in MATLAB. As outlined in Fig 3.

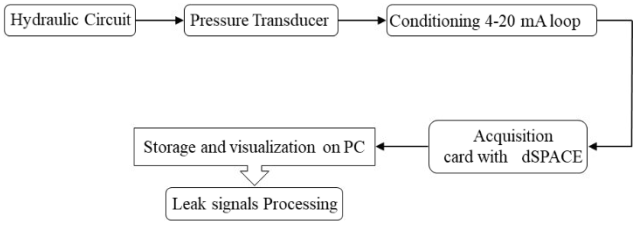


Fig. 3. Synoptic of the Acquisition System.

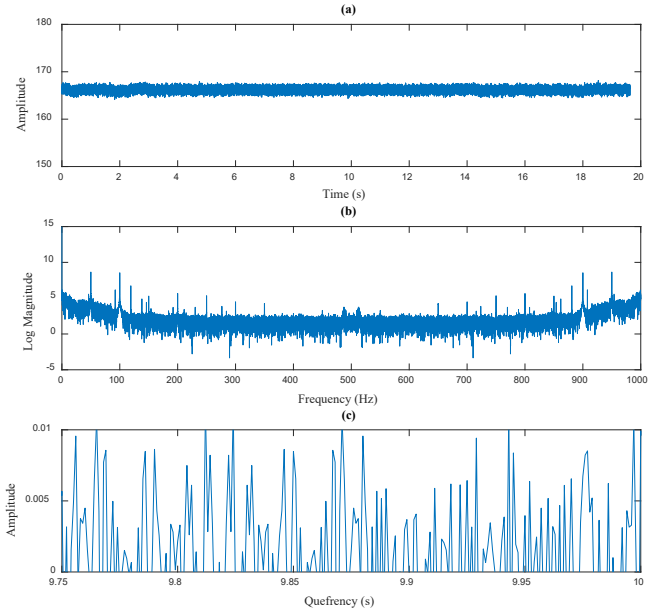


Fig. 4. Signals without leak. a) Original signal b) Log magnitude signal. c) Cepstrum signal.

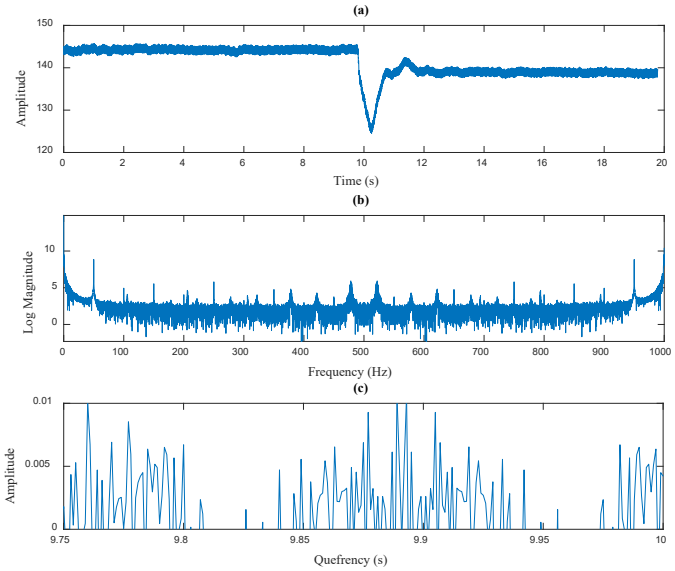


Fig. 5. With leak for distance 1.5 m. a) Original signal b) Log magnitude signal. c) Cepstrum signal.

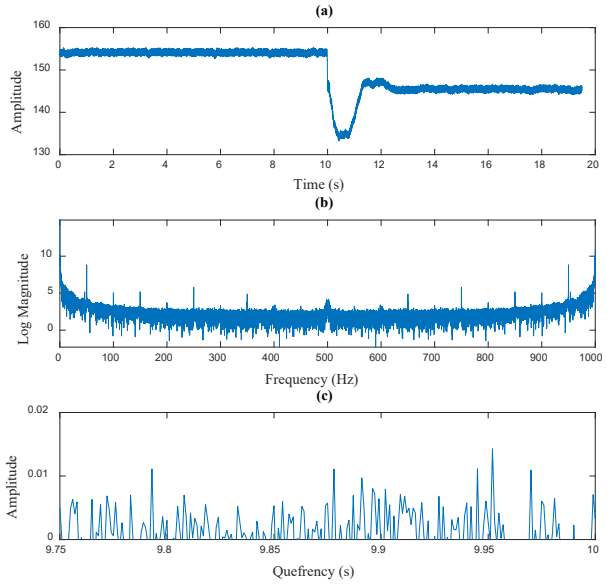


Fig. 6. Signal with leak for distance 76 m. a) Original signal b) Log magnitude signal. c) Cepstrum signal.

With the classical method such as FFT (Fast Fourier transform) we could not clearly distinguish the occurrence of any leak event during its creation Fig 7.

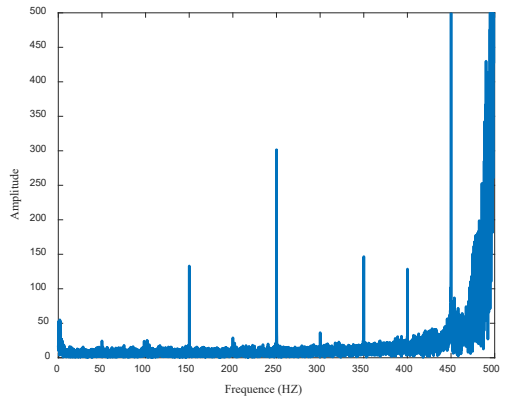


Fig. 7. FFT signal with leak for distance 76 m.

The application of the decimal logarithm on the amplitude also did not give information with which we can't judge the presence of the leak Fig 8 (log magnitude).

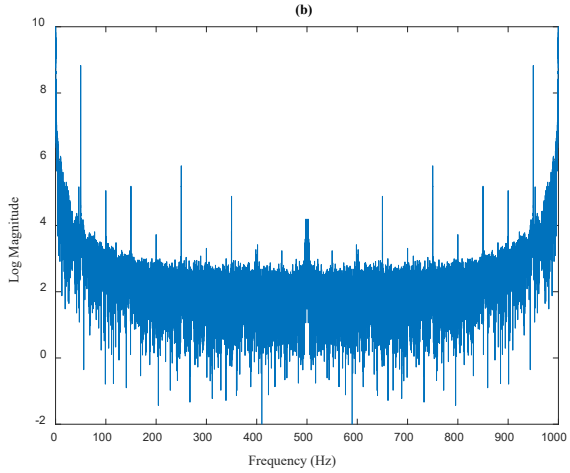


Fig. 8. Signal log magnitude with leak for distance 76 m.

On the other hand, with the cepster technique characterized by its expansion of the temporal axis, we note that in the vicinity of 9s a separation of a temporal space is created when the leak is stimulated Fig 5 and Fig 6 (c).

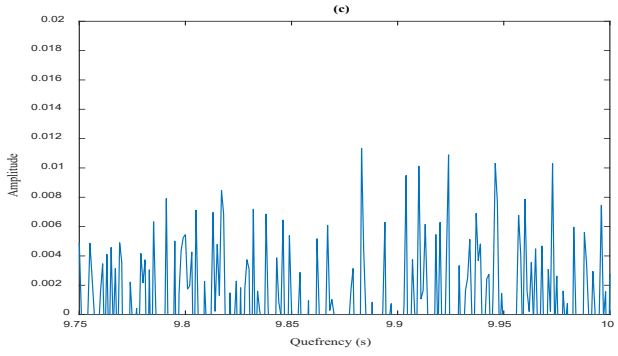


Fig. 9. The Cepstrum signal with leak for distance 1.5 m.

We observed this separation at different positions of the pressure transmitter. In this work, we only took two different positions of the transmitter at 1.5m and 76m. We also note that the modulus of the event is not necessarily high.

On the other hand, all along the axis we only have noises that are not useful.

IV. Conclusion

This study demonstrated the effectiveness of cepstrum analysis in detecting leaks within pipeline networks. By processing pressure signals and identifying distinctive features through the cepstrum domain, the method has proven to accurately locate leaks and assess their severity. This approach enhances the precision and reliability of leak detection compared to traditional techniques.

Acknowledgment

This study receives support from University research and training projects (PRFU) Project-N° 154/SDRF/2024 organized by the Algerian Ministry of Higher Education and Scientific and Research. The authors express gratitude to the general management of teachers and training and the management of doctoral training to support this work.

References

[1] Luis Romero-Ben, Débora Alves, Joaquim Blesa, Gabriela Cembrano, Vicenç Puig, and Eric Duviella, Leak detection and localization in water distribution networks: Review and perspective. Annual Reviews in Control, 2023. 55: p. 392-419.

[2] Adrià Soldevila, Giacomo Boracchi, Manuel Roveri, Sebastian Tornil-Sin, and Vicenç Puig, Leak detection and localization in water distribution networks by combining expert knowledge and data-driven models. Neural Computing and Applications, 2022. 34(6): p. 4759-4779.

[3] Ali Fares, Ia Tijani, Zhang Rui, and Tarek Zayed, Leak detection in real water distribution networks based on acoustic emission and machine learning. Environmental Technology, 2023. 44(25): p. 3850-3866.

[4] Rabab Al Hassani, Tarig Ali, Md Maruf Mortula, and Rahul Gawai, An Integrated Approach to Leak Detection in Water Distribution Networks (WDNs) Using GIS and Remote Sensing. Applied Sciences, 2023. 13(18): p. 10416.

[5] Huimin Yu, Sen Lin, Hua Zhou, Xiaodan Weng, Shipeng Chu, and Tingchao Yu, Leak detection in water distribution networks based on deep learning and kriging interpolation method. AQUA—Water Infrastructure, Ecosystems and Society, 2024. 73(8): p. 1741-1753.

[6] Miloud Bentoumi, Ahmed Bentoumi, and Haddi Bakhti, Welsh DSP estimate and EMD applied to leak detection in a water distribution pipeline. Instrumentation Mesure Métrologie, 2020. 19(1): p. 35-41.

[7] Miloud Bentoumi, Djamel Chikouche, Amar Mezache, and Haddi Bakhti, Wavelet DT method for water leak ‐ detection using a vibration sensor: an experimental analysis. IET Signal Processing, 2017. 11(4): p. 396-405.

[8] Alberto Martini, Marco Troncossi, and Alessandro Rivola, Automatic leak detection in buried plastic pipes of water supply networks by means of vibration measurements. Shock and Vibration, 2015. 2015(1): p. 165304.

[9] Aimé Lay-Ekuakille, Giuseppe Vendramin, and Amerigo Trotta, Robust spectral leak detection of complex pipelines using filter diagonalization method. IEEE Sensors Journal, 2009. 9(11): p. 1605-1614.

[10] M Taghvaei, Sbm Beck, and Wj Staszewski, Leak detection in pipeline networks using low ‐ profile piezoceramic transducers. Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures, 2007. 14(8): p. 1063-1082.

Performances Evaluation of Electrical Appliances Identification System using COOLL database in brand-independent mode

Fateh Ghazali
LMSE laboratory, University Mohamed El Bachir El Ibrahimi of Bordj Bou Arreridj, El-Anasser, Bordj Bou Arréridj, 34030, Algeria
laboauto@hotmail.fr

Abdenour Hacine-Gharbi
LMSE laboratory, University Mohamed El Bachir El Ibrahimi of Bordj Bou Arreridj, El-Anasser, Bordj Bou Arréridj, 34030, Algeria
gharbi07@yahoo.fr

Hamza Roubhi
ETA laboratory, University Mohamed El Bachir El Ibrahimi of Bordj Bou Arreridj, El-Anasser, Bordj Bou Arréridj, 34030, Algeria
hamza.roubhi@univ-bba.dz

Abstract*— In this work, we evaluate the Electrical Identification (IAE) system, based on the KNN classifier with the voting rule method and feature extraction approach based on Discrete Wavelet Energy (DWE). We also incorporate two other descriptors inspired from DWE: Log Wavelet Energy (LWE) and Wavelet Cepstral Coefficients (WCC). The evaluation is conducted using the COOLL database distributed in two modes: appliance brand-dependent mode and brand-independent mode. In the brand-dependent mode, for each appliance signal class, each brand will have examples in testing and training data. Conversely, in the brand-independent mode, the brand in testing data does not have any example in training data. This paper aims principally to evaluate the influence of the appliance brand on EAI system performance. To do this, we propose a comparative study between the EAI system performance in brand-dependent mode and those of brand- independent mode with different descriptors. In addition, in order to improve probably the system performance in appliance brand-independent mode, we have tested other classifiers such as Gaussian Mixture Model (GMM), Linear Discriminant Analysis (LDA) and Artificial Neural Network (ANN).

The results show that our EAI system performs well in brand-dependent mode and poorly in brand-independent mode, indicating that the appliance brand significantly influences the system performance. This highlights a robustness issue regarding appliance brands in the EAI system.

***KEYWORDS—Non-Intrusive Load Monitoring (NILM), Electrical Appliances Identification, Discrete Wavelets Analysis, Wavelet Cepstral Coefficient (WCC), k*-nearest neighbors (KNN), voting rules method, Appliance brand-independent mode.**

I. INTRODUCTION

The electrical appliances identification (EAI) problem has been extensively discussed by several research aimed at developing novel frameworks using machine learning methods in the field of Non-Intrusive Load Monitoring (NILM) also called energy disaggregation. The goal of NILM is to decompose aggregated energy consumption into its individual components non-intrusively, using a single energy meter.

Previous studies [1], [2], [3], [4], [5], [6], [7] proposed EAI systems using The Plug Load Appliance Identification Dataset (PLAID) [8] to evaluate their performances. In [1], the authors, propose an EAI system based on STFS feature extraction method combined with the HMM models evaluated on PLAID data set. This last work was followed by the work of [2], in which, the authors proposed an EAI system using wavelet cepstral coefficients as features. In [3] , the authors propose an EAI system based on KNN classifier combined with voting rule method using statistical harmonics as features. This work flowed by [5], [6] and [7], all these works used a EAI system based on KNN classifier combined with voting rule method and the use of PLAID dataset for the objective to improve the classification ratio CR. The obtained results give a maximal CR of 98.52 % with [6].

In this work, we propose to use the COOLL database [9] to evaluate the performance of these prior studies. In addition, we propose also to evaluate these systems in appliance brand-independent mode. The Controlled On/Off Loads Library (COOLL) is a Public Dataset of high-sampled electrical current and voltage measurements representing individual appliances consumption for Appliance Identification. The appliances are mainly controllable appliances. 42 appliances brands of 12 types were measured at a 100 kHz sampling frequency.

The EAI systems have employed this latest COOLL dataset, which is divided into two parts: one for training and the other for testing. These parts can be organized in either an appliance brand-dependent or appliance brand-independent mode. In the appliance brand-dependent mode, the testing part contains appliance classes and brands that are also represented in the training part. Conversely, in the appliance brand-independent mode, the testing part includes brands that are not present in the training part. Our second contribution is therefore to evaluate the identification system based on the KNN classifier and the use of the voting rule under the COOLL database organized in appliance brand-independent mode.

II. RELATEDWORKS

In [3], the authors proposed a novel flowchart of an EAI system based on KNN classifier combined with a voting rule strategy using statistical harmonic features derived from harmonic analysis. Each signal must be converted into a sequence of statistical harmonic features based on a fixed-duration time segment. The obtained results indicate that using the mean and standard deviation with 500 statistical features (250 for each) achieves a CR classification rate of 92.63%. Implementing the voting rule strategy further enhances this result to 94.97%.A flowchart of the system proposed in [3] is given in Figure 1.

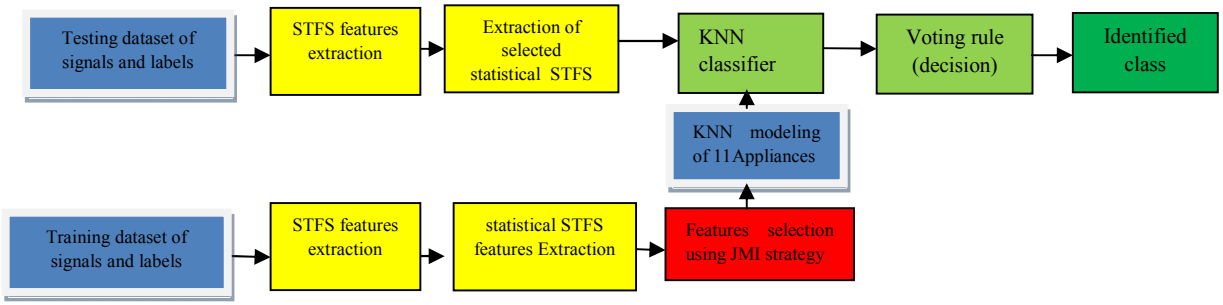


Figure 1 : IAE system proposed in [3]

This work is followed by the work of [4], when the authors propose to reduce the dimensionality by using a compact features representation based on the computing of energy at each decomposition level of discrete wavelet analysis (DWE: Discrete Wavelet Energy). The statistical features are computed over time using the coefficients of the DWE descriptor or coefficients of other descriptors such LWE(Log Wavelet Energy) and WCC(Wavelet Cepstral Coefficients) extracted from DWE coefficients. Figure 2 presents a flowchart of the features extraction method based on Wavelet Cepstral Coefficients for EAI proposed in [4].

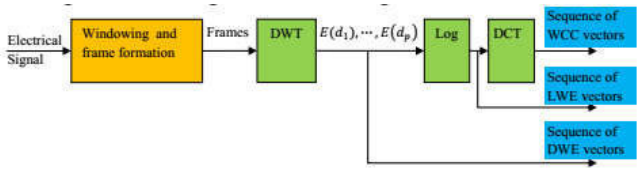


Figure 2: Process of DWE / LWE / WCC features extraction with hamming windowing [4]

III. THE EAI IN APPLIANCE BRAND-INDEPENDENT MODE

It should be noted that the operation of such EAI system requires a database. The latter is divided into two subsets, one for learning and the other for testing, so that each appliance brand will have examples in both bases (appliance brand-dependent mode). In this work, we are interested in the appliance brand- independent mode, ie each appliance brand will have examples in a single database (either for testing or training).

III-1 PRESENTATION OF COOLL DATABASE

This section provides a brief overview of the Controlled On/Off Loads Library dataset (COOLL) , which consists of high-frequency electrical current and voltage measurements for individual appliances. Collected at the PRISME laboratory at the University of Orléans, France, the dataset includes 42 controllable appliances across 12 types, sampled at 100 kHz. The COOLL dataset offers twenty turn-on transient signals for each appliance, each recorded at a distinct turn-on moment with a controlled delay relative to the mains voltage zero-crossing. This allows for tracking the variation in the turn-on waveform as the appliance connects to the power grid. Figure 3 summarizes the appliances found in the dataset with the different appliance types and the number for each type.

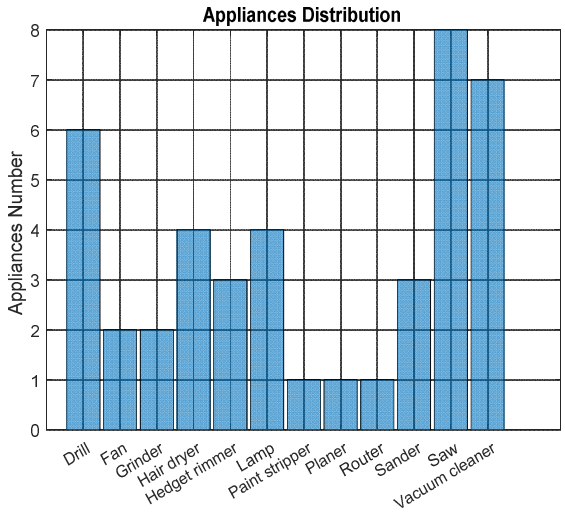


Figure 3: Distribution of appliances types and his number in COOLL dataset.

The COOLL database includes turn-on current and voltage measurements for 42 appliances, with 840 samples each at a frequency of 100 kHz. For each appliance, 20 controlled measurements are taken, corresponding to action delays from 0 to 19 ms in 1 ms increments. This range covers the entire 20 ms duration of the 50 Hz mains voltage cycle, indicating the delay of the turn-on action relative to the start of a specific mains voltage cycle. The Table 1 presents the COOLL dataset summary.

Table 1: COOLL dataset summary

NBR	Appliance type	Number of appliances	Numberof current signals (20 per appliance)
1	Drill	6	120
2	Fan	2	40
3	Grinder	2	40
4	Hair dryer	4	80
5	Hedge trimmer	3	60
6	Lamp	4	80
7	Paint stripper	1	20
8	Planner	1	20
9	Router	1	20
10	Sander	3	60
11	Saw	8	160
12	Vaccum cleaner	7	140
Overall		42	840

III-2 COOLL DATASET SUBDIVISION IN APPLIANCE BRAND INDEPENDENT MODE

The Table 2 present the subdivision of COOLL dataset for testing and training in appliance brand independent mode

Table2: COOLL database distribution in appliance brand independent mode (50% for tasting and 50% for training)

N°	Appliance type	Number of appliances		Training subset		Testing subset	
		Nbr	ins	Nbr	ins	Nbr	ins
1	Drill	6	120	3	60	3	60
2	Fan	2	40	1	20	1	20
3	Grinder	2	40	1	20	1	20
4	Hair dryer	4	80	2	40	2	40
5	Hedge trimmer	3	60	1	20	2	40
6	Lamp	4	80	2	40	2	40
7	Paint stripper	1	20	1	20	0	00
8	Planner	1	20	1	20	0	00
9	Router	1	20	1	20	0	00
10	Sander	3	60	1	20	2	40
11	Saw	8	160	4	80	4	80
12	Vaccum cleaner	7	140	3	60	4	80
Overall		**42**	**840**	**24**	**420**	**24**	**420**

IV- FEATURES EXTRACTION STEP

The extraction step converts each current signal into a sequence of feature vectors using wavelet analysis, which involves several sub-steps: pre-processing, windowing, and computing of energy at each decomposition level of discrete wavelet analysis (DWE: Discrete Wavelet Energy) or coefficients of other descriptors such as LWE(Log Wavelet Energy) and WCC(Wavelet Cepstral Coefficients) extracted from DWE coefficients.

IV-1 PRE-PROCESSING

The signal pre-processing step ensures the signal is suitable for classification. We extracted for each appliance the operating state intervals by identifying the start and end points from the COOLL database text files. Figure 3 shows the curve of scenario C1_1 which represents the current signal of the Drill1_1 appliance before and after the pre-processing step.

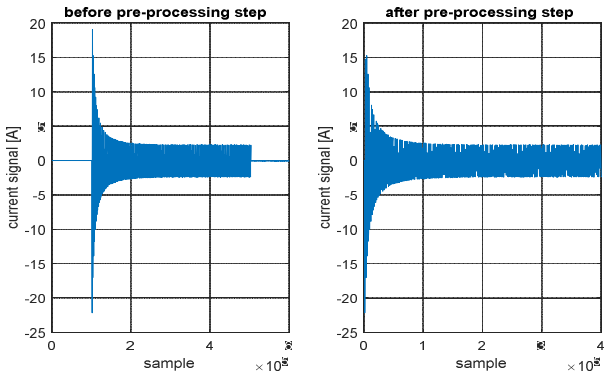


Figure 3 : scenarioC1_1 of the Drill_1 appliance before and after pre-processing step.

IV-2 WINDOWING

This step consists of dividing the signal into a sequence of 20 ms (1/50 Hz) analysis windows with a 50% overlap, and converting each analysis window into a vector of DWE coefficients based on the computing of energy at each decomposition level of discrete wavelet analysis. Thus, this step converts each signal into a sequence of DWE feature vectors or two others features vectors derived from the DWE such as LWE and WCC (Figure 2).

IV- 3 DISCRETE WAVELET ANALYSIS

The discrete wavelet transform is applied to each analysis window, calculating the energy at each level of dyadic decomposition to obtain DWE, LWE, and WCC coefficients.

The DWE coefficients defined as:

$$\begin{cases} E(d_j) = \frac{1}{E_T}\sum_{n=0}^{n=N_j-1}|d_j[n]|^2 \\ E(a_p) = \frac{1}{E_T}\sum_{n=0}^{n=N_p-1}|a_p[n]|^2 \end{cases} \quad \text{for} \quad j \ldots . p \quad (1)$$

N_j is the number of wavelet coefficients that are computed at each scale $jN_j = N/2^j$.

The LWE coefficients defined as:

$$\begin{cases} LE(d_j) = \log\left(\frac{1}{E_T}\sum_{n=0}^{n=N_j-1}|d_j[n]|^2\right) \\ LE(a_p) = \log\left(\frac{1}{E_T}\sum_{n=0}^{n=N_p-1}|a_p[n]|^2\right) \end{cases} \quad \text{for} \quad j = 1 \ldots . p \quad (2)$$

The WCC coefficients, defined as:

$$\begin{cases} WCC(d_j) = RDCT\left[\log\left(\frac{1}{E_T}\sum_{n=0}^{n=N_j-1}|d_j[n]|^2\right)\right] \\ WCC(a_p) = RDCT\left[\log\left(\frac{1}{E_T}\sum_{n=0}^{n=N_p-1}|a_p[n]|^2\right)\right] \end{cases} \text{for} \quad j = 1..p \ (3)$$

V. EXPERIMENTAL RESULTS AND DISCUSSION

In order to evaluate the EAI system based on KNN classifier combined with voting rule method, we tested it using the WCC descriptors extracted from the current signals of the COOLL database in appliance brand-dependent mode and also in appliance brand-independent mode. For this study, we present in this section the following experiments:

- To test the robustness of the proposed EAI systems in the previews works in terms of appliance brand, a performance comparative study between different descriptors such as DWE, LWE and WCC in appliance brand-dependent mode and appliance brand-independent mode, is carried out using the best configurations obtained in the previous works(the use of DB5 at level 5 with 4 cycles (one cycle =20ms) of window duration).

- To improve probably the accuracy CR of the EAI system spicily in appliance brand-independent mode, we are seeking the optimal configuration of the KNN classifier (K value and distance).
- Test another classifier such as GMM, ANN and LDA with COOLL database distribution in appliance brand-independent mode .

V-1 COMPARATIVE STUDY BETWEEN BRAND-DEPENDENT AND BRAND-INDEPENDENT MODE

This experiment evaluates the performance the EAI system based on the KNN classifier combined with a voting rule strategy, using the COOLL database and also tests the robustness of this EAI system in appliance brand-dependent mode and appliance brand-independent mode. The evaluation is carried out in terms of the CR classification rate. Taking the best configurations obtained in the previews work [6], which are the window durations of 4 cycle length using DB5 mother wavelet with a decomposition level equal to 5, and with KNN classifier's default configuration (k=1 , distance Euclidean). Table 3 presents the CR's obtained with different descriptors in different modes.

Table 3: CR's of appliance brand-dependent mode and appliance brand-independent mode

Descriptors	DWE	LWE	WCC
NBF	6	6	6
CR in Appliance brand-dependent mode	100	100	100
CR in Appliance brand-independent mode	35.95	35.95	35.95

From the results obtained in the Table 3, we can observe the following points:

- The EAI system has very good results in appliance brand-dependent mode achieving100% of classification ratio CR for all descriptors
- The classification rate (CR) in appliance brand-independent mode is relatively low for all descriptors type. These results show that the electrical appliance brand affects the performance of our EAI system.

To enhance identification results in appliance brand-independent mode, we aim to find the right alternative configurations of our system. The next experiment will focus on optimizing the KNN classifier by determining the ideal number of nearest neighbor vectors and the optimal distance metric.

V-2 OPTIMAL CONFIGURATION OF THE KNN CLASSIFIER

This experiment aims to seek the optimal configuration for the KNN classifier by determining the best number of nearest neighbors, K, and the most effective distance metric in terms of the classification rate (CR). Table 4 displays the CR for K values ranging from 1 to 200 across various distance metrics (Euclidean, Cosine, Correlation, Cityblock) using the WCC descriptor.

Table 4 : KNN Classifier Configuration

k	Distances			
	Euclidean	**Cosine**	**Correlation**	**Cityblock**
1	35.95	43.09	44.04	35.71
2	35.95	43.09	44.04	35.71
3	39.28	43.33	45.00	35.00
4	34.52	43.33	46.19	35.23
5	35.71	44.52	46.66	35.00
6	35.00	44.28	46.19	35.00
7	35.95	45.95	47.85	35.23
8	35.95	45.95	47.61	35.00
9	36.42	46.42	47.85	35.23
10	36.19	46.90	47.61	35.47
20	38.33	47.38	47.85	36.19
30	38.80	47.61	48.09	38.09
40	39.28	47.85	48.80	38.80
50	39.52	47.85	49.28	39.28
60	39.52	48.33	49.52	39.28
70	39.52	48.57	50.95	39.04
80	39.52	50.23	51.42	39.04
90	39.52	50.71	52.38	39.04
100	39.52	51.42	52.85	39.04
200	39.52	52.85	53.33	39.04

Table 4 shows that the maximum classification rate of 53.33% is achieved by using 'correlation' distances with k set to 200.

V-3 TEST OF THE SYSTEM WITH OTHER CLASSIFIERS

In order to improve the classification results in appliance brand-independent mode, we have tested our EAI system with other classifiers such as GMM, LDA and ANN. Table 5 presents the CR's obtained with various classifiers and descriptors in two modes with optimal configurations.

- The optimal configuration of the GMM classifier is to take a Gaussian Number Ng equal to 4.

- The optimal configuration of the LDA classifier is to take Discriminant Function Type DFT='mahalanobis"
- The optimal configuration of the ANN classifier is to take hidden Layer Size HLS=100.

Table 5 : *CR's for the different classifiers and different mode*

	Appliance brand-dependent mode			Appliance brand-independent mode		
	GMM	LDA	ANN	GMM	LDA	ANN
DWE	97.38	57.17	47.61	53.09	47.38	44.04
LWE	100	93.80	99.76	57.85	50.95	46.19
WCC	97.85	93.80	100	53.09	47.38	50.00

Table 4 shows that the EAI system performs well in appliance brand-dependent mode, but results in the brand-independent mode are low, with only a slight increase in the CR to 57.85% when using the GMM classifier with the LWE descriptor.

VI- CONCLUSION

This work evaluates the performance of the EAI system using a KNN classifier combined with a voting rule method, incorporating a compact feature representation through Discrete Wavelet Energy (DWE) as a discriminative descriptor. The system's performance is assessed using two modes of the COOLL database distribution, focusing on appliance brands to test the robustness of the EAI system. Firstly, a comparative study is conducted on different descriptors (DWE, LWE, and WCC) in two previous modes. Secondly, we evaluate other classifiers, including GMM, LDA, and ANN.

The obtained results indicate that our EAI system performs well in brand-dependent mode and poorly in brand-independent mode. This explains why the electrical appliance brand has a strong imprint in the current signal. This suggests an issue with the robustness of the EAI system regarding appliance brands.

REFERENCES

[1] Nait-Meziane M, Hacine-Gharbi A, Ravier P, Lamarque G, and Le Bunetel JC et al, "HMM-based transient and steady-state current signals modeling for electrical appliances identifcation.," , Rome, Italy, 2016.

[2] A Hacine-Gharbi and P Ravier, "Wavelet cepstral coefficients for," in *7th International Conference on Pattern Recognition Applications and Methods*, Funchal, Portugal, 2018.

[3] F. Ghazali, A. Hacine-Gharbi, P. Ravier, and T Mohamadi, "Extraction and selection of statistical harmonics features for electrical appliances identification using kNN classifier, combined with voting rules method ," *Turkish Journal of Electrical Engineering & ComputerSciences* , vol. Vol.27, , no. No.4, pp. pp.2980-2997, 2019.

[4] Fateh Ghazali, Hacine-Gharbi Abdenour, and Ravier Philippe, "Statistical features extraction based on the discrete wavelet transform for electrical appliances identification," in *the international conference of intelligent systems and pattern recognition.*, 16-18 October 2020, Hammamet , Tunisia, 2020.

[5] A. Hacine-Gharbi, P. Ravier, and M. and Nait-Meziane, "Relevant harmonics selection based on mutual information for electrical appliances identification," vol. Vol. 62, no. pp.102–115, DOI: 10.1504/IJCAT.2020., 2020.

[6] F. Ghazali, A Hacine-Gharbi, and P. Ravier, "Selection of statistical wavelet features using a wrapper approach for electrical appliances identification based on a KNN classifier combined with voting rules method," *Int. J.Computational Systems Engineering*, vol. Vol. 6, no. No. 5, pp. pp.220–230., 2021.

[7] Fateh Ghazali, Abdenour Hacine-Gharbi, Khaled Rouabah, and Philippe Ravier, "Performance Evaluation of the Electrical Appliances Identification System Using the PLAID Database in Independent Mode of House," in *13th International Conference on Pattern Recognition Applications and Methods*, Rome-Italy, 2024, pp. pp.880-885.

[8] J Gao, S Giri, EC Kara, and M Berges, "Plaid: A public dataset of high resolution electrical appliance measurements for load identification research: demo abstract.," in *In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy Efcient Buildings*, New York, NY, USA, 2014, pp. pp. 198–199.

[9] Thomas Picon et al., "COOLL: Controlled On/Off Loads Library, a Public Dataset of High-Sampled Electrical Signals for Appliance Identification," *arXiv:1611.05803v1 [cs.OH]* , Nov 2016.

ECG signal Denoising Techniques based on Wavelet Transforms : Assessment and Comparison

Saida BOUDINA
Department Engineering of Electrical Systems
Faculty of Technology, LIST Laboratory
35000 Boumerdes, Algeria
sa.boudina@univ-boumerdes.dz

Samia BELKACEM
Department Engineering of Electrical Systems
Faculty of Technology
35000 Boumerdes, Algeria
s.belkacem@univ-boumerdes.dz

Noureddine MESSAOUDI
Department Engineering of Electrical Systems
Faculty of Technology, LIST Laboratory
35000 Boumerdes, Algeria
n.messaoudi@univ-boumerdes.dz

Abstract - Electrocardiographic (ECG) signal quality is crucial for accurate diagnosis and monitoring of cardiovascular disease. However, various types of noise often corrupt ECG recordings, which can hamper their interpretation. So denoising ECG signals is an important task for diagnosing and monitoring heart disease. Wavelet transforms are effective tools for denoising ECG signals. In this study, we compared a few techniques for denoising ECG signals using wavelet transforms on real ECG signals contaminated with white Gaussian noise to determine their effectiveness and applicability in the clinical setting. ECG signals from the standard MIT-BIH database are adopted to verify the proposed techniques using MATLAB R2020a software. We used three performance evaluation criteria to measure the quality of denoised signals, namely mean square error (MSE), signal-to-noise ratio (SNR), and percentage root mean square difference (PRD). Experimental results indicate that some proposed strategies can reduce noise from contaminated ECG signals more efficiently than others.

Keywords: ECG, denoising, WT, SWT, SNR, MSE, PRD.

I. INTRODUCTION

To assess cardiac health, we record the electrical activity of the heart as an ECG signal, and this can be influenced by several elements resulting from the patient's motions, and breathing, muscular electrical activity, incorrect positioning of the heart and electrodes, or interference with the electrical network. The presence of these various external disturbances can greatly affect valid features that cannot be removed using simple filtering methods, and degrade the quality of the recorded ECG signal. Therefore, it is necessary to pre-process and denoise the ECG signals before diagnosis and analysis. The following are different types of noises: **Baseline Wandering**: It is an artifact in the low-frequency ECG, ranging from 0.15 to 0.3 Hz, typically generated by respiration, subject movement, electrically charged electrodes, or patient motions [1]. **Powerline Interference** It is a high-frequency noise that often ranges from 50 to 60 Hz, typically originating from power supply lines [1][2]. **Electromyographic Noise** EMG noise often arises from the contraction and relaxation of muscles distinct from cardiac muscles. Generally, occurs throughout the frequency spectrum of 5 to 500 Hz [1]. **White Gaussian Noise (WGN):** White Gaussian noise is utilized as artificial interference noise [3][4]. It is identified by a wide frequency range, as it merges all the aforementioned noises [3].

In this article, we present a comparative study of five techniques for denoising ECG signals using wavelet transforms, namely hard thresholding, soft thresholding, visual thresholding (Visu shrink), stationary wavelet transform (SWT) and adaptive SWT based on median absolute deviation (MAD). We evaluate the performance of these techniques on real ECG signals from MIT-BIH databases, using different types of wavelet bases (Db4, Db8, Db12 and bior3.1). Results are evaluated according to three common criteria: Mean Square Error (MSE), Signal-to-Noise-Ratio (SNR), Percentage Root Mean square Difference (PRD).

The rest of the paper is structured as follows: In section 2, we provide some works related to our topic. Section 3 offers a comprehensive explanation of the tools and methodology employed. Section 4 establishes the results derived from the study, followed by a discussion of the findings. Section 5 finishes the work by summarizing the principal insights and proposing directions for future research.

II. RELATED WORK

A great deal of work has been devoted to denoising ECG signals using wavelet transforms. Most wavelet denoising algorithms for ECG signals are based on the universal theory of Donoho [4][5].The followings are a few research projects related to our topic:

Madan and al. [1] proposed an ECG signal denoising method called STWaTV (Stationary Wavelet Total Variation). They introduced total variation (TV) regularization in the stationary wavelet domain (SWT), using a bivariate shrinkage rule for thresholding. In their research work [2], Kumar and al. proposed a comparison of different feature extraction and denoising techniques using the stationary wavelet transform, as well as various denoising techniques including low-pass filtering, high-pass filtering, empirical mode decomposition, Fourier decomposition method, wavelet transform. Gupta and al.[6] used a fractional wavelet transform (FrWT) method as a pre-processing tool.ECG signal denoising method based on wavelet energy and a sub-band smoothing filter have proposed in [7], the proposed method selects only those coefficients to be denuded based on their energy. A sub-band smoothing filter is applied to improve the quality advantage of the noise signal, particularly in the areas between QRS complexes. Al-Safi and al. [8] offers a variety of algorithms (LMS, NLMS and Leaky LMS) for the denoising real-time ECG signals using adaptive filters. The proposed method uses a delayed

version of the ECG signal itself as a reference. An adaptive predictive filter is then applied. Thamarai and al. [9] proposed a multi-scale wavelet-based noise suppression algorithm that preserves both the signal's temporal and frequency resolution. This algorithm is applied to noisy EEG, ECG and PPG signals. Gualsaquí and al. [10] presented an experimental study of ECG signal denoising using the discrete wavelet transform (DWT) and different thresholds (values and functions). Ziran and al. [11] have proposed a method for automatic analysis and detection of eigenvalues by electrocardiogram (ECG), using ECG eigenvalues to invert the myocardial action potential for automatic detection and diagnosis of heart disease. Mandala and al. [12] conducted a thorough simulation and analysis to evaluate the efficacy of various wavelet-based denoising techniques. They did so by incorporating adaptive white Gaussian noise (AWGN) into the ECG signal before the denoising process. They investigated three kinds of denoising techniques: soft thresholding, hard thresholding, and Visu Shrink (adaptive).

III. Materials and Methods

This section outlines the materials, tools, and experimental procedures used to conduct the study.

A. Evaluation criteria

The performance evaluated using MSE, SNR and PRD**.**

Mean Square Error (MSE); which measures the difference between the distorted signal and the original signal. [12] It is defined as follows (1):

$$\mathbf{MSE}=\frac{1}{N}\sum_{i=1}^{N}(x_i-\hat{x}_i)^2 \quad (1)$$

Where N is the number of signal points, x_i is the original signal, and $\hat{x}_i$ the denoised signal.

Signal-to-noise ratio (SNR), which measures the ratio between the power of the original signal and the power of the noise. It is defined as follows (2)[2]:

$$\mathbf{SNR}=10\log_{10}\frac{\sum_{i=1}^{N}x_i^2}{\sum_{i=1}^{N}(x_i-\hat{x}_i)^2} \quad (2)$$

Where N is the number of signal points, $x_i$is the original signal point, and $\hat{x}_i$ the denoised signal point.

Percentage-root-mean-square difference (PRD), is calculated to check the distortion in the denoised signal as compared to the original signal. It is defined as follows (3) [1]:

$$\mathbf{PRD}=100\times\frac{\sum_{i=1}^{N}(x_i-\hat{x}_i)^2}{\sum_{i=1}^{N}x_i^2} \quad (3)$$

B. Methods

In this section, we present the five techniques for denoising ECG signals using Wavelet Transforms that we studied,

1) Discrete Wavelet Transformations (DWT)

DWT works with discrete signals defined as:

$$DWT(s,\tau)=2^{\frac{-s}{2}}\sum x[n]\,\Psi^*(2^{-s}n-\tau) \quad (4)$$

Where, x[n] is the discrete time signal, Ψ* is the complex conjugate of the analyzing wavelet function (Ψ[n]), *s* and *τ* are the dilation and location parameters respectively [10].

The inverse discrete wavelet transform (*IDWT*) is given by:

$$IDWT(s,\tau)=A\sum\sum DWT(s,\tau)\,\Psi^*(2^{-s}n-\tau) \quad (5)$$

Where *A* is a constant that do not depend of **x** [n], and *IDWT* is the reconstructed signal [10].

2) Wavelet Thresholding

The well-known thresholds in the field are listed below:

a) Hard thresholding

Hard thresholding is a denoising technique in which wavelet coefficients below a certain threshold are cancelled, while coefficients above the threshold are retained. [12] Hard thresholding can be formulated as follows:

$$y_i=\begin{cases}x_i & if\ |x_i|\ \geq\lambda\\ 0 & \text{otherwise}\end{cases} \quad (6)$$

Where x_i is the wavelet coefficient to be processed, y_i is the denoised wavelet coefficients, and λ is the threshold.

b) Soft thresholding

Soft thresholding is a denoising technique in which wavelet coefficients below a certain threshold are reduced, while coefficients above the threshold are retained. Soft thresholding can be formulated as follows [12]:

$$y_i=\begin{cases}x_i-\lambda & if\ x_i>\lambda\\ 0 & if\ |x_i|\ \leq\lambda\\ x_i+\lambda & if\ x_i<-\lambda\end{cases} \quad (7)$$

Where x_i is the wavelet coefficient to be processed, y_i is the denoised wavelet coefficient, and λ is the threshold.

c) Visu shrink

Visual thresholding (Visu shrink) is a denoising technique that adaptively selects the threshold,[12] depending on the noise level and signal size. The visual threshold is defined as follows:

$$\lambda=\sigma\sqrt{2logN} \quad (8)$$

Where σ is the standard deviation of the noise, and N is the number of signal points.

d) Stationary wavelet transform (SWT)

Stationary Wavelet Transform (SWT) is a denoising technique that uses a modified version of DWT, which preserves the signal length at each decomposition level. SWT can be formulated as follows:

$$SWT(x)=\sum_{j=0}^{j-1}\sum_{k=0}^{N-1}c_{j,k}\phi_{j,k}(t)+\sum_{j=0}^{j-1}\sum_{k=0}^{N-1}d_{j,k}\psi_{j,k}(t) \quad (9)$$

Where$x$ is the signal to be processed, j is the number of decomposition levels, **N** is the number of signal points, $c_{j,k}$ and $d_{j,k}$ are the approximate and retail wavelet coefficients, respectively, $\phi_{j,k}(t)$ and $\psi_{j,k}(t)$ are the scaling and wavelet functions, respectively, defined as follows:

$$\phi_{j,k}(t)=2^{-\frac{j}{2}}\phi(2^{-j}t-k) \quad (10)$$

$$\psi_{j,k}(t)=2^{-\frac{j}{2}}\psi(2^{-j}t-k) \quad (11)$$

e) Adaptive SWT

Adaptive SWT is a denoising technique that uses an enhanced version of SWT, which adapts the threshold to each decomposition level and wavelet coefficient. MAD is defined as follows:

$$MAD=\mathbf{median}(|\,x_i-\mathbf{median}(x_i)\,|) \quad (12)$$

Where x_i is the wavelet coefficient to be processed. The adaptive threshold is defined as follows:

$$\lambda_j=k\,\sigma_j \quad (13)$$

Where j is the decomposition level, **k** is a proportionality factor, and σ_j is the noise standard deviation estimated from the MAD, according to the following relationship [13]

$$\sigma_j = \frac{MAD}{0.6745} \quad (14)$$

C. Wavelet function selection

In order to increase the efficiency of the proposed method, the selection of an appropriate wavelet function is carried out. In particular, we applied four wavelet functions "Db4", "Db8", "Db12", "bior1.3" to select the best one based on SNR, MSE and PRD values.

D. Bloc diagram

The algorithm employed in ECG signal denoising comprises five fundamental phases, as depicted in Fig.1.

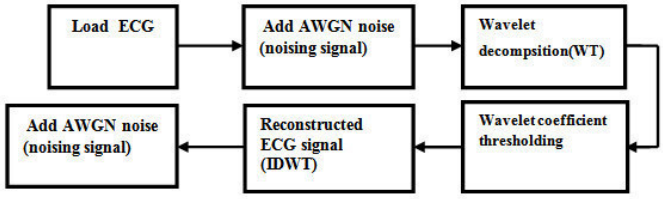


FIG. 1. THE ALGORITHM STEPS USED IN ECG SIGNAL DENOISING.

ECG Signal: The original ECG signal is used as input.
Add AWGN Noise: At this stage, artificial noise (AWGN) is added to the ECG signal to simulate disturbances that may occur in real-life conditions.
Wavelet decomposition: This step involves decomposing the noisy signal into a series of wavelet coefficients using the wavelet transform.
Wavelet coefficient thresholding: Here, the coefficients obtained during wavelet decomposition are processed by a thresholding process. The choice of threshold is an important step in the wavelet thresholding algorithm.
Reconstruct the signal: Once thresholding is complete, the signal is reconstructed using the inverse wavelet transform. This results in a clean signal, where noise has been reduced, while preserving important characteristics of the original signal.

IV. RESULT AND DISCUTIONS

The performance of the proposed techniques was evaluated on the MIT-BIH (Massachusetts Institute of Technology - Beth Israel Hospital) database. We used signals 101, 103, 115, 122 and 123 for our study.

The simulation parameters of our experiment are: clean and noisy signals, wavelet bases used for denoising, denoising techniques, decomposition level and thresholding value.

A representation of a pure signal (noise-free) and noisy signals is given in Fig. 2.

Wavelet bases: We use four types of wavelet bases that are most commonly used in wavelet signal processing, namely: Db 4, Db8, Db12 and bior3.1.

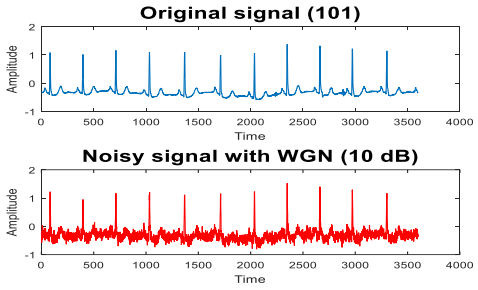


FIG. 2. (A) ORIGINAL ECG SIGNAL, (B) NOISED ECG SIGNAL.

Denoising techniques: The denoising techniques used are: hard thresholding, soft thresholding, Visu shrink, SWT and adaptive SWT.
Decomposition level: In our experiment, we use level four "4" decomposition for each of the signals to perform the denoising technique.
Threshold value: The optimal threshold value for thresholding methods is "0,1".

A. Results

The following results are given to demonstrate our findings. FIG.3 demonstrates the results when Hard Thresholding, soft Thresholding, Visu shrink thresholding, SWT, and SWT adaptive are used.

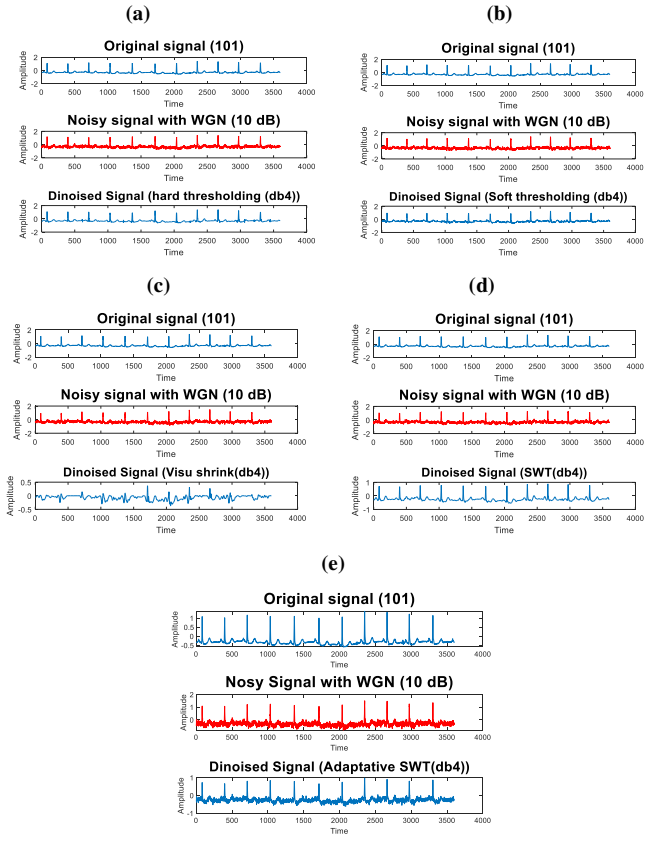


FIG. 3DENOISED SIGNAL WAVELET BASIS DB4 USING: (A) HARD THRESHOLDING, (B) SOFT THRESHOLDING, (C) VISU SHRINK THRESHOLDING, (D) SWT, (E)SWT ADAPTIVE(MAD)

TABLE I. EXPERIMENTAL RESULTS ON FIVE WAVELET TECHNIQUES

Wavelet basis	Signal number	Hard thresholding			Soft thresholding			Visu shrink			SWT			SWT adaptative(MAD)		
		MSE	SNR (dB)	PRD(dB)	MSE	SNR (dB)	PRD(dB)	MSE	SNR (dB)	PRD(dB)	MSE	SNR (dB)	PRD(dB)	MSE	SNR (dB)	PRD(dB)
Db4	101	0.0081	48.10	4.11	0.0081	52.17	3.49	0.0829	37.75	3.20	0.0032	51.87	3.88	0.0105	46.74	3.34
	103	0.0086	48.05	5.00	0.0086	53.00	5.04	0.0769	38.51	6.55	0.0034	52.08	5.00	0.0145	45.77	4.73
	115	0.0082	51.50	3.19	0.0082	55.79	3.38	0.1796	38.09	5.01	0.0033	55.50	3.11	0.0214	47.33	2.85
	122	0.0081	55.47	1.55	0.0081	59.68	1.50	0.3793	38.77	2.40	0.0034	59.19	1.38	0.0340	49.25	1.48
	123	0.0086	55.38	2.06	0.0086	**60.78**	1.97	0.4194	38.51	2.55	0.0031	59.83	2.14	0.0364	49.13	1.92
Db8	101	0.0012	56.31	3.66	0.0031	52.07	3.72	0.0859	37.59	4.19	0.0033	51.80	3.33	0.0089	47.43	3.39
	103	0.0011	56.86	4.86	0.0033	52.17	4.97	0.0842	38.12	5.91	0.0028	52.85	4.97	0.0094	47.65	4.65
	115	0.0011	60.15	3.25	0.0037	54.93	3.09	0.1903	37.83	3.85	0.0031	55.74	3.13	0.0150	48.86	3.01
	122	0.0014	63.11	1.38	0.0037	58.92	1.45	0.3799	38.76	2.67	0.0032	59.45	1.45	0.0317	49.54	1.57
	123	0.0012	**64.00**	1.98	0.0033	59.52	2.01	0.4386	38.32	2.39	0.0029	60.09	2.01	0.0300	49.96	1.90
Db12	101	0.0011	56.65	3.57	0.0035	51.50	3.86	0.0837	37.70	3.37	0.0031	51.99	3.41	0.0092	47.30	2.86
	103	0.0011	57.06	4.89	0.0035	51.96	4.90	0.0866	38.00	5.03	0.0029	52.81	4.94	0.0110	46.96	4.57
	115	0.0010	60.51	3.24	0.0035	55.25	3.37	0.1898	37.85	3.82	0.0032	55.63	3.11	0.0160	48.59	3.06
	122	0.0013	63.40	1.41	0.0036	58.98	1.44	0.4014	38.52	2.63	0.0033	59.38	1.44	0.0354	49.07	1.57
	123	0.0012	**63.87**	2.06	0.0033	59.57	1.99	0.4533	38.17	2.65	0.0028	60.27	1.95	0.0313	49.78	1.91
Bior3.1	101	0.0020	53.99	3.58	0.0050	49.95	3.59	0.0809	37.85	7.14	0.0043	50.59	3.88	0.0075	48.21	2.97
	103	0.0021	54.12	4.91	0.0049	50.49	5.02	0.0768	38.52	7.35	0.0040	51.39	5.00	0.0083	48.20	4.48
	115	0.0020	57.72	3.13	0.0050	53.65	3.16	0.1847	37.96	6.56	0.0043	54.33	3.31	0.0139	49.21	3.00
	122	0.0020	61.57	1.39	0.0054	57.26	1.46	0.4146	38.38	2.41	0.0042	58.34	1.50	0.0260	50.40	1.41
	123	0.0020	**61.64**	2.00	0.0054	57.38	2.19	0.4385	38.32	4.26	0.0045	58.21	2.14	0.0285	50.19	1.79

TABLE II. EXPERIMENTAL RESULTS ON FIVE WAVELET TECHNIQUE (AVERAGE MSE/SNR/PRD)

Wavelet basis	hard thresholding			Soft thresholding			Visu shrink			SWT			SWT adaptative(MAD)		
	Average MSE	Average SNR (dB)	Average PRD (dB)	Average MSE	Average SNR (dB)	Average PRD (dB)	Average MSE	Average SNR (dB)	Average PRD (dB)	Average MSE	Average SNR (dB)	Average PRD (dB)	Average MSE	Average SNR (dB)	Average PRD (dB)
Db4	**0.00832**	**51.7**	**3.182**	**0.00832**	**56.392**	**3.076**	**0.22772**	**38.32**	**3.982**	**0.00328**	**55.694**	**3.102**	**0.02336**	**47.644**	**2.864**
Db8	**0.0012**	**60.086**	**3.026**	**0.00342**	**55.522**	**3.048**	**0.2357**	**38.124**	**3.802**	**0.00306**	**55.986**	**2.978**	**0.019**	**48.688**	**2.8014**
Db12	**0.00114**	**60.298**	**3.034**	**0.00348**	**55.452**	**3.112**	**0.24296**	**38.048**	**3.5**	**0.00306**	**56.016**	**2.97**	**0.02058**	**48.34**	**2.794**
Bior3.1	**0.0020**	**57.808**	**3.002**	**0.00514**	**53.746**	**3.048**	**0.2391**	**38.206**	**5.544**	**0.00426**	**54.572**	**3.166**	**0.01684**	**49.242**	**2.73**

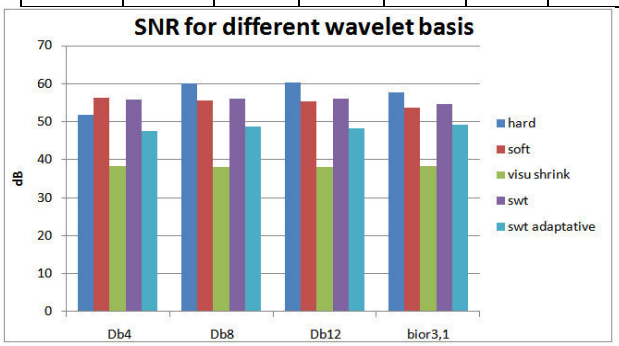


FIG. 4 SNR FOR DIFFERENT WAVELET BASIS

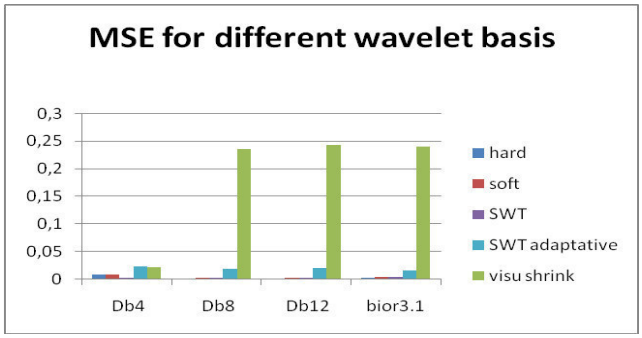


FIG. 5 MSE FOR DIFFERENT WAVELET BASIS.

B. Discussion

In Table Iwe compare the performance of different wavelet techniques for ECG signal processing using MSE, SNR and PRD.Table. IIshows the comparison of the mean values of each SNR, MSE and PRD indicator for different wavelet bases (Db4, Db8, Db12, and bior3.1) according to several thresholding methods (hard, soft, visu shrink, SWT, and adaptive SWT).

The results are examined according to threshold approaches and wavelet bases as follows:

1) Thresholding methods

From Fig.4 we can see that SNR values for hard thresholding are the highest for all bases (between 51.72dB and 60.29dB), and MSE values are the lowest (between 0.0011 and 0.008); relatively high PRD values, meaning that it retains the signal with the fewest errors. The soft thresholding and classic SWT techniques also perform very well, but the SNR and MSE results are slightly lower, although close to those of hard thresholding. As well as, the adaptive SWT method also performed well, with a slightly higher MSE and lower SNR than standard SWT. The Visu Shrink technique

consistently performed worst for all wavelet types tested: SNR (between 38.048and 38.32), and MSE (between 0.2277and 0.2429).

2) Comparison of wavelet bases

From Fig.5, we can confirm that the Db8 wavelet yields the highest signal-to-noise ratio (SNR) for hard thresholding, while the Db12 wavelet demonstrates slightly superior performance in terms of MSE using the identical technique. Wavelets Db4, Db8 and Db12 appear to perform similarly overall.

According to the results presented in Table II, the best technique of ECG denoising is that based on hard thresholding with the Db8 wavelet. The choice of thresholding technique and wavelet type therefore has a significant impact on performance. So, in conclusion, according to these experimental results, the best ECG signal processing technique among those tested is hard thresholding used in conjunction with the Db8 wavelet. This combination optimizes both noise reduction and signal distortion.

V. Conclusion

ECG processing using the wavelet method has been extensively studied in practice, with satisfactory results in terms of improving the reliability of clinical diagnosis. The performance of wavelet denoising depends on several factors, such as the choice of wavelet technique, the thresholding method and the threshold value. In overall, our results demonstrate the effectiveness of the hard threshold technique in improving the diagnosis of cardiac examinations, removing noise from the ECG signal without distorting the basic signal, compared with other denoising techniques in terms of SNR, MSE and PRD. The wavelet family (Db8) also delivered the best results in terms of SNR, MSE and PRD.

VI. References

[1] Madan, P., Singh, V., Singh, D. P., Diwakar, M., & Kishor, A. (2022). Denoising of ECG signals using weighted stationary wavelet total variation. Biomedical Signal Processing and Control, 73, 103478.

[2] Kumar, A., Tomar, H., Mehla, V. K., Komaragiri, R., & Kumar, M. (2021). Stationary wavelet transform based ECG signal denoising method. ISA transactions, 114, 251-262.

[3] Vitek, M., Smital, L., Kozumplík, J., & Provaznik, I. (2013). Adaptive wavelet Wiener filtering of ECG signals. IEEE Trans Biomed Eng 60:437–445

[4] El-Dahshan, E. S. A. (2011). Genetic algorithm and wavelet hybrid scheme for ECG signal denoising. Telecommunication Systems, 46, 209-215.

[5] Alfaouri, M., & Daqrouq, K. (2008). ECG signal denoising by wavelet transform thresholding. American Journal of Applied Sci-ences, 5, 276–281.

[6] Gupta, V., & Mittal, M. (2020). Arrhythmia detection in ECG signal using fractional wavelet transform with principal component analysis. Journal of The Institution of Engineers (India): Series B, 101(5), 451-461.

[7] Zhang, D., Wang, S., Li, F., Wang, J., Sangaiah, A. K., Sheng, V. S., & Ding, X. (2019). An ECG signal de-noising approach based on wavelet energy and sub-band smoothing filter. applied sciences, 9(22), 4968.

[8] Al-Safi, A. (2019). ECG signal denoising using a novel approach of adaptive filters for real-time processing. International Journal of Electrical and Computer Engineering (IJECE), 9(1).

[9] Thamarai, P., & Adalarasu, K. (2018). Denoising of EEG, ECG and PPG signals using wavelet transform. Journal of Pharmaceutical Sciences and Research, 10(1), 156-161.

[10] Gualsaquí, M., Vizcaíno, I., Rosero, V. G. P., & Calero, M. J. F. (2018). ECG signal denoising using discrete wavelet transform: A comparative analysis of threshold values and functions. Maskana, 9(1), 105-114.

[11] P.Ziran., & Wang, G. (2017). A novel ECG eigenvalue detection algorithm based on wavelet transform. BioMed research international, 2017.

[12] Mandala, S., Fuadah, Y. N., Arzaki, M., & Pambudi, F. E. (2017, May). Performance analysis of wavelet-based denoising techniques for ECG signal. In 2017 5th International Conference on Information and Communication Technology (ICoIC7) (pp. 1-6). IEEE.

[13]Hélène, L., Philippe, C., & Catherine, M. (1997, September). Utilisation de la transformée en ondelettes non décimée pour le débruitage du signal électrohystérographique utérin. In Seizième Colloque Gretsi (pp. 15-19).

Soft Sensing Based On Artificial Neural Network Regression And Data Reconstructed by Self-Organizing Maps for Water Quality Monitoring.

Mohamed Imed KHELIL
LGE, Laboratoire de Génie Electrique,
Department of Electrical Engineering
Faculty of Technology,
University of M'sila, Algeria
mohamed-imed.khelil@univ-msila.dz

Mohamed LADJAL
LASS, Laboratoire d'Analyse des Signaux et Systèmes, Department of Electronics,
Faculty of Technology, University of M'sila, Algeria
mohamed.ladjal@univ-msila.dz

***Abstract*— It has been found that the coagulation-flocculation unit is considered as a key step in the production of drinking water. Our work consists in developing a soft sensor to predict the optimal coagulant dose to be injected based on the raw water characteristics using an artificial neural network (ANN) model. This paper studies the application of the soft sensing modeling approach for surface water quality monitoring using advanced intelligence techniques such as self-organizing maps (SOM) and artificial neural networks (ANN). The reconstructed SOM data were subjected to a second level for the neural network regression algorithm. In water treatment processes, many monitoring parameters are expensive or difficult to measure in real time, which limits the possibility of effective control of water production processes. In this work and with this in mind, an intelligent soft sensor was developed to predict the optimal coagulant dosage based on the reconstructed outlier parameters.**

Keywords*— *coagulation, self-organizing map, ANN, prediction,

I. INTRODUCTION

Producing better quality drinking water at lower cost has always been our goal. Only computational intelligence techniques can describe and compare water quality parameters and make decisions in real time. We can use artificial intelligence techniques for analysis in a variety of water treatment applications, visualization and modeling of complex multivariate data. Today, it is unanimously recognized that the increased interest in artificial neural networks (ANNs) is most suitable for solving resource-related problems, which has led to an increase in research activities in this field. [1] Artificial neural networks are increasingly used in the modeling of hydrological processes, which is related to their ability to link input and output variables in complex systems without detailed knowledge of the physics of the processes involved [2]. In this context, there is a process that is considered one of the most important steps in water treatment. This coagulation process removes colloidal particles from water. To this end, our work is mainly based on the development of a hybrid system that leads us to two main phases. First, it is the reconstruction of outlier data of water parameters based on the self-organizing map (SOM) which is a tool for data analysis, visualization and interpretation, which is based on the principles of quantification and vector similarity measures. Although most SOM procedures can be considered exploratory, this method can be used to perform broad categories of operations such as function fitting, prediction or estimation, clustering, pattern recognition or noise reduction and classification. [3] The second phase is to predict the coagulant dose from the SOM reconstructed values using artificial neural networks such as multilayer perceptron (MLP) networks, which are powerful models for solving nonlinear mapping problems [4]. The objective of this study is the application of soft sensor development approach to water quality assessment, control and monitoring using SOM and ANN. This study focused on the assessment and interpretation of surface water of Cheliff Dam (Algeria). In the first phase of this paper, SOM is used for the reconstruction of outlier or missing data. The second phase is to predict the optimal coagulant dosage using the artificial neural network (ANN) method in soft sensor modeling, which is a powerful model, used after the SOM networks process. The numerical results of soft sensor modeling demonstrate the capability of our proposed approach. [5] The remainder of this paper is organized as follows. The next section describes the research domain and briefly describes the data collection selected for soft sensor modeling. Section 3 briefly describes the proposed methods used for the preprocessing of raw water characteristics and Section 4 presents and discusses the experimental results. Finally, the last section is devoted to conclusions.

II. MATERIALS AND METHODS

A. Study Area and dataset

The surface water of Tilesdit dam is geographically located about 20 km southeast of the city of Bouira and 122 km east of Algiers (Fig. 1). It is situated between the geographic coordinates: 36° 13' 22''N 4° 14' 23''E.

In this research, physical sensors installed in the Tilesdit production station provided us with several important physical and chemical parameters. Our mission was limited during three years (2016- 2018) to know the different treatment process and to collect data from this station. Descriptive measures of the used water quality parameters are shown in Table 1. [6] The dataset used for the purposes of this study includes more than 9000 results. The data are arranged in Table 1, which contains 05 variables selected after using an artificial intelligence technique adopted in previous research [6] and 1800 samples, while being variable in terms of time. Water measurements were carried out between 2016 and 2018, 3 full years of measurements were included at the Tilesdit dam in the wilaya of Bouira in Algeria. The indicators concerned were turbidity (Ntu), PH, temperature (°c), conductivity (µs/cm) and coagulant.



Fig. 1. The map shows the studied area: Tilesdit Dam – Bouira – Algeria.

Variables	Min	Max	Mean	Standard deviation
pH	7.07	8.76	7.8076	0.2789
Conductivity (µs/cm)	515	649	581.1567	22.7408
Temperature (°C)	9.8	27.2	17.1285	3.9287
Turbidity (NTU)	1.07	10.9	3.6631	1.3571
Coagulant (mg/L)	0.64	8.05	3.8148	1.6298

TABLE I. Descriptive statistics of water parameters

B. Self Organizing Map (Kohonen)

A SOM is kind of a neural network that is trained to create a 2-dimensional discrete representation (called a mapping) of the input space of training examples. More specifically, it is a nonlinear, ordered, smooth mapping from high-dimensional input data to elements of a regular low-dimensional array. SOM plots can be used to classify and visualize high-dimensional data because they can transform nonlinear statistical relationships between high-dimensional data into simple geometric relationships of image points on a regular two-dimensional grid of nodes [8][9]. Dissimilar of many other types of NN, SOMs do not require the specification of a target output. Instead, the regions of the grid where the node weights match the input vector are selectively optimized to better resemble the data of that class. Starting from an initial distribution of random weights, after many iterations, the SOM eventually stabilizes to a stable region map. Each region is actually a feature classifier. [8][9] The training process of the SOM can be described based on the representation of the input pattern and the fitting of the model vector. Each training iteration starts with a random selection of an input pattern x, x Є Rn, the pattern is presented to the SOM, and each unit determines its activation. Usually, the activation of a unit is calculated using the Euclidean distance between the input pattern and the model vector. In this case, the unit whose model vector has the smallest Euclidean distance with the input pattern is called the winner. We use the subscript c to denote **the** winner, see equation (1).

$$C(t)=\arg\min_i \{ \| x(t) - m_i(t) \| \} \qquad (1)$$

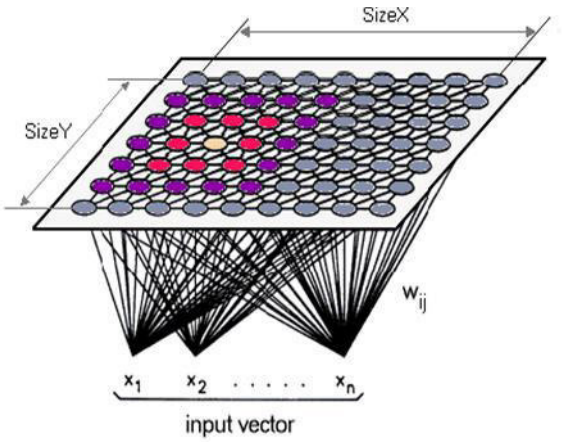


Fig. 2. SOM Training.

Finally, the model vector of the winner and the model vectors of the units in the vicinity of the winner are adjusted. This adjustment is achieved by gradually reducing the difference between the corresponding components of the input pattern and the model vector, as shown in equation (2). Note that we use discrete-time notation, where t represents the current training iteration.

$$M_i(t+1) = m_i(t)+\alpha(t) \cdot h_{ci}(t) \cdot [x(t) - m_i(t)] \quad (2)$$

Geometrically speaking, the model vector of the fitted unit is slightly shifted toward the input pattern. The degree to which the model vector is shifted is determined by the learning rate α, which decreases over time. The number of units affected by the adjustment and the strength of the adjustment, which depends on the distance of the unit from the winner, is determined by the neighborhood function hci. The number of training times is also reduced over time, so that only the winner is adjusted at the end of the training process.

The neighborhood function is usually a unimodal function that is symmetric with respect to the location of the winner and monotonically decreases as the distance to the winner increases. A Gaussian function is often used as the neighborhood function, as shown in equation (3).

$$h_{ci}(t) = \exp((-\| r_c - r_i \|^2)/(2.\delta(t)^2)) \quad (3)$$

In the formula, $\| r_c - r_i \|^2$ represents the distance between cells c and i in the output space, where ri represents the two-dimensional position vector of cell i in the grid. The time-dependent parameter **δ** controls the reduction of the neighborhood kernel during training. Usually, a sufficiently large neighborhood kernel is chosen at the beginning of learning to cover a wide range of the output space. During the training process, the spatial width of the kernel is gradually reduced until the end of the process, and only the winner is adjusted. [10]

C. Reconstruction data algorithm

If the vector prototypes provide a good representation of the data, any missing values of a given input variable can be estimated by the values of the corresponding components of the winning prototype. Let x be a new vector consisting of two parts, xo and xm, containing the observed and missing values, respectively. Mainly, xm is reconstructed based on the information provided by the Kohonen map

The proposed method is based on the similarity between this new vector x = (xo, xm) and the reference vectors mk. Given xo and xm, the subspaces, respectively, of the variables

xo and xm. mo and mm are the projections of these subspaces. Depending on the activation defined by Eq. 4, the closer xo is to mo, the more likely we are that xm will be closer. [7] [11]

$$K_i(x^o) = \exp\left(\frac{-1}{\sigma_i^2} \|x^o - m_i^o\|\right) \quad (4)$$

Where σ_i^2 is a parameter defining the size of the region of influence of the unit i. σ_i^2 Can be calculated as the mean empirical variance of the n input characteristics, among the samples associated with the unit i.

Approaches to estimating missing data use different techniques, generally assuming a probabilistic context. For example, heuristic methods (such as middle and middle replacement techniques) are often used and are simple and inexpensive solutions. Parametric maximization methods, such as the EM algorithm (expectation maximization), are widely used and proven to be effective, but they require knowledge or the laws of estimation of variable probabilities. We can use a simple approach in our case to estimate the missing data based on the values of the components corresponding to the winning prototype m_b:

$$\forall\, l \in M(x),\ \hat{x}_l = m_{bl} \quad (5)$$

Where: M (x) is the index set of missing values.

This method is very sensitive to the change of prototypes between two consecutive vectors x. To address this problem, we consider an alternative approach that takes into account the influence of the k closest prototypes. Each missing or invalid value j is estimated by the combination of the corresponding components in the k closest prototypes:

$$\hat{x}_j = \frac{\sum_{i=1}^{k} h(i) m_i(j)}{\sum_{i=1}^{k} h(i)} \quad (6)$$

Where m_i (j) denotes the component j of prototype i. [12],

D. Multi-Layer Perceptron

Fig.3 demonstrates the structure of a standard two-layer perceptron. The inputs form the network's input nodes; the outputs are processed by the output nodes. The intermediate layers of nodes are invisible to either the input or the output and are called hidden layers, which, unlike the input and output layers, are not physical. Hidden layers are often used to create bottlenecks, forcing the network to create a simple model of the system that generated the data and be able to generalize to previously unseen patterns.

The function of this network is defined as follows:

$$Y_i^{(H)} = f^{(H)}\left(\sum_j w_{ij}^{(HI)}\ xj\right)$$

$$Y_i = f^{(T)}\left(\sum_j w_{ij}^{(TH)}\ yj(H)\right)$$

This dictates how the input pattern vector *x* is mapped to the output pattern vector $y^{(0)}$ by the hidden pattern vector $y^{(H)}$ in a manner parameterized by two weight layers $w^{(HI)}$ and $w^{(TH)}$. The univariate function f(x) is usually fixed in each case.

$$f(x) = \frac{1}{1 + e^{-x}}$$

It varies smoothly from 0 at -∞ to 1 at ∞, as a threshold function would do abruptly.

When the number of hidden layer nodes is smaller than the number of degrees of freedom inherent in the training data, the activations of the hidden nodes tend to form a set of orthogonal variables that are linear or nonlinear combinations of the attribute variables, solving the problem as best as possible. When we add a small number of constraints to the network, these internal variables form a linear or nonlinear principal component representation of the attribute space. When noise that is not an inherent part of the generative system is added to the data, the principal component network acts as a filter of the noise signal with low variance, provided that the signal-to-noise ratio of the data is high enough. This property allows MLPs to generalize to previously unseen patterns by modeling only the important underlying structure of the generative system. Hidden nodes can be seen as detectors of abstract features in the attribute space [13]. Each hidden layer and output layer also has bias neurons (see Figure 3). A feedforward MLP must be trained using a training algorithm that creates a learning process to find the optimal set of weights for the connections. bias values of neurons. Backpropagation is the most common algorithm for training feedforward MLPs. In a feedforward MLP, the input patterns presented to the neurons in the input layer are propagated through the network from the input layer to the output layer. h. In the forward direction. The output of the network is compared to a target value and an error is calculated. The calculated error is propagated through the network to update the connection weights and bias values [14]. This process is repeated until convergence. Once training is complete, the network can generate outputs for unseen inputs. For a detailed review of ANN training algorithms and their applications to water resources and hydrology, see Maier and Dandy. [14][15][16][17]

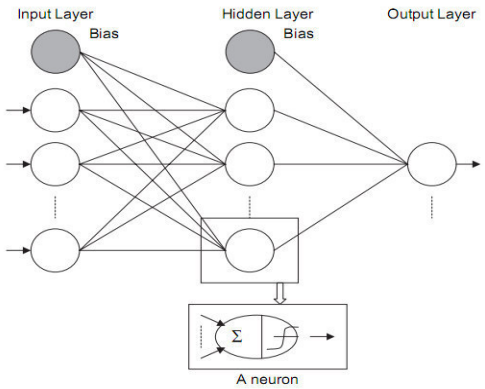


Fig. 3. Structure of a feed-forward multi-layer perceptron (MLP) [18]

III. Experimental Results and Discussion

The feature selection process is a well-known step, especially in the field of machine learning, to ensure better decisions, higher performance, and better decision-making capabilities for a given model. The first analysis is the actual process of selecting a subset of relevant and distinct variables based on the SOM to determine the new reduced input information for the proposed decision support model. Electrical conductivity (EC), turbidity (TU), temperature (T°), and pH. Four parameters were chosen as inputs to the proposed monitoring model. These same parameters can be more easily measured continuously and used on-site when sensors are installed directly at each stage of the production station. [6]

A. *Reconstruction of data*

In the previous work, we presented the application of the SOM algorithm to the detection of atypical data. We will now use this model to deal with invalid or missing data. This method allows representing the knowledge acquired from the data in the form of prototypes or reference vectors. The method proposes to estimate all the missing or outlier values of a vector in the same system, regardless of the number of available variables. However, it is obvious that the accuracy of the reconstruction will depend on the number of available variables compared to the number of missing variables. A defect was artificially introduced between the 500th and 510th day at the sensor that measures the temperature (Fig. 4). Figure 5 and Table 2 presents the reconstruction performance metrics for this method. After visualizing which descriptor should be declared precisely as defective (in our case the temperature sensor), the contributions of each of the components of the vector x are then also examined. These abnormal samples are then removed to compute a new winning prototype with only normal samples. Figure 4 shows the malfunction of the descriptor. The abnormally low or high variations were correctly identified as the faulty parameter. Figure 5 shows the reconstruction values for the faulty temperature values. The SOM procedure allows the rejection of atypical samples and therefore implements a kind of "novelty detection". However, this type of rejection can come from unreliable data acquisition sources, faulty sensors, data collection errors or simply a lack of completeness of the data. This is a conservative approach to data changes. Figures 4 and 5 summarize and illustrate the different erroneous data recognized using the SOM approach and in parallel, we present the estimated results of the latter. In addition, the SOM approach allows to estimate the variations of these samples. The reliability and robustness of this approach are justified by the validation-reconstruction process of the defective measurements.

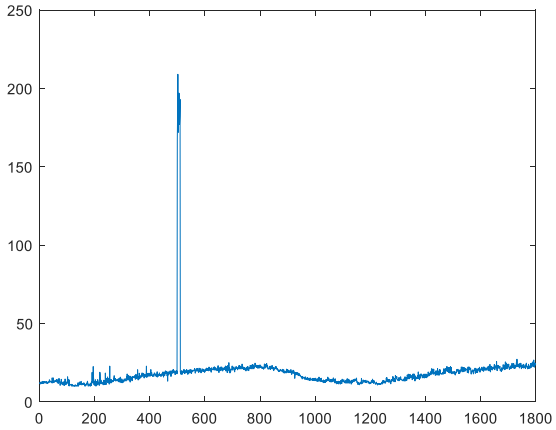


Fig.4. Temperature measurement with artificial fault.

	RMSE	**R^2**
Temperature	0.206	0.997

TABLE II. Performance of reconstruction data.

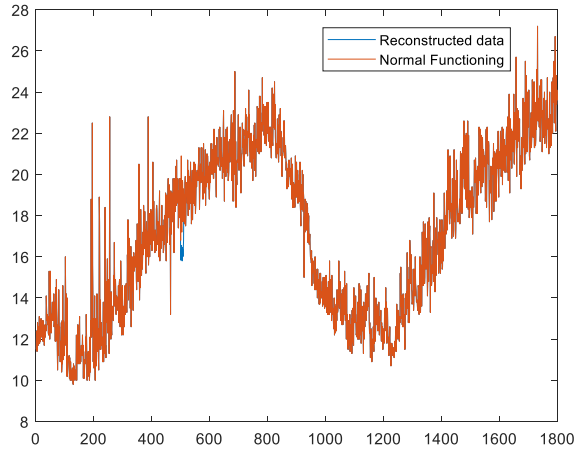
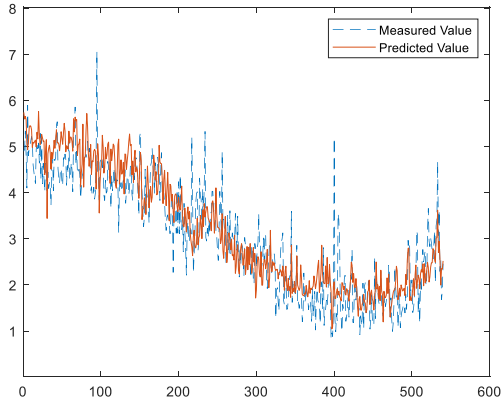


Fig.5. Reconstruction of Temperature faulty.

B. *Soft sensor*

In order to evaluate the robustness of this method adopted in this paper, a simulation study was carried out by artificially introducing a defect in the original data at certain time steps. The simulation was based on 10 days of real data between the 500th day and the 510th day at the temperature sensor as shown in Figure 4. After using the fault detection methods, the Temperature variable was correctly identified as the faulty parameter. Figure 5 shows the reconstruction of Temperature using the SOM approach. The obtained values are used to predict the optimal coagulant dosage to be injected into the raw water. Predicting the optimal coagulant dosage from water characteristics is a nonlinear regression problem that can be addressed using multilayer perceptrons (MLPs). We used a conventional MLP with a hidden layer of sigmoid units. To train the MLPs, a training set of 1800 complete measurement vectors was constructed by removing erroneous data. A total of 1260 samples (about 70%) were used for the training phase to build the model, with the rest used as an independent test set. The model training included: finding the best ANN structure and estimating the prediction accuracy. The prediction accuracy and confidence limits calculated on the validation set are shown in Figure 6. The ANN prediction and accuracy are shown in Figure 7 for the preprocessed data. This should be compared to the prediction results without preprocessing as shown in Figure 8. These results clearly demonstrate the robustness induced by the preprocessing module in our system.



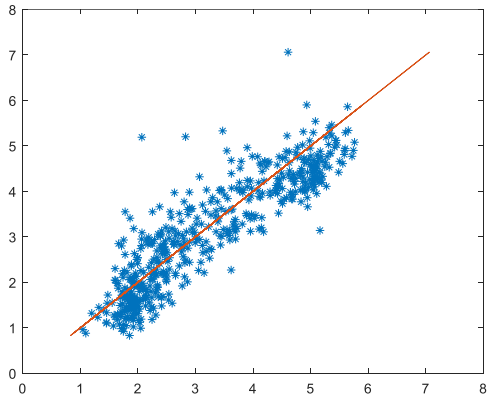


Fig.6. predicted and target coagulant dose.

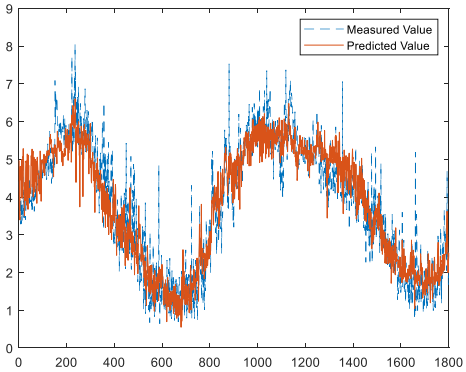


Fig.7. Actual (dotted line) and versus predicted (continue line) coagulant dosage with ANN model.

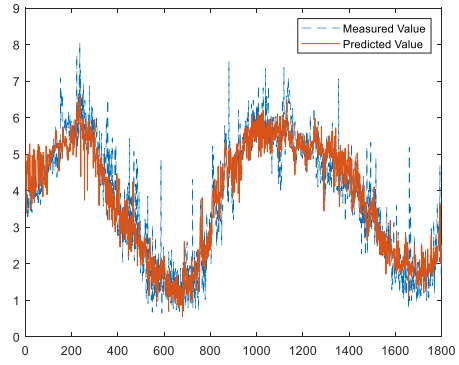


Fig.8. Actual (dotted line) and versus predicted (thick line) coagulant dosage with ANN model without pre-processing.

	ANN model with pre-processing.	**ANN model without pre-processing.**
RMSE	0.639	0.635
R^2	0.846	0.848

Table. III. Performance measurements of the ANN model of the optimal coagulant dose software sensor with and without pre-processing.

It can be easily seen that the training points in Figure 6 are distributed close to the perfect line, which means that the optimal coagulant dosage predicted by the proposed ANN method has good accuracy. The two evaluated indices of all samples with and without pre-processing are shown in Table III. In this table, the RMSE of the dataset with and without pre-processing is relatively stable and at a low level, and the R2 is at a high level, which means that the optimal coagulant dosage value predicted by the SOM-ANN-based model has good accuracy and stability.

IV. CONCLUSION

In this paper, we studied the integration of two approaches for outlier reconstruction by SOM and coagulant dose prediction by ANN. Experimental results using real data from the drinking water treatment process have shown the efficiency and robustness of these algorithms. A SOM-based algorithm is presented for outlier reconstruction. It is shown that the estimated measurements of a sensor identified as faulty can be reconstructed using an extension of the SOM algorithm. These reconstructed data are validated by an ANN-based software sensor. Experimental results using real data from a drinking water treatment study area demonstrate the effectiveness and accuracy of the approach used.

REFERENCES

[1] A.M. Kalteh*, P. Hjorth, R. Berndtsson. 2008. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. Environmental Modelling and Software 23. 835-845.

[2] Dawson, C.W., Wilby, R.L., 2001. Hydrological modelling using artificial neural networks. Progress in Physical Geography 25 (1), 80-108.

[3] S. J. Fraser, B.L. Dickson. A New Method for Data Integration and Integrated Data Interpretation: Self-Organising Maps. Data Visualization and Integration.2007, p. 907-910.

[4] Cornelius T. Leondes. Algorithms and Architectures. Volume 1, neural network systems techniques and applications. Academic Press. 1998.

[5] Mohamed Imed Khelil, Mohammed Assam Ouali, Mohamed Ladjal, Hamza Bennacer, Mohamed Djerioui, Soft Sensing Modeling Based on Support Vector Machine and Self-Organizing Maps Model Selection for Water Quality Monitoring.

[6] Mohamed Imed Khelil, Mohamed Ladjal, Youcef Brik, Mohammed Assam Ouali, Self-Organizing Maps-Based Features Selection with Deep LSTM and SVM Classification Approaches for Advanced Water Quality Monitoring, International Journal of Intelligent Engineering and Systems, Vol.15, No.3, 2022.

[7] Mohamed Imed Khelil, Mohamed Ladjal, Mohammed Assam Ouali, Bennacer Hamza, Sensor Anomaly Detection using Self Features Organizing Maps and Hierarchical Clustring for Water Quality Assessment

[8] Gavrilas M. Inteligenta artificiala si aplicatii in energetica, Editura Gh. Asachi, Iasi, 2002.

[9] Otilia Elena Dragomira, Florin Dragomirb, Marian Radulescuc. Matlab Application of Kohonen Self- Organizing Map to Classify Consumers' Load Profiles. Procedia Computer Science 31 (2014) 474 – 479.

[10] Andreas Rauber, Dieter Merkl, and Michael Dittenbach, The Growing Hierarchical Self-Organizing Map: Exploratory Analysis of High-Dimensional Data, IEEE TRANSACTIONS ON NEURAL NETWORKS.

[11] YING ZHAO, GEORGE KARYPIS. 2005. Hierarchical Clustering Algorithms for Document Datasets. Data Mining and Knowledge Discovery, 10, 141–168.

[12] B. Lamrini, El-K. Lakhal, M-V. Le Lann, L. Wehenkel. 2011. Data validation and missing data reconstruction using self-organizing map for water treatment. Neural Comput & Applic.

[13] D. Michie, D.J. Spiegelhalter, C.C. Taylor. Machine Learning, Neural and Statistical Classification. 1994.

[14] ASCE Task Committee on application of Artificial Neural Networks in Hydrology, 2000a. Artificial neural networks in hydrology. I:

preliminary concepts. Journal of Hydrologic Engineering 5 (2), 115-123.

[15] Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling and Software 15, 101-124.

[16] ASCE Task Committee on application of Artificial Neural Networks in Hydrology, 2000b. Artificial neural networks in hydrology. II: hydrologic applications. Journal of Hydrologic Engineering 5 (2), 124-137.

[17] Dawson, C.W., Wilby, R.L., 2001. Hydrological modelling using artificial neural networks. Progress in Physical Geography 25 (1), 80-108.

[18] Kalteh, A.M., Berndtsson, R., 2007. Interpolating monthly precipitation by self-organizing map (SOM) and multilayer perceptron (MLP). Hydrological Sciences Journal 52 (2), 305-317.

Fabric Fault detection based on image processing and SVM classifier

Boualem IKHLEF
Solid Mechanics and Systems Laboratory (LMSS)
University of Boumerdes
Boumerdes- Algeria
b.ikhlef@univ-boumerdes.dz

Mohammed Amine Sahraoui
Systems Engineering and Telecommunications Laboratory (LIST)
University of Boumerdes
Boumerdes- Algeria
ma.sahraoui@univ-boumerdes.dz

Chemseddine RAHMOUNE
Solid Mechanics and Systems Laboratory (LMSS)
University of Boumerdes
Boumerdes- Algeria
ch.rahmoune@univ-boumerdes.dz

Abstract´ Automated inspection of fabric quality in the textile industry is a very important task, in fact, it ensures good production quality and therefore leads a profitable economy for the company, unfortunately in nowadays we find that this task is always done manually. In this work we will present a method of detection and classification of fabric defects using image processing and artificial intelligence. The proposed technique consists of three steps, in the first step the color (RGB) images of fabric are converted into HSV images then into grayscale images before being filtered, the second step consists of features extraction using eleven statistical indicators on the images obtained from the first step, the last step consists of using the support vector machine technique (SVM) of the artificial intelligence family to detect and classify the different classes of defects. After several simulations and a comparison with other classifiers, the technique gave good results.

Keywords´ image processing, fault detection, textile industry, fabric fault, SVM classifier

I. INTRODUCTION

Production quality is one of the main objectives of any industry, so match so that companies deploy enormous material and human resources to guarantee good quality of their products, in the same context, the textile industry must also ensure good quality of the fabric produced.

However, we find that manual inspection is still applied today, although this inspection only detects 70% of production defects.

In cases where the fabric is basic, fault detection may be acceptable, but in complex patterned fabrics, fault detection remains a complicated task, there is a trade-off in manual checking between quality and quantity of production, in fact, the meticulous inspection of the product leads to good quality but not a significant quantity and a superficial inspection of the product leads to significant production but a poor quality.

The use of automated inspection of fabric quality has become essential to have good production quality and sufficient quantity.

In recent years, much research has been carried out on fabric inspection techniques based on image processing and artificial intelligence.

Following the chronology of the work, Wood et al. used the Fourier power spectrum of the images to determine carpet defects [1]. Campell et al. used the discrete Fourier transform for error detection on denim fabric with a multi-layer artificial neural network [2]. Tsai et al. Used the Fourier transform-based spherical image refinement scheme to detect defects on fabric [3]. Sari-Sarraf et al. used edge fusion and wavelet transform in preprocessing of images [4]. Kumar et al performed Gabor wavelet feature selection with a guided error approach in detecting fabric faults [5]. [6] and [7] used deep learning technique. [8] Shahrabadi and al used Convolutional Neural Network (CNN) based on known architectures such as AlexNet or Visual Geometry Group on defect analysis

The proposed method used the image processing to pre-trait the image of fabric and a combination of eleven statistical indicators to extract features from the images, after that random forest (RF), decision tree (DT), K near-neighbor (KNN) and support vector machine (SVM) classifiers are used to classify the defects into different classes.

The flowchart of proposed technique is shown in the following figure.

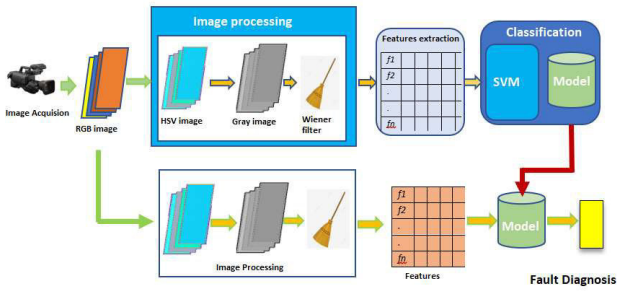


Fig1 : Flowchart of proposed method

II. FABRIC DEFECTS

Fabric defects are closely related to 85% of defects observed in the textile industry [9]. There are many types of fabric defects, part of them is caused by machine malfunctions. Others are caused by defects in threads or machine waste. waste machine oil is often with management along the chain direction, and they are wide and irregular.

In this work we have used three defects and healthy fabric under different contrast which are shown in the following figures.

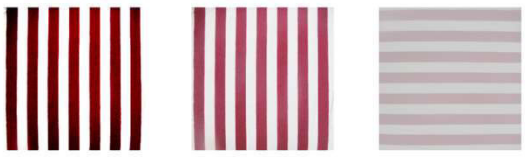


Fig2. Healthy fabric

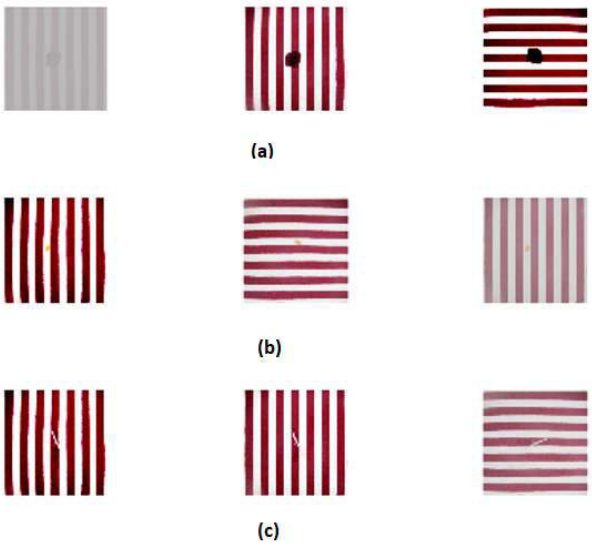


Fig3. Fabric defects

The images of (a) group simulate wide waste machine oil defect with different contrast level and in different position, the images of (b) group simulate a small and clear waste machine oil and in different position, the images of (c) group simulate a tear defect with different contrast level and in different position.
We have used 240 images of different fabric defects, 60 images for each class, so we have four classes considering healthy one.

III. PROPOSED TECHNIQUE

A. preprocessing image

The color R, G and B in RGB are all correlated to color luminance. we cannot separate color information from luminance. HSV or Hue Saturation value is used to separate image luminance from color information. This makes it easier when we are working or need the luminance of the image.
The Following figure shows an image RGB and HSV image correspondant

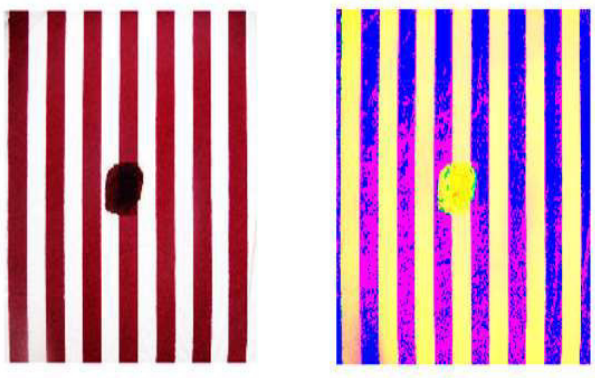


Fig4. RGB and HSV Image

To eliminate all forms of chromatic information and leave only different shades of gray, Grayscale image is used and result is shown in the following figure



Fig5. Grayscale Image

Before the step of extraction features, we have used the Wiener filter to filter the gray image and the result is send to the extraction phase.

B. Features extraction

The features are extracted by using eleven statistical indicators on the all images, these statistical indicators are shown in the table below

Table 1: Table statistical indicators.

Name	formula
Root mean square	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert^2}$
Crest factor	$C=\frac{\lVert x\rVert_\infty}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert^2}}$
Peak to peak	$\mathrm{Max}(x)-\mathrm{Min}(x)$
skewness	$E\left(\frac{X-m}{s}\right)^3$
kurtosis	$E\frac{(x-m)^4}{s^4}$
Mean	$\frac{1}{N}\sum_{i=1}^{N}A_i$
Std	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\lvert A_i-m\rvert^2}$
Var	$\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^3$
Root sum square	$X_{rss}=\sqrt{\sum_{n=1}^{N}\lvert X_n\rvert^2}$
Max	$\mathrm{Max}\lvert x_i\rvert$
Min	$\mathrm{Min}\lvert x_i\rvert$

C. Support Vector Machine (SVM)

SVM is part of supervised learning methods, is used to answer regression and classification problems.
its principle is to look for a hyperplane which separates classes by maximizing the distance separating data points from this hyperplane as shown in the following figure.

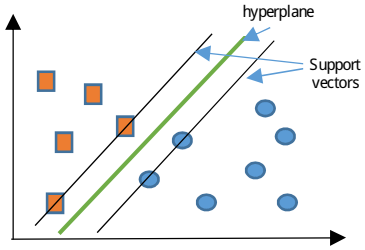


Fig 6. Example of Support Vector Machine

The optimized hyperplane can be mathematically expressed as [10].

(1)

where w is the vector of weights, x is an input vector, and b represents the bias. The equations of the support vectors of each class are given as :

for di=+1
for di=-1 (2)

where di corresponds to the respective class, i.e.,
di=+1 for class A,
and
di=Б1 for class B.
The optimization problem for training sample to find the optimal hyperplane is given as

(3)

Such as

The final decision function can be obtained as follows:

(4)

where x denotes the input vector to be classified and N is the number of support vectors obtained in training phase. The non- negative parameters are used to define support vectors among input vectors.
The linearly non-separable patterns are transformed into a higher dimension feature space, using a mapping function – (x), allowing for classification of the data using the linear hyperplane. The decision function in equation (4) can be modified to

(5)

The inner-product kernel function, defined as is used to reduce the complexity of numerical optimization in high-dimensional space. The decision function can be updated as follows:

(6)

For multi-class classification problems, the SVM can be used in a one-versus-rest manner [11], [12], [13].

IV. Result and Discussion

The data are divided into two groups, 70% used for training and 30% used for testing, so we have 168 images that will be used for training and 72 images for testing.

The simulations gave the result shown in the following figures.

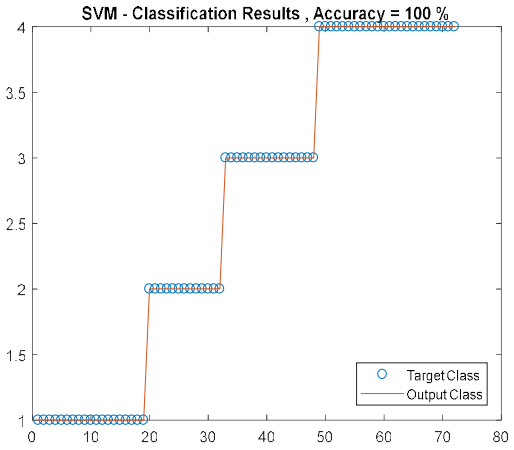


Fig 7. Classification result with SVM as classifier

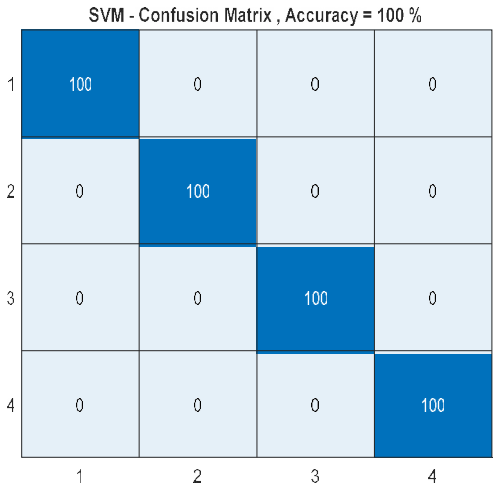


Fig 8. Confusion matrix with SVM as classifier

These two figures above show that the accuracy is 100%, which means that there is not misclassified defect and the technique is very good.

To verify the superiority of the SVM classifier on the other classifiers, we have used three others classifiers which are: Random Forest, K-near neighbor and Decision Tree, the result is shown in the following figures:

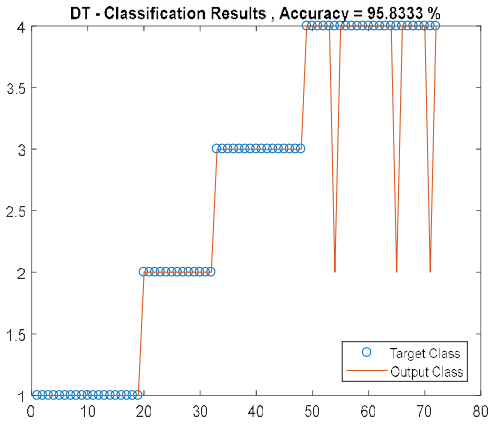


Fig 9. Classification result with DT as classifier

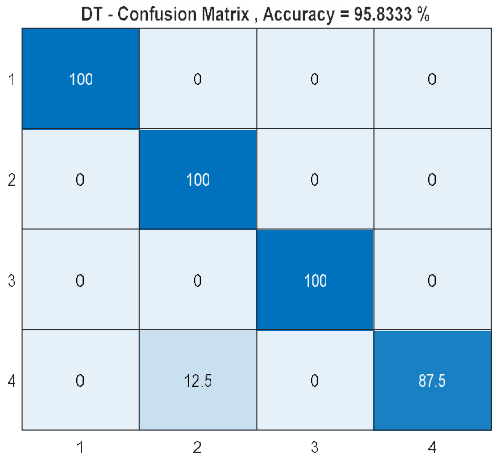


Fig 10. Confusion matrix with DT as classifier

From the figures 9 and 10, we see that DT technique gave an accuracy of 95.83% which is less than the accuracy with SVM as classifier, there are only three misclassifications, they were predicted in the fourth class while they belong to the second class.

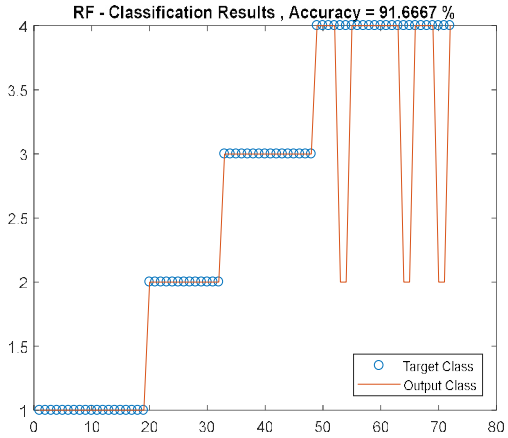


Fig 11. Classification result with RF as classifier

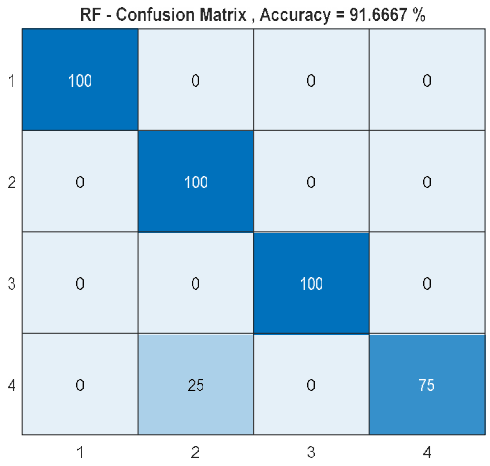


Fig 12. Confusion matrix with RF as classifier

From the figures 11 and 12, we see that RF technique gave an accuracy of 91.66% which is less than the accuracy of SVM and DT techniques, there are six misclassifications, they were predicted in the fourth class while they belong to the second class.

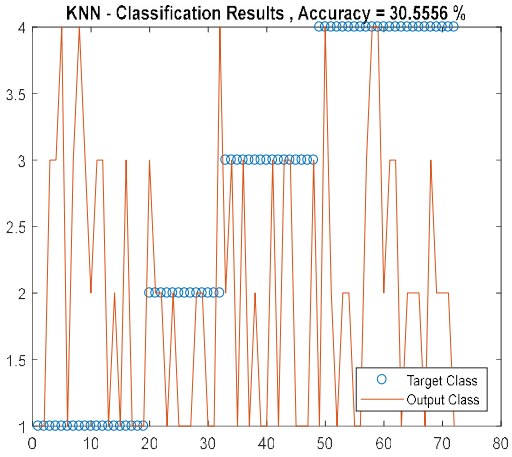


Fig 13. Classification result with KNN as classifier

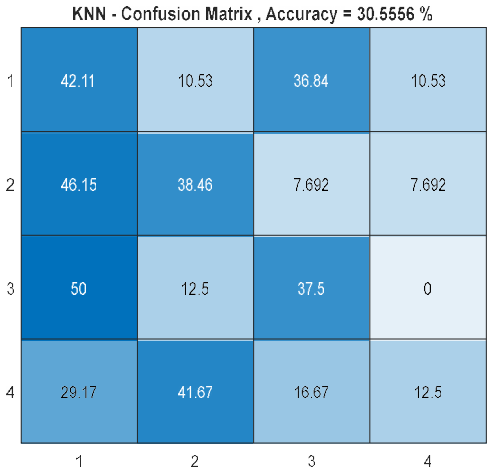


Fig 14. Confusion matrix with KNN as classifier

From the figures 13 and 14, we see that KNN technique gave a bad result, with an accuracy equal to 30.55% which is far

behind RF and DT techniques, there are twenty-two misclassifications.

A. Stability study

To verify if the SVM technique gives always the best result, we have done fifteen experiences with the different techniques and we have computed the mean, and the standard deviation of the accuracy for each classifier, the result is shown in the following table

Table 2: Classification result

	SVM	DT	RF	KNN
Mean	97.50	95	95.27	22.96
Std	3.07	4.22	3.47	4.16

From the table, we can see that the DT is not stable even it had a good accuracy in the first simulation, we remark also that the SVM is stable and has better accuracy than the other techniques as shown in the following figure

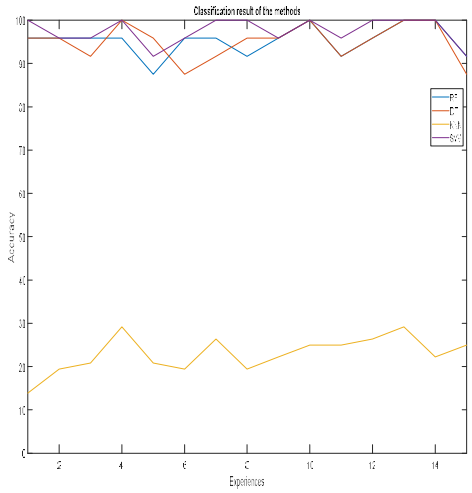


Fig 15. Variation of the accuracy of the all classifiers

The figure 15 shows that SVM classifier gives the best accuracy and is stable than the other classifiers, the second good classifier is DT with a good accuracy but not stable followed by RF and far behind by KNN.

V. CONCLUSION

Fabric fault is becoming increasingly important through time because of the fast development of industries. This is to help company to have good quality and profitable economy. In spite of the great interest this field has gained, only a little focus was devoted for automated fabric fault. In this paper, an automated fabric inspection method based on image processing and SVM classifier is proposed to detect and classify fabric fault. The effectiveness of the proposed method is validated by recognizing 3 fault types of fabric. Compared with other existing classification methods, the obtained simulation results using SVM classifier indicate that the proposed method provides an alternative way for automated inspection of fabric fault.

Under the premise of the same input, the SVM classifier is always higher than that of DT, and the classification effect is better. By comparing with RF and KNN, the proposed method has higher classification accuracy and can be better used fabric fault inspection.

VI. REFERENCES

[1] Wood, E. J., 1990. Applying Fourier and Associated Transforms to Pattern Characterization in Textiles. Textile Research Journal, 60 : 212-220.

[2] Campbell, J. G., Hashim, A. A., Murtagh, F. D., 1997. Flaw Detection in Woven Textiles using Space-dependent Fourier Transform. Ulster, UK. INFM-97-004, 16 May 1997, University of Ulster. UK. s. 1-8.I. S.

[3] Tsai, D. M., Hsieh, C. Y., 1999. Automated Surface Inspection for Directional Textures. Image and Vision Computing, 18 : 49-62.

[4] Sari-Sarraf, H., Goddard, J. S., 1999. Vision System for On-loom Fabric Inspection. IEEE Transactions on Industry Applications, 35(6) :1252-1259.

[5] Kumar, A., Pang, G., 2002. Defect Detection in Textured Materials Using Gabor Filters. IEEE Transactions Industry Applications, 38 : 425¯440.

[6] A. Heker, K. A. Peker, A. G. Yaksek and E. DelibaQ "Fabric defect detection using deep learning," 2016 24th Signal Processing and Communication Application Conference (SIU), Zonguldak, Turkey, 2016, pp. 1437-1440, doi: 10.1109/SIU.2016.7496020.

[7] Q. Liu, C. Wang, Y. Li, M. Gao and J. Li, "A Fabric Defect Detection Method Based on Deep Learning," in IEEE Access, vol. 10, pp. 4284-4296, 2022, doi: 10.1109/ACCESS.2021.3140118.

[8] P. Sengottuvelan, A. Wahi and A. Shanmugam, "Performance of Decamouflaging Through Exploratory Image Analysis," 2008 First International Conference on Emerging Trends in Engineering and Technology, Nagpur, India, 2008, pp. 6-10, doi: 0.1109/ICETET.2008.232.

[9] Tiwari, Vikrant, and Gaurav Sharma. "Automatic fabric fault detection using morphological operations on bit plane." Int. J. Comput. Sci. Netw. Secur.(IJCSNS) 15.10 (2015): 30.

[10] S. Haykin, Neural Networks: A Comprehensive Foundation, Upper Saddle River, NJ, USA:Prentice-Hall, 1994.

[11] W. Aoudi, "Support vector machines: A distance-based approach to multi-class classification", Proc. IEEE Int. Multidisciplin. Conf. Eng. Technol. (IMCET), pp. 1-6, Sep. 2016.

[12] W. Z and X. Xue, "Multi-class support vector machines" in Support Vector Machines Applications, Springer, pp. 23-48, 2014.

[13] G. Zheng and R. Craven, "Multiclass support vector machines for power system disturbances classification based on wide-area frequency measurements", Proc. Conf. IEEE SOUTHEASTCON, pp. 68-72, Oct. 2011.

Sound traffic light and smart glasses for the Visually impaired people

Ms.Benahmed Ines Nadia
IBN KHALDOUN University
Department of Electrical Engineering
Tiaret,ALGERIA
inesbenahmed84@gamil.com

Ms.khelil Ines
IBN KHALDOUN University
Department of Electrical Engineering
Tiaret,ALGERIA
Khelilines2003@gamil.com

Mr.Belfedhal Seifeddine
IBN KHALDOUN University
Department of Electrical Engineering
Tiaret,ALGERIA
seifeddine.belfedhal@univ-tiaret.dz

Mr. Drias Souhil
IBN KHALDOUN University
Department of Electrical Engineering
Tiaret,ALGERIA
souhil.drias@univ-tiaret.dz

The aim of this project is to focus on a category that has a significant presence in society: the blind. Where it has been forgotten that this category is full of life, and hopes. The visually impaired remain marginalized in society, which fails to recognize their skills and enable them to succeed in various fields, particularly in their studies. Our study focuses on the realization of a sound traffic light and smart glasses using a microcontroller. We were able to design our pedagogical circuit according to the performance requirements. This study enabled us to design a system suitable for the visually impaired, and to choose the appropriate components. We wrote the code in the Arduino Uno program, then simulated it in the program (ISIS), and finally carried out simulations on a real model.

Keywords—smart glasses; sound traffic light; microcontroller; visually impaired; Arduino uno; ISIS;

I. INTRODUCTION

Blindness is a highly widespread impairment among people all over the world. Worldwide, 285 million people are visually impaired, 39 million are blind, and 246 have low vision, according to the World Health Organization (WHO). Around 90% of people who are blind or visually impaired reside in developing nations. [1] Vision impairment or blindness may be caused by several reasons, such as, cataract (94 million), unaddressed refractive error (88.4 million), glaucoma (7.7 million), corneal opacities (4.2 million), diabetic retinopathy (3.9 million), trachoma (2 million), and others. [2]

Blindness is a curse for those who are in need. They require assistance with all daily tasks and walking outside. In order to make them independent and free from the curse of blindness, a method is thus described in the paper.

Several solutions exist to such problems, but in our project, we focused on traffic accidents and loss rates for the blind, we found a solution to these issues by creating an audio traffic light and smart glasses that help them to cross the road safely and adapt to their surroundings easily.

This paper presents a method designed to assist visually impaired individuals through the use of smart glasses equipped with ultrasonic sensors. The proposed system aims to enhance environmental accessibility and independence for blind individuals. The technology can detect objects within a specific range, and provide guidance and alerts to help users navigate their surroundings safely. The ultrasonic smart glasses are designed to sense objects within a distance of one meter. If an object is detected within this range, the system alerts the user and guides them to avoid obstacles.

In addition, it is equipped with a GPS system that minimizes lost-and-found accidents.

The smart glasses have an infrared sensor that is fully connected to the audio traffic light. As soon as the wearer reaches the crossing area, he receives a signal from the glasses to wait until he hears the traffic signal saying, “The color is green, so you can cross.” The sound traffic light will help blind people to cross safely and boost their self-confidence.

This method has been tested with visually impaired volunteers, demonstrating its effectiveness in helping users identify and verify items independently.

II. LITERATURE SURVEY

The prevalence of blindness and visual impairment is a major global concern affecting millions of individuals and leads to significant challenges in activities of daily living. This paper presents the design of an acoustic traffic light and ultrasonic smart glasses aimed at enhancing the independence and safety of blind individuals.

The acoustic traffic light is equipped with a high-quality loudspeaker that can alert the blind to stop or cross,

When the color is red for pedestrians, the traffic light sounds “Stop, the color is red”.

When the color is green for pedestrians, the traffic light sounds “You can cross now. The color is green.”

The smart glasses are equipped with ultrasonic sensors that detect obstacles within three-meters range and provides audio feedback to the user in real time.

This system aims to improve spatial awareness and navigation for blind users, facilitating safe navigation in different environments. It involves the development process includes the integration of sensors and microcontrollers, along with the development of software for obstacle detection and sound output. The glasses have an infrared sensor

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

connected to an audio traffic signal, so that when the wearer reaches the crossing area, they receive an audio message from the glasses to wait.

Initial tests indicate that the system is effective in real-world scenarios, demonstrating its potential to reduce dependence on assistance and improve the quality of life for the visually impaired.

People with low vision face many challenges in mobility and daily tasks, often requiring constant assistance. This study explores the development of low-cost ultrasonic smart glasses designed to provide a practical and affordable solution for the blind. The glasses incorporate an ultrasonic sensor that continuously scans the surrounding environment for obstacles and delivers audio signals to the user via a built-in speaker.

The system is controlled by an Arduino microcontroller, ensuring a cost-effective approach without compromising functionality. The design focuses on affordability, ease of use and reliability, making it accessible to the widest possible population.

User experiments revealed that the glasses effectively enhance spatial awareness and independence, proving to be a viable tool to improve mobility and reduce the need for external support for visually impaired individuals. In case of vision loss, the GPS system will be activated, making it easier for the blind person's relatives to easily reach them via their GPS app, and this app will be downloaded to the smartphone of one of the blind person's relatives, thus connecting to the glasses.

Navigational challenges significantly affect the independence of visually impaired individuals, often requiring assistance for safe and efficient movement. This paper presents a model for the development of traffic lights by making them acoustically oriented for blind people to help them navigate safely and sensitize society to the need to pay attention to this group, this paper also presents advanced smart glasses using ultrasonic sensors and a microcontroller to help visually impaired users. The system uses sensor fusion techniques and machine learning algorithms to detect and accurately classify obstacles, while providing haptic and auditory feedback to guide the user. The glasses are designed to be lightweight and easy to use, with a focus on enhancing user safety and confidence in different environments. Extensive testing and user feedback have proven that smart glasses improve mobility and reduce dependence on assistance, representing a meaningful advancement in assistive technology for the visually impaired. [3]

III. OBJECTIVE

The main objective of our present work is to provide a reliable, solution for a blind people which would help them to move almost like any other normal pedestrian. It consists of developing a traffic signal by adding sound and producing smart glasses for the blind to help them adapt to daily life and provide them with safety and protection, in addition to the ability to locate the location of the glass's holder in relation to relatives to stay informed. Developing smart systems that allow blind people to recognize obstacles and directions by converting information into audio signals, enabling them to navigate more confidently and safely.

This system will serve as a third eye for the blind.

IV. COMPONENTS

To facilitate the adaptation of visually impaired individuals to daily life, a team decided to develop smart glasses connected to an audio traffic light that includes different components. These components were selected after thorough research and aligned with the project's objectives, we used :

- Red light module
- ISD1820 voice recording module
- Arduino UNO
- IR emitting LED
- Arduino NANO
- GRY infrared sensor
- Buzzer
- HCSR04 ultrasonic sensor
- Vibration motor
- Mini GPS GF-07 Magnetic

a) Firstly, for the traffic light audio

1. ISD1820 voice recording module

ISD1820-based voice recorder module for recording and playback of sound from 8 to 20 seconds.

Microphone and 0.5 W loudspeaker included, this module can be used on its own or connected to the digital output of an Arduino compatible board. Controls: recording, playback, loop playback.

2. Arduino UNO

The Arduino UNO board is an open-source microcontroller (ATmega328 clocked at 16 MHz), which can be used to interactively control a system using a program defined and stored in its memory. This board is based on a simple input/output interface, enabling it to be connected to a computer or to be powered. It can be programmed using Arduino software. It offers 14 digital input/output pins , 6 of which can generate PWM (Pulse Wirth Modulation).

3. IR emitting LED

An infrared (IR) emitting LED is a light-emitting diode designed to emit light in the infrared spectrum.

b) Secondly, the smart glasses

1. GRY infrared sensor

Infrared radiation is part of the electromagnetic spectrum and includes visible light, radio waves, ultraviolet rays and gamma rays. Sensors, called IR sensors, detect this radiation. An infrared transceiver emits a beam of infrared waves which, when it strikes an object, is reflected and picked up by a receiver. The receiver converts the reflected waves into an electrical signal, which is analyzed to determine information such as the presence or distance of the object.

2. HCSR04 ultrasonic sensor [3]

Ultrasonic sensors utilize ultrasonic waves to measure distance. They emit ultrasonic waves that
bounce back to them after hitting an object, allowing the ultrasonic sensor to determine the distance to
the object. These sensors can sense distances from 2 to 450 cm.

3. Arduino NANO

is an open-source microcontroller board based on the Microchip ATmega328P microcontroller (MCU). It offers the same connectivity and specifications as the Arduino Uno board in a smaller
format. The Arduino Nano is equipped with 30 male I/O connectors, and can be programmed using Arduino software.

The board can be powered by a mini-USB type B cable or by a 9 V battery.

4. Mini GPS GF-07 Magnetic

The GF-07 GPS module is a compact and efficient tracking device that uses GPS technology to locate vehicles, objects or people.

V. PROPOSED MODEL

Blind as a special group in society, the needs of society to give them more care and attention, so that they are better able to live independently. However, how safe walking blind life is the biggest problem. Traditional navigation device mostly blind cane, blind by tapping the ground or walking around the object to determine the direction, the structure is simple, single function, easy to use, but the secondary effect is not very obvious, in fact, will encounter many problems when using the blind such as poor road conditions, uneven, hanging in front of obstacles, ordinary cane cannot be proven accurate, such a serious impact on the safety of blind travelers.

A smart ultrasonic glasses for blind people comprises of a pair of wearable glasses, ultrasonic sensors for detection of obstacles in the way of blind man, a buzzer to give the sound as per the direction of the obstacle from the man, a central processing unit comprising of Arduino NANO which takes the information from the sensor about the obstacle distance and processes the information according to the coding done and sends the output through the buzzer, power supply is given to the central unit which distributes the power to different components. The sensor is mounted in between of the top bar and bridge present in optical glasses as shown in the figure. All the components are connected to the central unit using single strand copper wires and the power is given to the central unit using a USB cable.

The best sensors that can be used will be ultrasonic sensors because ultrasound is a strong point, the energy consumption of slow wave propagating in the medium relatively far distance. Therefore, often it is used to measure the distance over big length. At the same time, ultrasound for the object in the dark, dust, smoke, electromagnetic interference, toxic and other harsh environments have a certain ability to adapt, with a wide range of applications.

According to claim 1, as the blind man goes closer to the obstacle the distance sent by the sensors to the central unit will decrease. Hence the beeping of the buzzer will take shorter intervals and hence the beeping will be faster. But as the man will go far away the beeping will take long intervals and hence decrease.

At present, many of the navigation device using seeing-eye guide dogs, guide dogs by seeing some extent, although the trip to ensure the safety of the blind. But there are still some problems, training a guide dog larger difficulty, generally have to spend 3-6 months, training a skilled guide dogs will need to spend about two years, with dog the daily life of consumer spending, the cost it takes to reach the million, while the limited life cycle of guide dogs.

The glasses will carry a small GPS device placed to the left of them, this device is connected to an application in the phone of a blind relative in case of loss, the glasses holder will be found without any effort or cost and as soon as possible. The glasses also contain an infrared sensor placed on the front of the glasses, this part is the first transmitter, while the other part is placed on the traffic light, as soon as the holder of the glasses arrives at the crossing area, infrared rays are emitted from the glasses to the sensor part in the traffic light, when it receives it, it also emits a ray, so the blind person receives an audio warning to stop, that is, he has reached the crossing area. This process happens in a very short period of time.

According to claim 5, the ultrasonic glasses according to specifications mentioned in claim 1 are very cheap resulting in a very cheap device affordable by all. These smart glasses are very easy to use and very simple to understand. If a blind uses it for 2-3 times, then he/she will understand the working and can handle it easily.

Now we will talk about the fixed part of our project, which is the audio traffic light, we made a small model of the traffic light, we added a high-quality speaker to it and inserted an audio recording into its memory ISD1820 (voice recording module), the latter is connected to the colors of the pedestrian traffic light. When the color is red, the recording plays and the blind person receives the voice message “Wait, the color is red”.

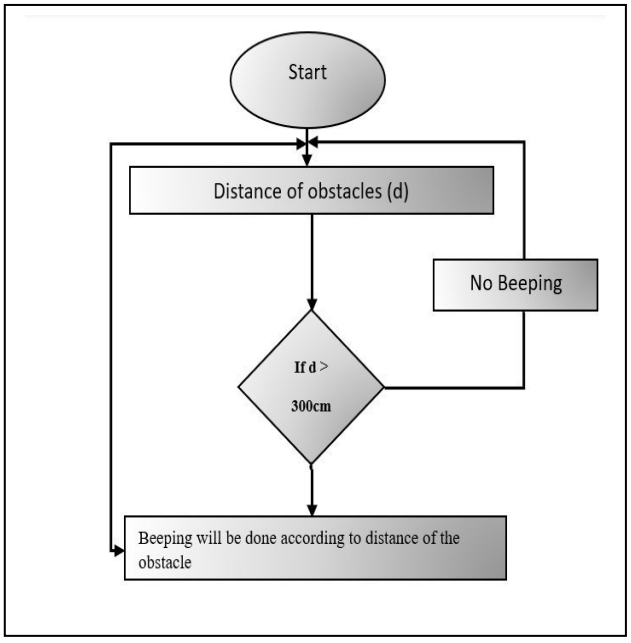
When the color is green, the recording plays “You can cross, the color is green.”

This way they can cross independently and safely, in addition to the glasses part that we talked about earlier. This step will be very important for the blind as it will further sensitize the community to the need to take care of this group in various fields.

VI. PROPOSED PROTOCOL

In this protocol sensor find out the object from distance, if it found within 3 meters then it gives sound and aware the user. Also, if it more nearer it give more sound effect.

a. *Field of Invention* [5]



There has always been a need for the person with disability to live a normal life and get opportunity to excel in the world. There have been many inventions so far to bring such peoples on the same grounds like others. Some have failed but some

have made it well. The science and technology in today's world has always tried to serve the mankind in the health and safety field and same is the motto of this project.

b. *Description of the invention*

Our project offers a service and a product As for the service, we would like to provide a service that consists of an audio traffic signal for the blind and visually impaired, providing them with the ultimate protection for crossing the road.

As for the product, we would like to provide smart glasses that provide adequate protection for the blind by providing features that the blind or visually impaired lack, which minimizes the accidents and damages that he is exposed to daily.

The glasses also add security for the parents of the blind by sending their location.

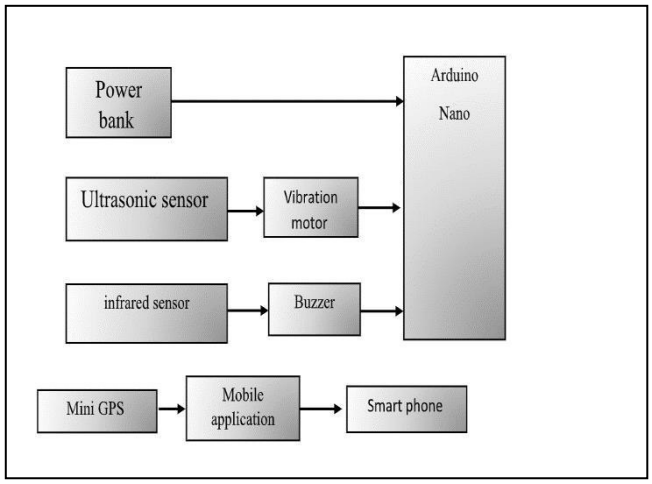
This device is very easy to use and very simple to understand this makes it user-friendly. Also, it is very cheap thus affordable.

In other words, the project focuses on developing innovative solutions to improve the daily lives of blind individuals by integrating advanced technologies. Key initiatives include creating smart glasses equipped with sensors to help users recognize obstacles and navigate safely through audio signals. The glasses will also feature GPS tracking to allow relatives to locate the user and an interactive audio map to provide directions and information about nearby locations, such as restaurants. Additionally, the project aims to design an audio-based traffic signal system offering clear voice directions to enhance safety. Emphasis is placed on battery efficiency to ensure prolonged usage and reliability. Collaboration with organizations, government agencies, and technical developers is a cornerstone of the project to deliver integrated and sustainable solutions.

c. *Diagram and Description*

The Obstacle detection module and the output device is connected to the processing unit. The power supply is used to supply power to the central processing unit. The obstacle detection module basically consists of an ultrasonic sensor, processing unit consist of a control module and the output unit consists of a buzzer. The control unit controls the ultrasonic sensors and get the information of the obstacle present in front of the man and processes the information and sends the output through the buzzer accordingly.

1. *Block Diagram of the Smart Glass*



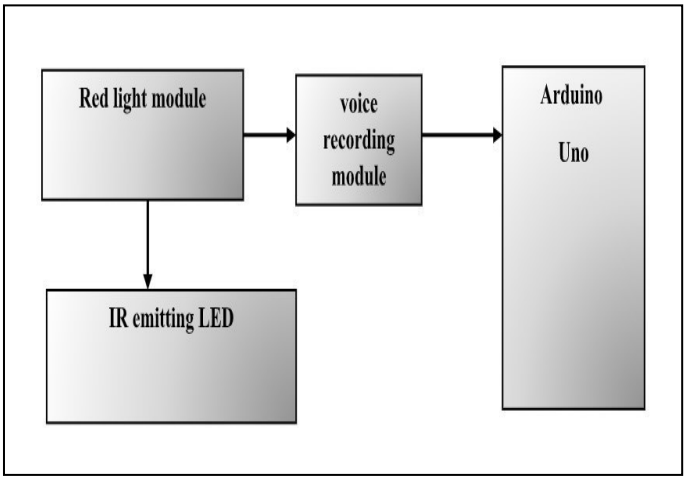
Description

The diagram shows an integrated system using an Arduino Nano as the main controller, powered by a power bank. Several sensors and actuators are connected to the Arduino to create a working device.

An ultrasonic sensor measures the distance to obstacles and triggers a vibration motor to provide haptic feedback when an obstacle is approached. An infrared sensor detects blind people near a transmitter connected to a traffic light and triggers a buzzer. In addition, a small GPS module provides

real-time location data that is sent to the mobile phone application for tracking.

1. *Block diagram of the audio traffic light system*



Description

The diagram illustrates an electronic system consisting of a number of components linked together for integrated operation. The system uses an Arduino Uno microcontroller as the central control unit. The red-light module, connected to an infrared (IR) emitting LED and a voice recording module for recording voice messages, and also connected to the Arduino Uno, lights up when the light is red, allowing the blind person to cross the street in complete safety. The IR LED

used for person detection is synchronized with the red-light module.

VII. RESULT ANALYSIS

A. This is a simulation of a traffic light with an ISD1820 voice recording module in ISIS, combining hardware and software to manage traffic lights and emit voice messages.

We connected the LEDs to the Arduino's digital pins, and used resistors to limit the current.

And the ISD1820 component is connected to the Arduino visa the (REC) pins and powered by the Arduino's 5V and GND pins.

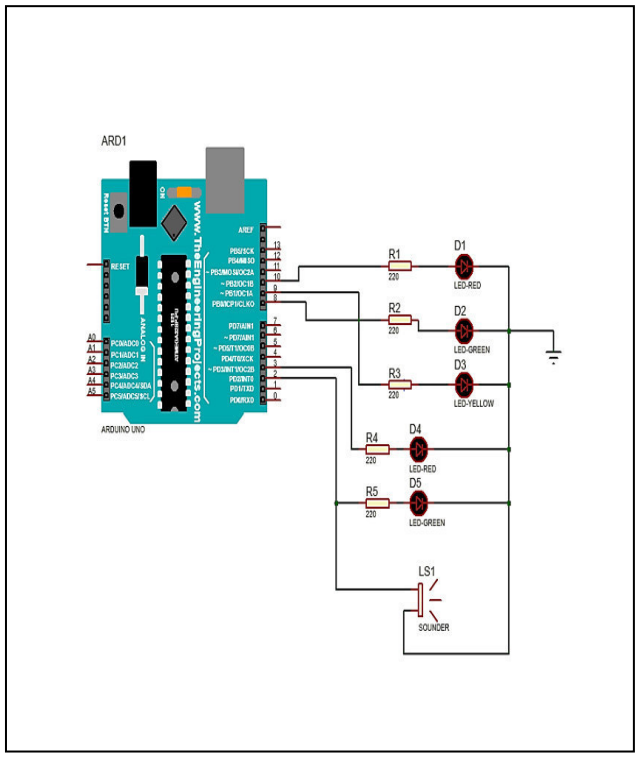


Figure.2 Simulation of a traffic light with a voice message.

B. This protocol is a microcosm of an audio traffic signal supported by a loudspeaker and an audio recording, When the color is red, the recording plays and the blind person receives the voice message “Wait, the color is red”.

When the color is green, the recording plays “You can cross, the color is green.”

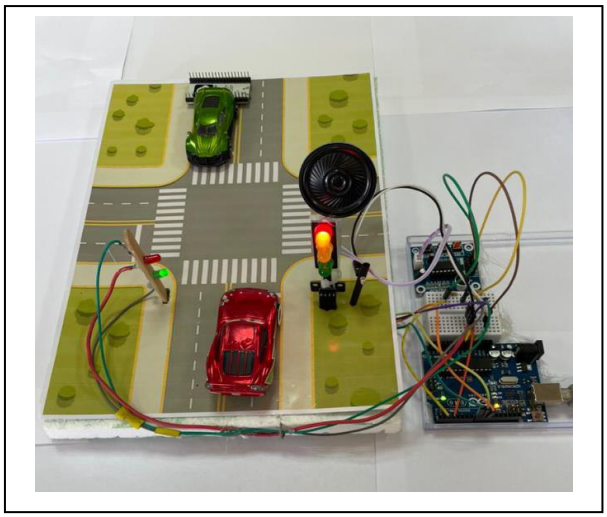


Figure3. Image shows the rrototype of sound traffic light

C. In this protocol when find object but distance greater than 3 meters then it not senses, if distance less than 300 cm then it senses and create sound. When the distance between object and user are closer then sound effect is high gradually.

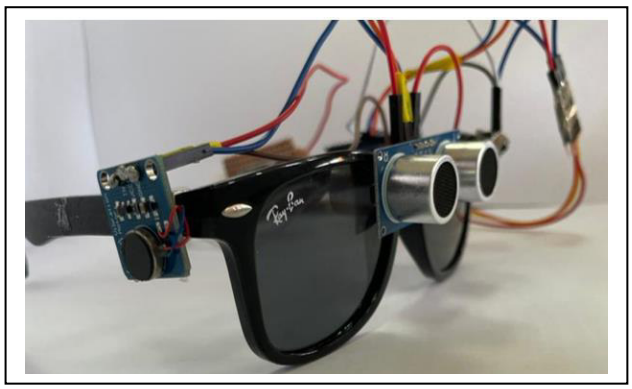


Figure.3 Sensor Implementation

D. This invention holds a GPS device that has its own app downloaded on a smartphone



Figure.4 GPS app

E. This part is the link between the glasses and the voice traffic signal. Both consist of an infrared sensor that activates when the wearer reaches the crossing area and automatically receives a voice message from the glasses telling them to stop.

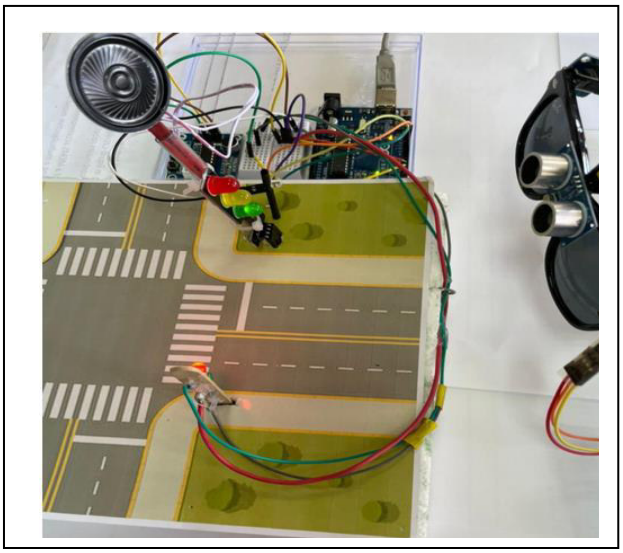


Figure 5. Realization of the global solution for our project

VIII. CONCLUSION

In conclusion, this project represents a major step in the field of technological assistance for blind and partially-sighted people. Through the development of the traffic light audio and smart glasses, it meets an essential need improving the autonomy and safety of the visually impaired. The audible light enables the visually impaired to cross the road in complete safety, by transforming light signals into audible signals, while the smart glasses offer navigation assistance by detecting obstacles and improving perception of the environment. The use of Arduino Uno and Nano boards has enabled the design of an accessible, modular prototype, paving the way for affordable, scalable solutions. This project shows the importance of innovation in creating solutions that are not only physical but also. intelligent, responding to the specific needs of users, with the integration of a mobile application and GPS, making the system even more adapted and effective. In short, this work proves that simple, accessible technologies can transform the daily lives of visually impaired people, offering them greater independence and safety.

References

[1] V. Lingawar, M. Nilakhe, M. Kamble and P. M. Shinde, "ULTRASONIC SMART GLASSES FOR VISUALLY IMPAIRED PEOPLES," *International Research Journal of Modernization in Engineering Technology and Science,* vol. 05, no. 03, 2023.

[2] M. Mukhiddinov and J. Cho, "Smart Glass System Using Deep Learning for the Blind and Visually Impaired," *Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 13120, Korea.*

[3] T. Priyan, N. Kumar, Sudharsanan, T. Selvan, D. EKamalanaban and Mrs.Saraswathi, "Ultra Sonic Glasses for the Blind," *International Jouranl of Advances in Engineering and Management (IJAEM),* vol. 5, no. 6, june 2023.

[4] J. Naskar, P. Ghosh, S. Ghosh, S. Nag and S. S. K. Sinha, "Electronic Obstacle Detector Spectacles for the Visually-Impaireds without Microcontroller,"*International Journal of Engineering Research and Applications,* vol. 14, no. 07, July 2024.

[5] H.N. Saha, S. Dey and R. Dey, " Low Cost Ultrasonic Smart Glasses for Blind," October 2017.

Dynamic Viewpoint Alignment and Sequence-Specific Optimization in MV-HEVC

1st Amel Belbel
Department of Electronics
University Badji Mokhtar of Annaba
Annaba, Algeria
amel.belbel@univ-annaba.org

2nd Amara Bekhouch
Department of Computer Science
University of Souk Ahras
Souk Ahras, Algeria
a.bekhouche@univ-soukahras.dz

3rd Noureddine Doghmane
Department of Electronics
University Badji Mokhtar of Annaba
Annaba, Algeria
ndoghmane@univ-annaba.org

***Abstract*—This paper presents a novel framework for efficient multi-view video encoding based on MV-HEVC, utilizing Dynamic Viewpoint Adaptation (DVA) with two strategies: Adaptive Alignment Estimation (AAE) and Hierarchical Encoding Adjustment (HEA). AAE refines alignment parameters to improve inter-view prediction and reduce redundant computations, while HEA optimizes resource allocation by prioritizing key views. Implemented on HTM 16.2, experimental results with six HD-MVV sequences show a reduction in encoding time by up to 29.26% with minimal impact on rate-distortion performance. The proposed framework offers substantial computational savings, making it suitable for real-time applications in resource-constrained environments.**

***Index Terms*—MV-HEVC, Coding Complexity, Adaptive Estimation, Hierarchical Adjustment.**

I. Introduction

Immersive multimedia applications, such as virtual reality (VR), 3D video streaming, and free-viewpoint television (FVV), are revolutionizing the domain of digital video technologies. These applications depend on the seamless compression and transmission of multi-view video sequences, where content is captured from multiple viewpoints to create interactive and realistic user experiences. At the core of these advancements lies the Multi-View High Efficiency Video Coding (MV-HEVC) standard [1], an extension of HEVC, which achieves remarkable compression efficiency by leveraging both temporal and inter-view correlations.

MV-HEVC provides significant compression gains compared to conventional single-view or multi-view solutions by incorporating advanced coding structures, including disparity-compensated prediction and inter-view residual prediction as shown in Fig.1 . This makes it a foundational technology for applications such as immersive teleconferencing, 3D gaming, and UHD 3D video delivery [2]. However, the encoding process faces substantial challenges arising from the computational complexity of processing multi-view sequences.

The complexity of MV-HEVC encoding stems from variations in camera configurations, viewing angles, and scene depth across multiple views, as shown in Fig.2 . These factors introduce inconsistencies between dependent views, complicating the alignment of corresponding regions. In high-motion scenes or sequences with significant viewpoint disparities, this misalignment is further exacerbated, leading to increased

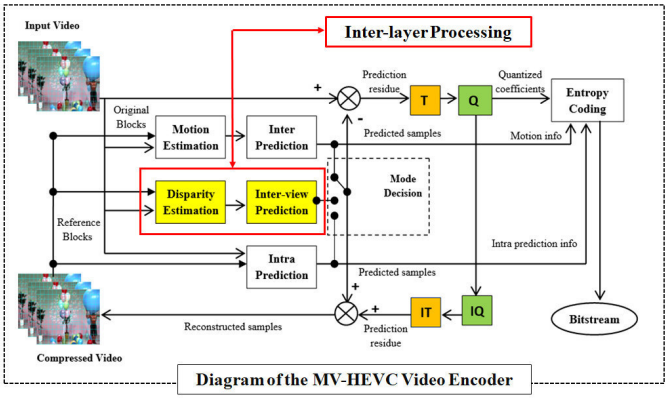


Fig. 1. MV-HEVC Encoding Concept.

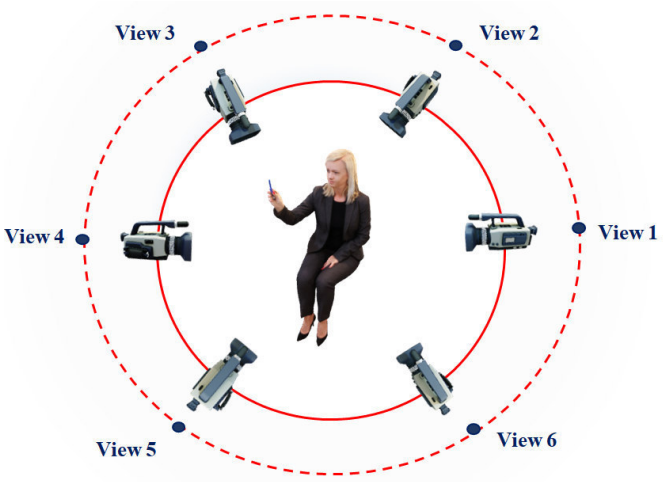


Fig. 2. Multi-view Video with 6 Views.

computational demands for disparity and motion estimation. Such inefficiencies hinder real-time implementation and limit the scalability of MV-HEVC in computationally constrained environments.

Traditional MV-HEVC encoding approaches adopt static processing frameworks that fail to adapt to the specific characteristics of individual video sequences. Consequently, these

methods often result in suboptimal performance, particularly for sequences exhibiting non-standard motion patterns, varying inter-view dependencies, or atypical scene dynamics. This limitation becomes increasingly critical as immersive applications expand to encompass diverse content types and operational conditions. Previous research has explored various methods to address these challenges. For instance, one approach focuses on reducing the computational complexity of the HEVC encoder by selecting the best motion vector correlation among different prediction modes. This method, applicable independently on different views of the MV-HEVC, results in an average 20% reduction in motion estimation encoding time. However, this technique is limited to the simulcast prediction structure of MV-HEVC. Additionally, G. Jiang et al. proposed a perceptual distortion threshold model (PDTM), which aims to reduce the complexity of the dependent view by terminating unnecessary mode selection through a threshold comparison of the sum of squared errors (SSE). This method achieved a 52.9% gain, but it is focused solely on the dependent view [3], [4].

In this paper, the proposed framework integrates two key strategies: Adaptive Alignment Estimation (AAE) and Hierarchical Encoding Adjustment (HEA). AAE dynamically refines alignment parameters based on motion characteristics and viewpoint variations to reduce redundant computations and enhance inter-view consistency. HEA prioritizes computational resources for key views while minimizing redundancy in dependent views, ensuring an optimal balance between encoding efficiency and computational complexity. Experimental results show that the proposed approach significantly reduces encoding time, up to 29.26%, without compromising rate-distortion performance, making it ideal for real-time applications and systems with stringent computational constraints.

The rest of this paper is organized as follows: Section 2 outlines the technical details of the proposed DVA framework, while Section 3 describes the experimental setup and results. Section 4 provides an in-depth discussion of the findings, and Section 5 concludes the paper with potential directions for future research.

II. Enhancing Multi-View Coding: Challenges and Opportunities in Inter-View Prediction

The demand for efficient video compression has grown with the advent of UHD and 3D video applications, requiring encoding frameworks capable of addressing diverse spatial and temporal redundancies. High-Efficiency Video Coding (HEVC) has introduced innovative structural features, including larger block sizes and a flexible quad-tree-based partitioning architecture. These features enable efficient coding of video frames through hierarchical divisions into Coding Units (CUs), Prediction Units (PUs), and Transform Units (TUs). However, HEVC's traditional design does not directly account for redundancies present in multi-view video sequences, particularly those arising from inter-view correlations.

The MV-HEVC extension addresses this gap by introducing inter-view prediction mechanisms alongside standard inter-frame and intra-frame prediction. This layered structure enables the efficient encoding of dependent views by exploiting inter-view similarities. The key innovation lies in Disparity Compensated Prediction (DCP), which utilizes disparities between adjacent views to infer motion information and reduce redundant encoding operations. Unlike Motion Compensated Prediction (MCP), which operates within a single view, DCP incorporates cross-view dependencies to enhance compression performance. Despite these advancements, significant challenges remain in optimizing MV-HEVC for varying scene complexities and view configurations.

To assess the efficiency of MV-HEVC's inter-view prediction mechanism, an experimental analysis was conducted employing a diverse set of multi-view video sequences [5]. These sequences, detailed in Table I, were selected to represent a broad range of real-world encoding scenarios, varying in resolution, frame rate, and camera spacing. The evaluation focused exclusively on the IP structure (two views). Coding efficiency was analyzed to understand the influence of this configuration on performance. The findings, as summarized in Table II, provide the following key insights:

1) **Strong Redundancies in Adjacent Views:** Inter-view prediction demonstrates high efficiency for adjacent views, with Coding Unit (CU) utilization rates exceeding 40% for many sequences, such as Dog and Kendo. This highlights the substantial redundancies present between neighboring views, which can be effectively exploited to minimize encoding complexity.
2) **Performance Degradation for Distant Views:** As the distance between views increases, inter-view CU rates decline, reflecting the reduced correlation between non-adjacent views. For instance, CU rates for distant views in the Sharek sequence drop by nearly 30% compared to adjacent views. This underscores the limitations of traditional DCP approaches in handling larger disparities.
3) **Resolution and Content Dependencies:** High-resolution sequences, such as Sharek (1920×1088), exhibit significantly higher inter-view prediction rates compared to lower-resolution sequences. Additionally, content-specific characteristics, including motion dynamics and texture complexity, influence the efficiency of inter-view prediction.

The diminishing CU rates with increasing view distances expose critical challenges in disparity estimation and motion prediction. Disparities caused by varying camera angles introduce offsets between views, as illustrated in Fig. 3.

These offsets complicate block matching processes, particularly in scenes with high motion or complex textures. Existing disparity estimation techniques struggle to maintain precision in such scenarios, leading to inefficiencies in CU utilization and increased computational overhead.

Based on these challenges, the next section introduces the proposed algorithm, which aims to address these limitations by enhancing inter-view prediction accuracy and optimizing computational efficiency.

TABLE I
HD MULTI-VIEW VIDEO SEQUENCES EMPLOYED IN EXPERIMENTAL ASSESSMENT

Sequences	Resolution	Frames	Frame Rate (fps)	Cameras spacing (cm)
Champagne-tower	1280x960	300	30	Stereo
Panatomim	1280x960	300	30	Stereo
Dog	1280x960	300	30	Stereo
Balloons	1024x768	300	30	5
Kendo	1024x768	300	30	5
Sharek	1920x1088	300	30	Stereo

TABLE II
PROPORTION OF PREDICTED UNITS BASED ON SHIFTED SEGMENTS

	Dependent Views					
Sequences	View 1	View 2	View 3	View 4	View 5	View 6
Champagne_tower	0.60	0.72	1.32	2.17	2.89	4.08
Pantomim	0.79	1.17	1.58	1.70	2.80	3.69
Dog	1.18	2.79	3.66	4.47	4.66	5.18
Balloons	2.75	3.22	3.31	4.88	6.43	9.92
Kendo	1.08	1.77	4.47	5.01	6.09	7.2
Sharek	0.03	018	0.40	0.68	0.84	0.95

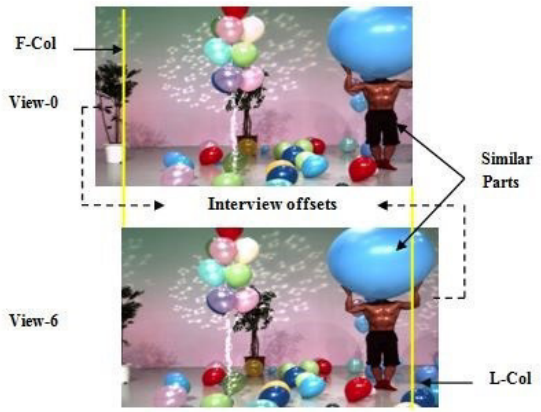


Fig. 3. Inter-view offsets between two views.

III. OPTIMIZED VECTOR ADJUSTMENT FOR IMPROVED INTER-VIEW CODING EFFICIENCY IN MV-HEVC

This paper introduces a novel Dynamic Viewpoint Alignment (DVA) framework that enhances the efficiency of MV-HEVC encoding by integrating sequence-specific optimizations. The proposed framework incorporates two complementary strategies to dynamically adjust encoding processes while reducing computational overhead:

1) **Adaptive Alignment Estimation (AAE):** A dynamic adjustment mechanism that tailors alignment parameters to the motion characteristics and viewpoint variations within each sequence. By refining the alignment process based on content-specific features, AAE effectively reduces redundant computations and enhances inter-view consistency. This approach integrates an initial analysis of inter-view similarities, focusing on shifted regions to eliminate offsets caused by camera angle variations. By prioritizing alignment for these relevant regions, redundant processing is minimized, and disparity estimation becomes more efficient. The elimination of these offsets not only simplifies the block matching process but also ensures that disparity estimation aligns accurately with the base view. The content-aware adjustments ensure that each view's unique characteristics, such as motion intensity and view spacing, are taken into account, leading to more accurate predictions and better utilization of computational resources. Fig. 4 illustrates this step, highlighting the elimination of offsets and the alignment process between views.

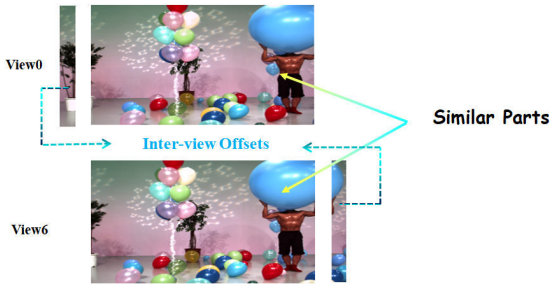


Fig. 4. Eliminating Inter-view Offsets.

2) **Hierarchical Encoding Adjustment (HEA):** A multi-layered optimization strategy that prioritizes computational resources for key views while minimizing redundancy in dependent views. The HEA approach ensures efficient resource allocation, carefully balancing compression efficiency and computational complexity. By focusing resources on the most critical views—typically the central or most significant ones—this strategy reduces unnecessary processing for less influential views, leading to a more efficient use of computational power. This hierarchical structuring aligns well with multi-view

video sequences, where certain views are more crucial for maintaining visual quality and user experience , as shown in Fig.5.

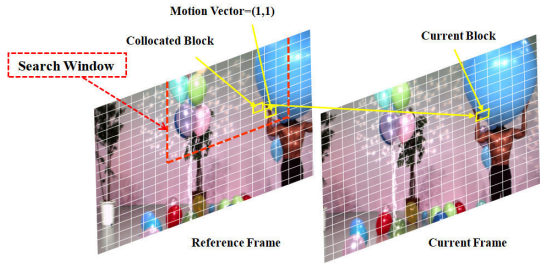


Fig. 5. Decreasing the Search Window Size.

3) **Earlier Coding Unit (CU) Splitting and Adaptive Search Window (ACUS)**: A critical enhancement to motion estimation and disparity reduction, this approach combines an earlier decision for CU splitting with an adaptive search window. By predicting and adjusting the CU splitting earlier in the encoding process, the framework significantly reduces the computational cost typically incurred during motion estimation, especially for dependent views, as shown Fig. 6. Additionally, the adaptive search window limits the area of motion search based on the inter-view similarities, narrowing down the focus to the most relevant regions. This method ensures a reduction in search time while maintaining accuracy in motion and disparity estimation, resulting in faster processing and reduced overall complexity.

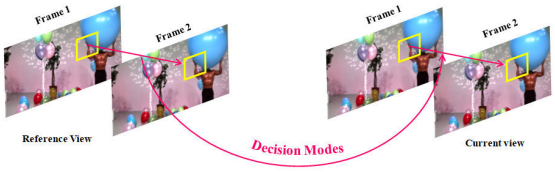


Fig. 6. Reducing the computation time of ADE.

The DVA framework is designed to operate effectively across a wide range of multi-view video sequences, addressing the unique demands of various immersive applications. By integrating content-aware adaptations into the encoding pipeline, the proposed approach achieves substantial reductions in computational complexity while preserving compression quality. This makes it particularly suitable for real-time applications and systems with stringent computational constraints. Experimental validation demonstrates the robustness of the DVA framework across diverse multi-view video datasets, showcasing significant reductions in encoding time while maintaining competitive rate-distortion performance. The proposed strategies pave the way for scalable, efficient, and adaptive multi-view video compression systems, aligning with the growing demands of next-generation immersive multimedia applications.

IV. Experimental results

The proposed Dynamic Viewpoint Adaptation (DVA) framework was implemented on the MV-HEVC reference software, HTM version 16.2. The experimental evaluation utilized six high-definition multi-view video (HD-MVV) sequences—Champagne-lower, Panatomim, Dog, Balloons, Kendos, and Sharek—as outlined in Table I. These sequences, captured using a linear camera arrangement with seven views per sequence, provided a diverse dataset for testing. Encoding was performed using quantization parameter (QP) values ranging from 25 to 40, enabling a comprehensive analysis of the framework's performance across varying levels of compression.

All simulations were conducted under the standard conditions defined by the JCT-3V common test configuration (CTC) [6], ensuring consistency and reproducibility. The parameters used for encoding are summarized in Table III.

TABLE III
Encoder configurations

Setting	Value
Prediction Structure	IP and PIP
GOP length	8
Intra period	24
Coding tree block	64x64 pixels
CU depth level	4
Search range	Variable (4,8,12 and 16)
Entropy coder	CABAC

The evaluation focused on three key performance metrics: percentage change in encoding time elta time $\Delta time$ (%), percentage change in bitrate $\Delta Bitrate$ (%), and change in peak signal-to-noise ratio $\Delta PSNR$ (dB), calculated using Eq.1. Here, the results of the proposed framework (DVA) are compared against the baseline performance of HTM.

$$gain = \frac{Proposed - HTM}{HTM} \quad (1)$$

Table IV provides a detailed comparison of the results achieved by the DVA framework, incorporating AAE and HEA, against the HTM reference software for two selected views, V-0 and V-1. The findings demonstrate that the proposed method achieves a significant reduction in total encoding time, ranging between 20.37% and 29.26% for a QP value of 25 across the six tested multi-view video sequences. As the QP value increases, the time savings become more pronounced due to changes in the depth levels of Coding Units (CUs) and the corresponding adjustment of the search window size. However, higher QP values result in larger CUs, which slightly affects encoding quality. The size of the search window also decreases as QP increases, further contributing to computational efficiency but potentially impacting the precision of motion and disparity estimation.

The variations in time across different sequences, as shown in Table IV, can be attributed to unique sequence characteristics and the spacing of the camera setup. For instance, sequences such as "Kendos" and "Balloons," which have

TABLE IV
EXPERIMENTAL RESULTS IN COMPARISON WITH THE HTM ENCODER USING THE IP PREDICTION SCHEME

Sequences	QP	$\Delta Time$		$\Delta PSNR$
Champagne_tower	25	-26.87	1.39	-0.05
	30	-30.82	1.17	-0.03
	35	-35.05	1.02	0.02
	40	-38.11	0.89	0.02
Pantomime	25	-29.26	1.79	-0.09
	30	-33.06	1.53	0.00
	35	-37.17	1.36	-0.03
	40	-40.07	1.23	0.07
Dog	25	-24.53	1.27	-0.07
	30	-28.26	1.09	0.02
	35	-32.25	0.98	0.01
	40	-35.12	0.88	0.04
Balloons	25	-20.37	0.89	-0.05
	30	-24.49	0.73	-0.03
	35	-28.91	0.63	-0.02
	40	-31.93	0.55	0.02
Kendo	25	-23.03	1.02	0.01
	30	-27.47	0.82	0.07
	35	-31.56	0.69	-0.01
	40	-34.47	0.59	0.02
Sharek	25	-26.54	1.22	-0.04
	30	-30.45	0.98	-0.03
	35	-34.68	0.80	-0.05
	40	-37.74	0.66	0.00
Average	25	-23.31	1.04	-0.02
	30	-27.47	0.84	0.00
	35	-32.73	0.70	-0.02
	40	-34.71	0.60	0.01

relatively smaller distances between cameras, show less time improvement compared to sequences with wider camera arrangements. Despite these differences, the proposed framework maintains a negligible impact on rate-distortion (RD) performance, as evidenced by the minimal changes in PSNR and Bitrate recorded in Table IV. These results highlight the framework's ability to deliver significant reductions in computational overhead without compromising compression efficiency, making it particularly suitable for real-time multi-view video systems.

V. CONCLUSION

In this work, we proposed an optimized MV-HEVC framework that significantly reduces computational complexity while maintaining high compression efficiency. The integration of Adaptive Alignment Estimation (AAE) and Hierarchical Encoding Adjustment (HEA) leads to a substantial reduction in encoding time, improving the feasibility of real-time applications such as virtual reality and 3D video streaming. Experimental results confirm that the proposed method offers an efficient solution for the challenges of multi-view video compression, demonstrating competitive rate-distortion performance. Future work will focus on further enhancing the motion and disparity estimation processes and exploring advanced machine learning techniques for dynamic encoding adjustments.

REFERENCES

[1] G. J. Sullivan, J. M. Boyce, Y. Chen, J. R. Ohm, C. A. Segall and A. Vetro, Standardized Extensions of High Efficiency Video Coding (HEVC), IEEE Journal of Selected Topics in Signal Processing, vol. 7, no. 6, pp. 1001- 1016, Dec. 2013.

[2] K. M¨uller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee, G. Tech, M. Winken and T. Wiegand, 3D High-Efficiency Video Coding for Multi-View Video and Depth Data, IEEE Transactions on Image Processing, vol. 22, no. 9, pp.

[3] G. Jiang, B. Du, S. Fang, M. Yu, F. Shao, Z. Peng, F. Chen, Fast inter-frame prediction in multi-view video coding based on perceptual distortion threshold model, Signal Process., Image Commun. 70 (2019) 199–209.

[4] Y. Wang, T. Jiang, S. Ma, W. Gao. Novel spatio-temporal structural information based video quality metric.IEEE Transactions on Circuits and Systems for Video Technology, 22(7) (2012) 989-998. .

[5] http://mcl.usc.edu/mcl-3d-database/

[6] K. Müller and A. Vetro, "Common test conditions of 3DV core experiments," JCT-3V document JCT3V-G1100, Jan. 2014.

Automatic Extractive Text Summarization

Sabrina Karboua
Departement of Computer Science
Mechatronics Laboratory (LMETR)
Ferhat ABBAS University Setif 1 Setif, Algeria
sabrina.karboua@univ-setif.dz

***Abstract*—In the context of crisis management, the ability to efficiently process and summarize vast amounts of information is crucial for timely decision-making. This thesis presents an innovative approach to Automatic Text Summarization (ATS) designed to assist stakeholders in navigating complex datasets. We developed a query-based multi-document summarization system that extracts the most relevant information from extensive collections of documents. The proposed methodology consists of two phases: the first phase employs a semantic machine learning algorithm combined with the BM25 OKAPI information retrieval method to generate extractive summaries for individual documents. The second phase integrates these summaries using the TextRank algorithm to produce a cohesive global summary. Our evaluation demonstrates the effectiveness of the system, achieving an average recall of 66.72%, precision of 69.69%, and F-measure of 67.68%. This work highlights the potential of ATS to enhance situational awareness and support informed decision-making in crisis scenarios.**

***Index Terms*—Automatic text summarization, information retrieval, data mining, natural language processing, semantic similarity, Text-Rank, BM25 OKAPI, WordNet**

I. Introduction

Crisis management is a critical field that encompasses the preparation for, response to, and recovery from emergencies, such as pandemics, natural disasters, and humanitarian crises. In these high-stakes scenarios, decision-makers face immense pressure to act swiftly and accurately, often relying on vast and rapidly growing volumes of information from diverse sources. These include scientific research, official reports, media coverage, and social media updates. While this abundance of data is invaluable, it also presents a challenge: how to process and extract actionable insights in a timely manner without being overwhelmed by the sheer scale of information [9].

Text summarization has emerged as a transformative solution to this challenge. At its core, text summarization refers to the automated process of condensing a large body of text into a concise summary that preserves its most important content. This technology is vital in crisis management, where access to precise and relevant information can significantly impact decision-making, resource allocation, and public safety. Summarization not only reduces the cognitive load required to process information but also ensures that critical details are not lost in the flood of data [8].

Two primary types of text summarization—extractive and abstractive—play a key role in addressing these needs. Extractive summarization selects the most relevant sentences or phrases directly from the original text, ensuring that the generated summary maintains factual consistency. Abstractive summarization, on the other hand, generates new sentences that paraphrase the key ideas of the original text, often improving the summary's coherence and readability. While abstractive methods are more flexible, extractive summarization remains more commonly used in applications requiring high accuracy and computational efficiency.

Query-based summarization extends these capabilities by focusing on specific information needs. In this approach, the system generates summaries tailored to a given query, ensuring that the output is directly relevant to the user's question. This makes query-based summarization especially valuable in crisis contexts, where stakeholders often require answers to specific, time-sensitive questions. For instance, during a natural disaster, emergency responders may need rapid updates on infrastructure damage, while public health officials during a pandemic may seek summaries of the latest research on disease transmission [8].

The integration of text summarization into crisis management workflows provides several key benefits:

Efficiency: Summarization reduces the time required to sift through large volumes of text, enabling faster decision-making. Relevance: By prioritizing key information, summarization ensures that decision-makers can focus on what matters most.

Adaptability: Advanced summarization techniques can be tailored to diverse domains, making them applicable to various types of crises.

Scalability: Automated summarization systems can handle enormous datasets, far beyond the capacity of manual review.

In this paper, we propose a two-step query-based multi-document summarization framework designed to address these challenges. The system first performs extractive single-document summarization based on a query, producing individual summaries of relevant documents. These summaries are then combined through a second step of multi-document summarization, creating a comprehensive summary that captures the most critical insights.

To evaluate the effectiveness of the framework, we apply it to a case study focusing on the COVID-19 pandemic. This global health crisis, marked by an unprecedented surge in scientific research and public health communications, exemplifies the need for robust tools to manage information

overload. By summarizing COVID-19-related literature, the study demonstrates the system's ability to provide concise, relevant insights, supporting decision-making in real-world crisis scenarios.

II. Definition of AlgorithmS Used

Okapi BM 25: The algorithm okapi BM25 is a probabilistic information retrieval ranking method that ranks documents based on a relevance score, which represents the probability of a document being relevant to the input query. Its popularity stems from its efficiency and simplicity. The BM25 score is calculated using two primary components: Term Frequency (TF) and Inverse Document Frequency (IDF). Additionally, BM25 incorporates heuristic techniques for document length normalization and for satisfying the concavity constraint of term frequency, which allows it to achieve better performance compared to traditional TF-IDF in many scenarios. However, BM25 has a notable limitation: it relies heavily on the explicit presence of query words within a document to identify the most informative sentences. This dependency restricts its ability to detect implicit relationships between terms. For example, BM25 would treat terms like "sick" and "ill" as unrelated entities unless explicitly linked in the text. This limitation can lead to a loss of coherence and introduce ambiguity in generated summaries, as it fails to capture semantic similarity or relatedness between terms.

Addressing this challenge requires incorporating external knowledge sources to enrich the system's understanding of implicit relationships between terms in a document. One approach is to apply semantic reasoning by measuring the semantic distance between text units, such as words, phrases, sentences, or paragraphs. Semantic distance quantifies how closely related two units are in terms of their meanings, allowing the system to better interpret the underlying relationships in a document.

To resolve the problem of machine understanding and improve coherence, we propose enhancing the system with knowledge repositories constructed by humans. These repositories provide a structured semantic understanding of language and allow the system to detect implicit connections between terms. Among these repositories, WordNet has been widely used to enrich the understanding of text documents for various applications. WordNet's extensive lexical database enables the system to perform semantic reasoning, bridging the gap between explicit text and implicit meanings, without requiring additional training once the necessary features are extracted.

Semantic Similarity using WordNet

WordNet is the result of a research project conducted at Princeton University, designed to model the lexical knowledge of native speakers and store it in a machine-readable dictionary [**?**]. It is an online lexical database that includes nouns, verbs, adjectives, and adverbs grouped into sets of cognitive synonyms, known as **synsets**. Each synset represents a specific meaning of a word and includes the word itself, its definition, and its synonyms. Words in WordNet are typically represented in a standardized format, with each word tagged according to its **Part-of-Speech (POS)**. WordNet uses the following four POS tags:

- **n**: nouns
- **v**: verbs
- **a**: adjectives
- **r**: adverbs

Word senses and synsets are interconnected through a variety of relationships. The relationships linking word senses are referred to as **semantic relations**, while those connecting synsets are known as **lexical relations** [**?**]. For example, nouns in WordNet exhibit the following semantic relations:

- **Hyponym/Hypernym (IS-A, HAS-A)**: These relations define hierarchical relationships where a hyponym is a more specific term, and a hypernym is a more general one.
- **Meronym/Holonym (Member-of, Has-member, Part-of, Has-part)**: These relations describe part-whole relationships, where a meronym is a part or member of something, and a holonym is the whole entity.

To measure the similarity between two words, we use **Wu and Palmer (WUP) similarity**. This measure extends similarity by incorporating the depth of the **Least Common Subsumer (LCS)**. The LCS represents the most specific concept that two concepts share as an ancestor in the WordNet hierarchy, effectively serving as the closest common ancestor of two concepts, C_1 and C_2, starting from the root node.

The WUP similarity between two concepts is calculated using the following formula:

$$\text{sim}_{wup} = \frac{2 \times \text{depth}(\text{lcs}(C_1, C_2))}{\text{depth}(C_1) + \text{depth}(C_2)}$$

In this formula:

- **depth(lcs(**C_1**,** C_2**))**: The depth of the Least Common Subsumer in the WordNet hierarchy.
- **depth(**C_1**)** and **depth(**C_2**)**: The depths of the two individual concepts.

This measure ensures that concepts sharing a deeper and more specific common ancestor have higher similarity values compared to those sharing a more general ancestor. WUP similarity is particularly useful in natural language processing (NLP) tasks, as it provides a precise measure of how semantically related two terms are within the WordNet hierarchy. Its integration into systems enhances their ability to detect implicit relationships, improving performance in tasks like semantic analysis and information retrieval.

TextRank Algorithm: TextRank is an unsupervised graph-based algorithm used for text summarization and keyword extraction. Inspired by Google's PageRank algorithm, TextRank represents the input text as a graph, where sentences or words act as nodes, and edges between nodes represent their similarity or co-occurrence. The algorithm assigns scores to the nodes based on their importance within the graph, calculated iteratively. In text summarization, the most highly scored sentences are selected to create a concise summary. Similarly, in keyword extraction, the top-ranked words are

identified as key terms, making TextRank a versatile tool for natural language processing tasks.

III. Architecture Overview

Our proposed query based multi-document summarization system works in two steps, the output of the first step serves as an input for the second step:

FIRST STEP: This is a query-based single-document summarization process, where each document in the corpus is summarized extractively and saved temporarily.

SECOND STEP: This is an extractive generic multi-document summarization process, where the set of temporary summaries generated in the first step is used as input to produce a single extractive summary.

The figure below illustrates how the two steps interact:

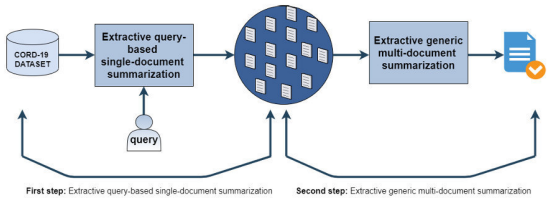


Fig. 1. The proposed system architecture

A. *FIRST STEP: Extractive Query-Based Single-Document Summarization*

This step involves query-based single-document summarization, where each document in the corpus is processed individually and independently. In this phase, the model applies an algorithm to each document, assigning a score to every sentence based on its relevance to the input query or question. The top N sentences with the highest scores are then selected to generate a summary that addresses the query.

The figure below illustrates the entire process of this phase, which will be explained in detail in the following paragraphs:

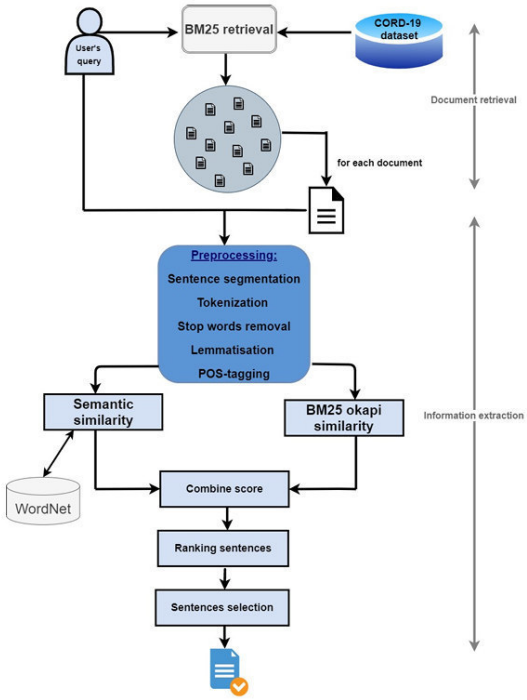


Fig. 2. First Step Architecture: Query Based Single Document Summarization

Phase 1: Document Retrieval

Document retrieval is the first phase in our proposed system, as we are working with a large corpus of over 50,000 papers. To manage this extensive dataset, it is necessary to reduce the corpus size by employing a document retrieval function.

The goal of document retrieval is to identify and select the top matching articles for a given query. For this purpose, the BM25 OKAPI index is used as a similarity measure between the query and the documents. This method ranks documents in the dataset by assigning a relevance score to each document and retrieves those with a score greater than 0. The resulting subset of documents is then used to extract the most relevant information.

a- Cleaning and Preprocessing Before further processing, the text undergoes a cleaning phase to ensure consistency and quality. The steps include:

1) Removing empty and duplicated data: Focus on cleaning the `"body_text"` column.
2) Discarding Non-English Articles: Retain only English articles for analysis.
3) Removing Citation Markers: Remove square brackets containing numbers (e.g., `[6,11]`).
4) Eliminating Punctuation and Special Characters: Discard symbols like ”?”, ”!”, ”;” and others.
5) **Lowercasing:** Convert all text to lowercase for uniformity.

This cleaning process ensures that the corpus and input query are properly preprocessed before further analysis.

Preprocessing Workflow

Preprocessing is a critical phase that converts unstructured data into structured data, minimizing noise that could interfere with the system's performance. Indexing engines and automatic summarization systems are especially sensitive to text noise.

1) Sentence Segmentation: Divides text into individual sentences.
2) Tokenization: Breaks sentences into individual words or tokens.
3) Stop-Word Removal: Removes uninformative words like *”a”, ”an”, ”the”*, and prepositions.
4) Lemmatization: Reduces words to their base or dictionary form. Example: ”causes” becomes ”cause”.
5) POS-Tagging: Assigns grammatical labels (e.g., noun, verb, adjective) to words based on context.

Input:

```
['Coronaviruses', 'family',
'enveloped', 'RNA', 'viruses',
'cause', 'diseases']
```

Output:

```
(('Coronaviruses', 'n'), ('family',
'n'), ('enveloped', 'adj'), ('RNA',
'n'), ('viruses', 'n'), ('cause',
'v'), ('diseases', 'n'))
```

B. *Phase 2: Information Extraction*

1. Sentence Scoring

Each sentence is assigned an importance score, which reflects its relevance to the query and serves as a measure of its overall usefulness. These scores can be used to rank sentences, allowing the system to prioritize and select the most important ones.

After preprocessing the documents and the input query, the sentences in the original document are scored based on two key measures: semantic similarity score and BM25 score. The semantic similarity score evaluates the contextual alignment between the sentence and the query, while the BM25 score measures term-based relevance. The final score for a sentence is calculated using these measures to determine its significance. each sentence within the document and the input query. We calculate the score of a sen tence by using the following measures

1) **BM25 Scoring:** Adapts the BM25 algorithm, typically used for document retrieval, to rank sentences based on relevance to the query.
2) **Semantic Similarity:** Measures the similarity between a sentence and the query using WordNet. The algorithm described below calculates the semantic similarity between two sentences, s_1 and s_2, by taking two English sentences as input and producing a similarity score as output. First, each sentence is tokenized into individual words, and these tokens are lemmatized to reduce them to their base forms, removing inflectional or derivational variations. Next, part-of-speech (POS) tagging is applied to assign grammatical roles, ensuring the correct synsets (specific meanings) are selected from WordNet. The synsets for words in s_1 and s_2 are then extracted, representing their semantic meanings. The algorithm iterates through all possible word pairs from the two sentences, computing the similarity between each pair using the Wu and Palmer (WUP) similarity measure, which considers the depth of the least common subsumer in the WordNet hierarchy. Finally, the individual similarity scores are aggregated to produce a single semantic similarity score that reflects the contextual alignment between the two sentences, which is then returned as the output.

The final score for a sentence is the average of its semantic similarity and BM25 scores:

Score(sentence_i, query) =

$$\frac{\text{BM25_Score} + \text{Semantic_Similarity_Score}}{2} \quad (1)$$

Algorithm : Semantic_ similarity_score

Input: sentence 1, sentence 2

Output: similarity_score

Begin

Tokenization,lemmatization and tag

1: sentence1 ← **pos_tag**(**lemma**(**word_tokenize**(sentence1)))

2: sentence2 ← **pos_tag**(**lemma**(**word_tokenize**(sentence2)))

Get the synsets for the tagged words

3:synsets1 ← [**Synset**(tagged_word) for tagged_word in sentence 1]

4: synsets2 ← [**Synset**(tagged_word) for tagged_word in sentence2]

5: score←0, count ← 0

For each word in the first sentence

6: for synset in synsets1 :

Get the similarity value of the most similar word in the other sentence

7: best_score ← max([synset.**WUP_similarity**(ss) **for** ss in synsets2])

Check that the similarity could have been computed

8: If best_score is not None:

9: similarity_score ←similarity_score + best_score

10: count ←count+ 1

12: End if

13: End for

15: return similarity_score/count

End

2. Sentence Selection

Sentences are ranked in descending order based on their scores. The top N highest-scoring sentences are selected and combined into a summary.

SECOND STEP: Extractive Generic Multi-Document Summarization

The second step is a continuation of the first step. During this phase, the set of summaries generated in the first step is used as input to produce a single, final extractive summary.

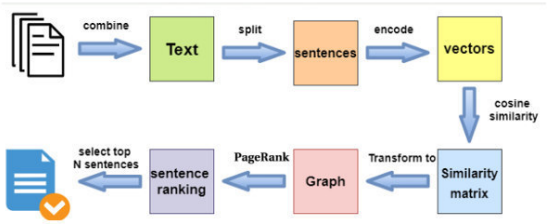


Fig. 3. Flow of TextRank Algorithm

Preprocessing: This step is similar to the preprocessing performed in the first step but excludes part-of-speech (POS) tagging.

Concatenate Sentences: The sentences are then combined into a single document.

Apply TextRank Algorithm: The TextRank algorithm is applied to the combined document to generate the final summary. The image below illustrates how the TextRank algorithm works. First, all text from the documents is combined into a single text, which is then split into sentences and preprocessed. Each sentence is tokenized into a collection of words and represented as a vector. Using the cosine similarity approach,

the similarity between sentence vectors is calculated and stored in a matrix. This matrix is transformed into a graph where sentences are represented as nodes, and the similarity between two sentences is represented as an edge. The PageRank algorithm is applied to this graph to calculate scores for each sentence, which are then ranked based on their importance. Finally, the top N sentences are selected based on their rank to generate the summary.

IV. Evaluation and Discussion

TABLE I
AVERAGE RECALL, PRECISION, AND F-MEASURE VALUES OF OUR SYSTEM USING BM25 OKAPI.

Metric	Average Precision (%)	Average Recall (%)	Average F-Measure (%)
Rouge-1	58.08	59.56	58.09
Rouge-2	36.39	38.72	37.31
Rouge-L	38.41	39.27	38.52
Rouge-W	24.06	7.84	11.70

TABLE II
AVERAGE RECALL, PRECISION, AND F-MEASURE VALUES OF OUR SYSTEM USING BM25 AND WORDNET

Metric	Average Precision (%)	Average Recall (%)	Average F-Measure (%)
Rouge-1	69.69	66.72	67.68
Rouge-2	54.16	51.29	52.28
Rouge-L	58.19	56.19	56.90
Rouge-W	45.18	13.18	20.52

A good summary balances precision and recall, ensuring relevance while capturing all critical information. The tables above compare the performance of two approaches—BM25 alone and WordNet + BM25 The proposed method, which combines WordNet with BM25, demonstrates superior results compared to BM25 alone. By enriching BM25 with semantic understanding from WordNet, the system overcomes key limitations inherent to BM25's design, thereby producing summaries that are more aligned with human-written counterparts.

Limitations of BM25 and How WordNet Addresses Them

BM25 is a probabilistic ranking algorithm that evaluates term relevance based on term frequency (TF) and inverse document frequency (IDF). While highly effective for ranking based on keyword matches, BM25 has several notable limitations:

Reliance on Exact Matches: BM25 depends on the exact overlap of query terms and document content to calculate relevance. It cannot identify relationships between terms with similar meanings, such as synonyms or semantically related words (e.g., "disease" and "illness").

WordNet's Contribution: WordNet resolves this limitation by providing a rich lexical database that captures semantic relationships such as synonyms, hypernyms, and hyponyms. This allows the system to identify and rank sentences containing contextually related words, even when exact matches are absent.

Lack of Contextual Understanding: BM25 treats terms independently, failing to account for word meaning based on context. For example, it would treat the word "virus" in "computer virus" the same as in "biological virus," potentially ranking irrelevant sentences higher.

WordNet's Contribution: WordNet enhances contextual understanding by incorporating part-of-speech tagging and synset identification. This ensures that terms are interpreted correctly based on their usage, improving the relevance of selected sentences.

Ambiguity in Term Usage: BM25 struggles with polysemy (words with multiple meanings), as it does not differentiate between them. For instance, the term "patient" could refer to a medical patient or someone showing patience.

WordNet's Contribution: Through its synset-based approach, WordNet assigns the appropriate sense to ambiguous words, ensuring that the sentences retrieved align with the intended meaning of the query.

Inability to Handle Implicit Relationships: BM25 focuses on explicit term matches and cannot infer implicit relationships, such as those between "treatment" and "therapy" or "vaccine" and "immunization."

WordNet's Contribution: WordNet's semantic similarity measures, such as Wu-Palmer similarity, enable the system to detect related terms based on their shared ancestry in the lexical hierarchy. This helps retrieve sentences that capture the essence of the query, even when specific terms are absent.

a) Evaluation Metrics:: Recall, Precision, and F-Measure The performance of the proposed system was evaluated using standard metrics:

Recall: Measures the proportion of relevant sentences retrieved. High recall indicates the system's ability to cover all important information. As shown in Table IV, the combination of WordNet and BM25 achieves significantly higher recall (66.72%) compared to WordNet alone. By addressing BM25's reliance on exact term matches and enhancing it with semantic understanding, the system retrieves a broader range of relevant sentences.

Precision: Evaluates the proportion of retrieved sentences that are actually relevant. Higher precision reflects the system's ability to avoid irrelevant content. Table IV demonstrates that WordNet + BM25 outperforms WordNet alone in precision. WordNet's semantic relationships allow the system to prioritize contextually appropriate sentences, reducing noise and redundancy in the summary.

F-Measure: Combines precision and recall into a single metric to evaluate overall performance. Results shows that the combined system achieves the highest F-measure (67.68%), striking a balance between retrieving relevant sentences (recall) and ensuring their quality (precision).

As a conclusion, the integration of WordNet into BM25 enhances its capabilities by addressing the limitations of exact term matching, lack of contextual understanding, polysemy, and implicit relationships. By leveraging WordNet's lexical and semantic resources, the system generates summaries that are more comprehensive and contextually accurate. This

synergy between BM25's ranking efficiency and WordNet's semantic richness ensures that the summaries are both precise and meaningful, closely aligning with human-written summaries.

V. Conclusion

The exponential growth of academic literature poses significant challenges for researchers and professionals seeking to extract relevant information efficiently. This paper presents a two-step framework for query-based multi-document summarization designed to address these challenges. In the first step, each individual document is processed through extractive summarization based on a given query, leveraging algorithms like BM25 and semantic similarity measures.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). Automatic text summarization: A comprehensive survey. Expert systems with applications, 165, 113679.

[9] Hristidis, V., Chen, S. C., Li, T., Luis, S., & Deng, Y. (2010). Survey of data management and analysis in disaster situations. Journal of Systems and Software, 83(10), 1701-1714.

Investigating the Electronic Properties of Doped SrO for Advanced Spintronic Applications

line 1: Mokaddem khadidja
line 2: *Department of Science and Technology*
line 3: *university belhadj boucchaib*

line 4: ain temouchent algeria
line 5: khadidja.mokaddem@univ-temouchent.edu.dz

line 1: bensaid djillali
line 2: *Department of Science and Technology*

line 3: *university djillali lyabes*
line 4: sidi belabas

Abstract*—This study investigates the half-metallic ferromagnetic properties of strontium oxide (SrO) doped with vanadium (V) at a concentration of 0.125. Using advanced density functional theory (DFT) calculations with a 16-atom supercell and applying the generalized gradient approximation (GGA-PBE), we demonstrate that doping SrO with vanadium results in nearly 100% spin polarization, establishing a robust half-metallic state. Our findings reveal that vanadium induces significant total magnetic moments and strong ferromagnetic interactions at this concentration. This research highlights the potential of doping to manipulate the electronic and magnetic properties of SrO, offering promising opportunities for the design of advanced spintronic devices.

***Keywords*—Half-metallicity, Ferromagnetism, Doping,Spin polarization, Strontium oxide (SrO)**

I. INTRODUCTION (*HEADING 1*)

Research into transition metal-doped semiconductors has attracted significant attention due to their potential in developing advanced spintronic materials. Spintronics, which exploits the spin of electrons alongside their charge, offers transformative possibilities for electronic devices, including increased transistor density, reduced energy consumption, and non-volatile features[1-3]. These innovations are essential for addressing challenges related to device miniaturization and the rising demand for energy-efficient technologies.

II. In this study, we focus on the electronic and magnetic properties of strontium oxide (SrO) doped with vanadium (V) at a concentration of 0.125. By employing a 16-atom supercell model and using the full potential linearized augmented plane wave (FP-LAPW) [4]method within density functional theory (DFT), we explore how vanadium doping affects SrO. Our primary goal is to demonstrate that this doping can induce half-metallic ferromagnetic behavior with nearly 100% spin polarization. The findings are expected to provide valuable insights into optimizing SrO for spintronic applications, contributing to the development of new materials with enhanced performance.

III. EASE OF USE

A. *Selecting a Template (Heading 2)*

I. CALCULATION DETAILS:

All calculations were performed using the full-potential linearized augmented-plane-wave (FP-LAPW) method [4] as implemented in the WIEN2k package developed by Blaha, Schwartz, and their collaborators [5]. The perdew etal version of generalized gradient approximation (GGA-PBE) for the exchange-correlation functional in density functional theory (DFT) was employed [6]. The muffin-tin sphere radii were selected to minimize the interstitial space and ensure that the muffin-tin spheres did not overlap. The value of RMT × Kmax was set to 7 for the plane-wave expansion, and the wave function expansion was truncated at lmax = 10 within the muffin-tin spheres. The charge density was expanded in Fourier series up to Gmax = 12 a.u.–1, where Gmax represents the largest vector in the Fourier series expansion. A 12 × 12 × 12 k-point mesh was used for the integration over the Brillouin zone.dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.

II. RESULTS AND DISCUSSION

1): Structural Properties.

Strontium oxide (SrO) crystallizes in the rocksalt (NaCl) structure, characterized by a face-centered cubic arrangement. This configuration promotes strong ionic bonding between strontium (Sr) and oxygen (O) atoms.

The table below summarizes the calculated structural properties for SrO and its vanadium-doped analogue. The lattice constant for SrO is 5.1442 Å, which closely corresponds to the experimental value of 5.16 Å and the theoretical predictions of 5.21 Å obtained using the GGA-WC method. This agreement validates the computational methods used. The bulk modulus of SrO is determined to be 94.06 GPa, indicating substantial resistance to compression, with a pressure derivative (B') of 3.92, reflecting the rigidity of the lattice.

Doping SrO with vanadium leads to significant changes in the structural properties. Specifically, the lattice constant for $Sr_{0.875}V_{0.125}O$ increases to 10.06 Å, suggesting a lattice expansion due to the larger atomic radius of V compared to Sr. Furthermore, the bulk modulus of the vanadium-doped compound is measured at 101.38 GPa, exceeding that of undoped strontium oxide. This improvement implies increased resistance to compression and a potentially stiffer lattice,

attributed to the electronic interactions between vanadium and the SrO matrix..

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

	a (A°)	*B(Gpa)*	*B'*	*method*
SrO	5,1442	94 ,06	3,92	GGA-PBE
$Sr_{0,875}V_{0,125}O$	10,069	101,38	4,35	GGA-PBE
SrO	5,21[7]	83,04	4,21	GGA-WC
SrO	5,16[8]	91,00	4,30	expérimental

Table 1 Calculated lattice constant (a), bulk modulus (B), and its pressure derivative (B′) for SrO and $Sr_{0,875}V_{0,125}O$

2)Electronic Structures.

The analysis of the total and partial density of states for the material SrO and its vanadium-doped analogue (SrVO) reveals distinct electronic characteristics.

For SrO, the spin-up and spin-down channels display semiconductor behavior, meaning they possess an energy gap around the Fermi level (EF). This gap indicates that, although there are electronic states below EF, those above it are empty. In contrast, for SrVO, the material exhibits half-metallic character, which means it is conductive for one spin channel (spin-majority) and insulating for the other (spin-minority). In SrVO, the majority spin states (spin-up) are primarily localized in the D-t2g states, which are lower in energy compared to the D-eg states. This energy separation is crucial, as it allows the majority spin electrons to contribute to the conductivity of the material.

Overall, the transition from SrO to SrVO illustrates how doping with transition metals, such as vanadium, can significantly modify the electronic properties of a material, paving the way for potential applications in spintronics and other advanced technologies.

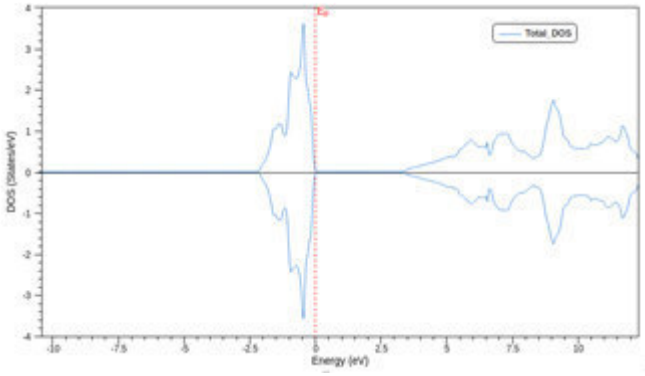
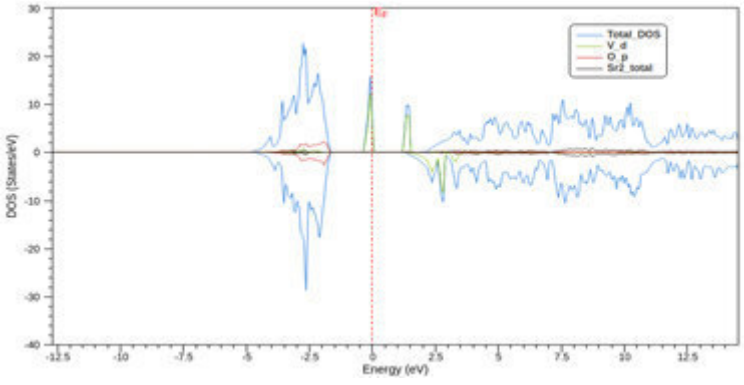


Figure 1: Contributions of total DOS in SrO compond



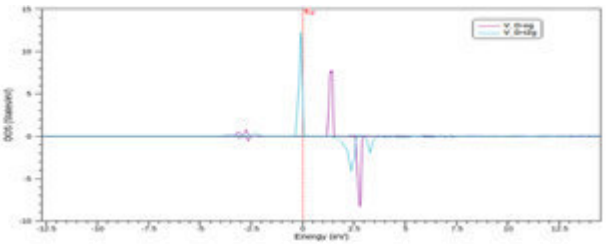


Figure 1: Contributions of Majority and Minority Spin States in the D-t2g and D-eg Orbitals of SrXO Compounds

CONCLUSION

In this study, we examined the electronic properties of strontium oxide (SrO) doped with vanadium (V), highlighting its half-metallic character. Using density functional theory (DFT) calculations with the generalized gradient approximation (GGA) and the Perdew-Burke-Ernzerhof (PBE) functional, we demonstrated that the hybridization between the t2g states of the vanadium impurities and the 2p states of the oxygen atoms plays a crucial role in the formation of the half-metallic state.

Simulations were conducted on 16-atom supercells with an impurity concentration of 0.125, allowing for a precise representation of the effects of doping on material properties. The results show that SrO doped with vanadium exhibits metallic conduction for majority spin electrons while also displaying high spin polarization.

The ability to modify doping with vanadium offers opportunities to finely tune the electronic and magnetic properties of SrO. These adjustments allow for the design of materials with specific characteristics suited to particular applications, such as magnetic data storage devices and spin sensors.

For future research, it would be relevant to further explore the effects of impurity concentrations and experimental conditions, as well as to conduct experimental studies to validate our theoretical predictions. Investigating the thermodynamic properties and stability of these compounds could also provide additional insights for their practical integration into devices.

[1] Sato, K., Katayama-Yoshida, H.: Material design of GaN-based ferromagnetic diluted magnetic semiconductors. Jpn. J. Appl. Phys. 40, L485–L487 (2001)

[2] Matsukura, F., Tokura, Y., Ohno, H.: Control of magnetism by electric fields. Nat. Nanotechnol. 10, 209–220 (2015)I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[3] Kaminska, M., Twardowski, A., Wasik, D.: Mn and other magnetic impurities in GaN and other III–V semiconductors-perspective for spintronic applications. J. Mater. Sci. Mater. E9. Rao, A.S., Kearney, R.J.: Phys. Stat. Sol. (b) 95, 243 (1979) R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] Kohn, W., Sham, L.J.: Self-consistent equations including exchange and correlation effects. Phys. Rev. 140, A1133–A1138 (1965)M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[5] K. Schwarz, P. Blaha, and G. K. H. Madsen, Comput. Phys. Commun. 147, 71 (2002)..

[6] Kohn, W., Sham, L.J.: Self-consistent equations including exchange and correlation effects. Phys. Rev. 140, A1133–A1138(1965)

[7] S. Berri, A. Kouriche, D. Maouche, F. Zerarga, M. Attallah,
Sci.Semicond.Process.38,101(2015)

[8] L.G. Liu, W.A. Bassett, J. Geophys. Res. 77, 4934 (1972) prescribed.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

Self-similar solitons in cubic-quintic media

Amel Mecelti
Radiation and Matter Physics Laboratory, Matter Sciences Department
Mohamed-Cherif Messaadia University
Souk-Ahras, Algeria
amelmecelti3@gmail.com

Houria Triki
Radiation Physics Laboratory, Department of Physics
Faculty of Sciences, Badji Mokhtar University
Annaba, Algeria
trikihouria@gmail.com

***Abstract*— We present different types of chirped self-similar waves for a generalized derivative nonlinear Schrödinger equation with varying dispersion, self–steepening effect, cubic–quintic nonlinearity, and gain or loss. The equation arises in modeling femtosecond light propagation in an optical fiber with spatial parameter variations. The newly found self-similar structures take the shape of gray and kink pulses. . The dynamical behaviors of the chirped self-similar gray and kink waves are analyzed in a periodic distributed amplification system. The acquired self-similar structures display rich dynamical evolutions that are important in practical applications**

Keywords— Cubic-quintic, Solitons,Chirp

Introduction

Studies of light pulses that propagate self-similarly in nonlinear optical materials have been addressed in recent years because of their relevance to understand widely various physical phenomena [1]. In nonlinear optics, significant results have been reported by studying

the self-similar optical waves through parabolic intensity profiles in an optical fiber amplifier [2,3].

the self-similar regime of collapse for spiral laser beams [8], and stimulated Raman scattering [9]. It should note that these self-similar pulses can sustain their shapes but allow their widths and amplitudes to vary via the parameters, such as the dispersion, nonlinearity, inhomogeneity and gain or loss [10]. Because of their remarkable properties, optical self-similar waves are useful for diverse applications in fiber-optic telecommunications and photonics [11].

The nonlinear Schrödinger equation (NLSE) with distributed coe cients is normally employed to model self-similar wave propagation in inhomogeneous Kerr-type media [12–20]. More specifically, the variable coe cient NLSE can be considered as a practical model for the dynamics of the soliton in inhomogeneous physical systems [12–22]. Such model possesses exact tanh- and sech-shaped self-similar optical waves which are linearly chirped [12]. The chirping property of these self-similar waves makes them useful in optical fiber amplifiers and optical pulse compressors [12–21].

A. Solitary waves of the model of the nonlinear Schrödinger equation derived and generalized

The generalized derivative nonlinear Schrödinger reads [10]

$$iU_Z = \beta U_{TT} + \sigma|U|^2U + \epsilon|U|^4U + i[\lambda + \mu|U|^2U]_T \qquad (1)$$

where $U(T,Z)$ is the electric field, β is the group velocity dispersion parameter, σ and ϵ are the quintic and cubic nonlinearity parameters sequentially, λ is the parameter associated with the group velocity, and μ is related to the self-steepening effect. This equation including higher order effects is relevant to govern the dynamics of femtosecond optical pulses in optical fibers.

To obtain exact traveling wave solutions of the model (1), we consider the ansatz solution [11]:

$$U(T,Z) = u(\xi)e^{[i(kZ - \Omega T) + i\varphi(\xi)]} \qquad (2)$$

where $\varphi(\xi)$ is the phase shift, $u(\xi)$ is the amplitude, $\xi = T - qZ$ is the traveling coordinate, and $q = V^{-1}$ is the inverse velocity of the wave. In addition, Ω and k are real parameters that sequentially represent the frequency shift and the propagation constant.

Plugging (2) into (1) gives the two coupled ordinary differential equations:

$$u(k - q\varphi') + \beta[u'' - u(\varphi' - \Omega)^2] + \sigma u^3 + \epsilon u^5 + \lambda u(\Omega - \varphi') + \mu u^3(\Omega - \varphi') = 0 \qquad (3)$$

and

$$(q + \lambda - 2\beta\Omega)u' + \beta(u\varphi'' + 2u'\varphi') + 3\mu u^2 u' = 0 \qquad (4)$$

Equation (4) can be written as an ordinary differential equation for the amplitude function $u(\xi)$ as:

$$u'' + \frac{(q+\lambda)^2 + 4\beta(K - q\Omega) + 2\mu G}{4\beta^2}u + \frac{2\beta + \mu(q+\lambda)}{2\beta^2}u^3 + \frac{3u^2 + 1}{1}\frac{\epsilon}{\beta^2}u^5 - \frac{G^2}{\beta^2 u^3} = 0 \qquad (5)$$

The latter equation describes the dynamics of the wave amplitude when the pulse propagates through an optical fiber medium governed by the generalized derivative nonlinear Schrödinger (1).

Using the change of variable $Y = u^2$, we obtain the following first-order differential equation:

$$(Y')^2 + aY^4 + bY^3 + cY^2 + dY + H = 0 \qquad (6)$$

where

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

$$a = \frac{3u^2+1\ \epsilon}{1\ \beta^2}\ ,\quad b = \frac{2\beta\ +\mu(q+\lambda)}{\beta^2}\ ,\quad c = \frac{(q+\lambda)^2+4\beta(K-q\Omega)+2\mu}{\beta^2}\ ,$$

$$c = 8K\ ,\quad H = \frac{4G^2}{\beta^2} \tag{7}$$

with K is a constant of integration.

In the following, we present the exact solitary wave solutions of the model (1) obtained by solving the nonlinear differential equation (6).

B. Chirped gray solitude waves

We have found that for equation (7) admits a gray-type solitary wave solution of the form

$$Y(\xi) = \rho - \frac{\Lambda}{1+D\quad \hbar[\eta(\xi-\xi_0]} \tag{8}$$

Therefore, we obtain the following the chirped solitary wave solution on a constant coefficient equation for the generalized derivative nonlinear Schrödinger (1):

$$U(Z,T) = \sqrt{\rho - \frac{\Lambda}{1+D\quad \hbar[\eta(T-q(Z-Z_0)]}}\, e^{i(K\ -\Omega T)+i\iota\ (\xi)} \tag{10}$$

With $\rho > \Lambda/(1+D)$. We observe that this nonlinear waveform solution has three free parametersρ, Λ and η if the fiber parameters $\beta, \sigma, \epsilon, \lambda$ and μ are known. Figure 1 shows an example of solitary wave propagation (10).

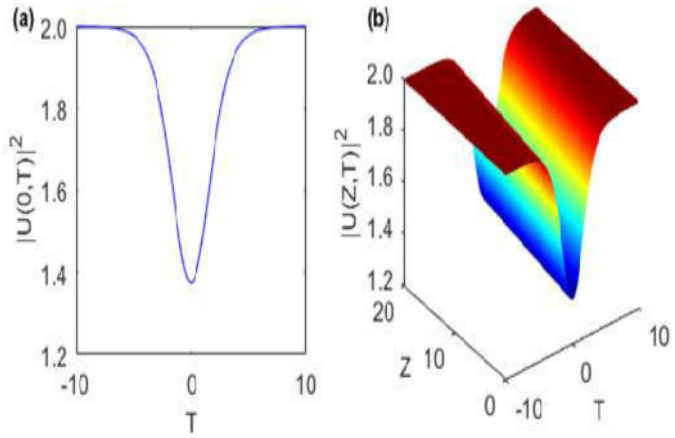


Figure 1: Intensity profile of the chirped gray solitary wave

C. Chirped gray solitary waves

We have found a kink-type solitary wave for equation (7) as

$$Y(\xi) = P - \frac{Q}{R+t_{\iota}\quad \hbar[W(\xi-\xi_0]} \tag{11}$$

Therefore, we get the following the chirped solitary wave solution on a constant coefficient equation for the generalized derivative nonlinear Schrödinger (1):

$$U(Z,T) = \sqrt{P - \frac{Q}{R+t_{\iota}\quad \hbar[w(T-q(Z-Z_0))]}}\, e^{i(k\ -\Omega T)+i\iota\ (\xi)} \tag{12}$$

Solution (1é) represents a chirped kink pulse that exists in the presence of all linear and nonlinear effects (figure 2).

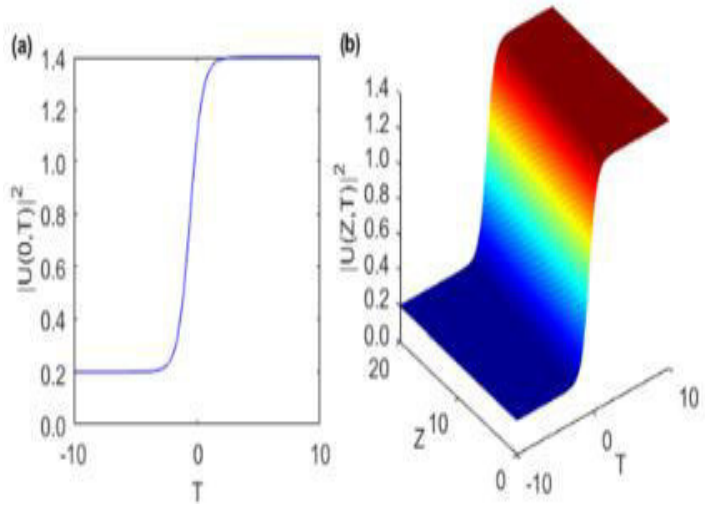


Figure 2: Intensity profile of the chirped kink solitary wave.

D. Similarity transformation of GD–NLSE with distributed coe cients

We have found that the propagation of a femtosecond pulse in an homogeneous fiber medium is governed by the derived and generalized nonlinear Schrödinger equation (1), for which the carrier wave is influenced by high-order nonlinear effects such as quintic nonlinearity and self-steepening. However, in reality, this kind of nonlinear medium does not exist, since the creation of materials without impurities and without mechanical and structural defects is very expensive and requires advanced technology. In the following, considering the inhomogeneity of the fiber system, we will examine the propagating properties of solitary waves in the presence of quintic nonlinearity and the self-steepening effect. For this, it is convenient to write the coefficients of the model (1) as functions dependent on the propagation distance z such that Eq. (1) in the presence of the loss/gain term becomes:

$$i\psi_z = \alpha(z)\psi_{t\iota} + \gamma(z)|\psi|^2\psi + \delta(z)|\psi|^4\psi + i[(s(z) + \upsilon(z)|\psi|^2)\psi]_t + il\ (z)\psi \tag{13}$$

where $\alpha(z)$, $\gamma(z)$, $\upsilon(z)$ and $\delta(z)$ are the variable group velocity, cubic nonlinearity, self-steepening, and quintic non linearity parameters sequentially. In addition, $\Gamma(z)$ represents the distributed gain (> 0) or loss (< 0) coefficient. Below we examine the existence condition for self-similar waves within the framework of Eq. (13).

To find self-similar waves with equation (13), we use the transformation

$$\psi(z,t) = A(z)\, U[T(z,t),\, Z(z)]e^{i\emptyset(z,t)} \tag{14}$$

here $\emptyset(z,t)$ and $A(z)$ are the real phase and amplitude functions sequentially, while U is an unknown complex function of $Z = Z(z)$ and $T = T(z,t)$.

Substituting (14) into 13), the generalized nonlinear Schrödinger equation (13) reduces into the model (1) with the results :

$$\beta = \frac{\alpha(T_t)^2}{Z_z}, \mu = \frac{\upsilon A^2 T_t}{Z_z}, \quad \epsilon = \frac{\delta A^4}{Z_z} \tag{15}$$

Therefore to determine the general self-similar wave type solution of the generalized derivative nonlinear Schrödinger equation (13) in the form:

$$\psi(z,t) = A_0 U\Big\{ k\Big[t + \Big(2p - \frac{\lambda}{\beta}\Big)\int_0^z \alpha(\zeta)d\zeta + \int_0^z s(\zeta)d\zeta\Big] + t_0\frac{k^2}{\beta}\Big[\int_0^z \alpha(\zeta)d\zeta\Big]\Big\} \times e^{\,[\int_0^z \Gamma(\zeta)d\zeta + i\emptyset(z,t)]} \tag{16}$$

The similarity transformation (16) allow us to determine different types of self-similar waves of equation (13) is the exact analytical solutions of model (1) are known.

E. Gray self-similar solitary wave

The general self-similar solution (16) of the generalized derivative nonlinear Schrödinger equation with varied coe cients (13) can be considered as

$$\psi(z,t) = A_0\Big[\int_0^z \Gamma(\zeta)d\zeta\Big]\sqrt{Y(\xi)}\, e^{\,[i\emptyset(z,t)]} \tag{17}$$

Note that the associated chirp is given by

$$\delta\ (z,t) = \frac{3\mu}{4\beta}\Big(\rho - \frac{A}{1+D\ \ h[\eta(\xi-\xi_0)]}\Big) - \frac{k}{\beta}\Big(\rho - \frac{A}{1+D\ \ h[\eta(\xi-\xi_0)]}\Big)^{-1}\frac{(q+\lambda)k-2p}{2\beta} \tag{18}$$

It is clear that the chirp presents a nontrivial form which depends on the intensity of chirped self-similar pulses. Note that the general form of the chirp can be written as

$$\delta\ (z,t) = \frac{3\mu}{4\beta}u^2 - \frac{k}{\beta u^2} + \frac{(q+\lambda)k-2p}{2\beta} \tag{19}$$

This last expression shows that the chirp (z,t) depends on the pulse intensity $|U(T,Z)|^2 = |u(\xi)|^2$. We also note that the parameter can be used to control the amplitude of the chirp, which means that the self-steepening effect is a crucial process for self-similar pulses to be chirped nonlinearly.

Substituting equation (8) into equation (17), we obtain a new gray self-similar solitary wave for the model (13) as

$$\psi(z,t) = A_0\Big[\int_0^z \Gamma(\zeta)d\zeta\Big]\Big[\rho - \frac{A}{1+D\ \ h[\eta(\xi-\xi_0)]}\Big]^{1/2} e^{\,[i\emptyset\ (z,t)]} \tag{21}$$

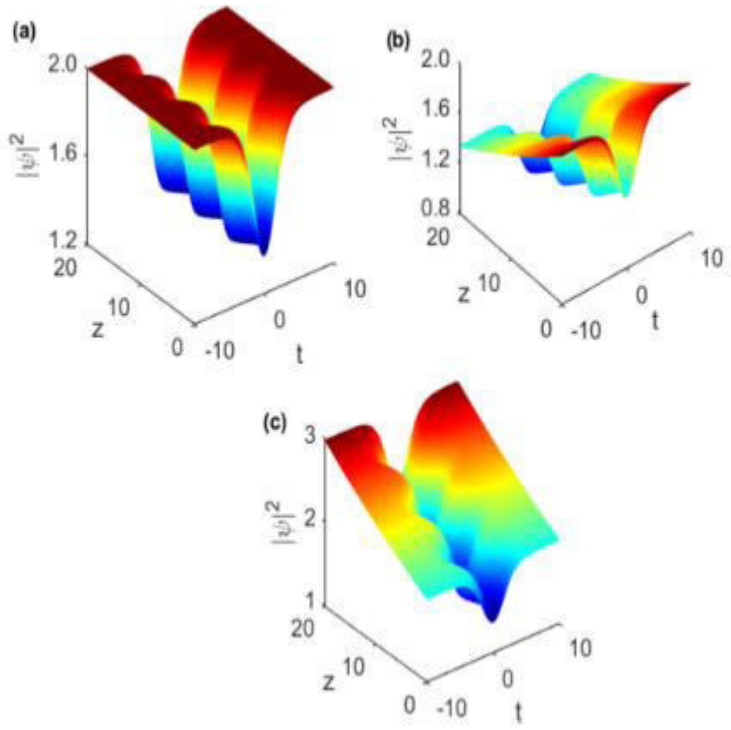


Figure 4: Evolution of the chirped gray self-similar wave, for (a) h=0 (b) h=-0.02 (c) h=0.02.

By inserting equation (11) into (17), we can write the second new self-similar gray solitary wave of the generalized model (22) in the following form:

$$\psi(z,t) = A_0\Big[\int_0^z \Gamma(\zeta)d\zeta\Big]\Big[P - \frac{Q}{R+t\ \ h[w(\xi-\xi_0)]}\Big]^{1/2} e^{\,[i\emptyset\ (z,t)]} \tag{22}$$

The corresponding chirp takes the following form

$$\delta\ (z,t) = \frac{\exists\mu}{4\beta}\left(P - \frac{Q}{R+t_i\ \ h[w(\xi-\xi_0)]}\right) - \frac{\kappa}{\beta}\left(P - \frac{Q}{R+t_i\ \ h[w(\xi-\xi_0)]}\right)^{-1} + \frac{(q+\lambda)\kappa-2p}{2\beta} \quad (22)$$

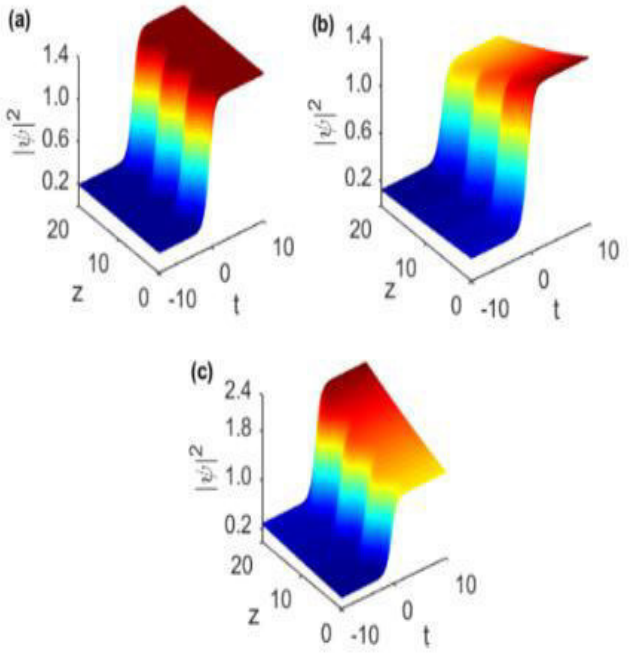


Figure 5: Evolution of the chirped self-similar kink wave for (a) h=0, (b) h=-0.02 (c) h=0.02.

F. Conclusion

The generalized derivative nonlinear Schrödinger equation with distributed coe cients modeling the dynamics of femtosecond solitary waves in an inhomogeneous optical fiber is considered. The model contains distributed self-steepening effect, group velocity dispersion, cubic-quintic nonlinearity and gain or loss. Based on the similarity transformation algorithm, the chirped self-similar waves are constructed. In particular, new self-similar solitary waves with a nonlinear chirp, including gray and kink pulses. The nonlinear chirp accompanying the acquired self-similar solitary waves results from the self-steepening process. The chirped self-similar waves are adopted via a periodic distributed amplification system. The dynamical evolution of the chirped self-similar optical waves can be governed by choosing the dispersion and cubic nonlinearity parameters appropriately. Undoubtedly, the novel gray and kink solitary waves presented here are of advantage in studying the transmission properties of light pulses since they possess four important characteristics in the same time: propagate self-similarly, localized structurally, chirped nonlinearly, and sited on a nonzero background.

References

[1] G.I. Barenblatt, Scaling, Self-Similarity, and Intermediate Asymptotics, Cambridge University Press, Cambridge, UK, 1996.
[2] M.E. Fermann, V.I. Kruglov, B.C. Thomsen, J.M. Dudley, J.D. Harvey, Self-similar propagation and amplification of parabolic pulses in optical fibers, Phys. Rev. Lett. 84 (2000) 6010.
[3] V.I. Kruglov, A.C. Peacock, J.M. Dudley, J.D. Harvey, Self-similar propagation of high-power parabolic pulses in optical fiber amplifiers, Opt. Lett. 25 (2000) 1753–1755.
[4] F.Ö. Ilday, J.R. Buckley, W.G. Clark, F.W. Wise, Self-similar evolution of parabolic pulses in a laser, Phys. Rev. Lett. 92 (2004) 213902.
[5] T. Hirooka, M. Nakazawa, Parabolic pulse generation by use of a dispersion-decreasing fiber with normal group-velocity dispersion, Opt. Lett. 29 (2004) 498–500.
[6] J.F. Zhang, Q. Tian, Y.Y. Wang, C.Q. Dai, L. Wu, Self-similar optical pulses in competing cubic-quintic nonlinear media with distributed coe cients, Phys. Rev. A 81 (2010) 023832.
[7] J.F. Zhang, L. Wu, L. Li, D. Mihalache, B.A. Malomed, Vortices and ring dark solitons in nonlinear amplifying waveguides, Phys. Rev. A 81 (2010) 023836.
[8] V.I. Kruglov, A.C. Peacock, J.D. Harvey, Exact self-similar solutions of the generalized nonlinear Schrödinger equation with distributed coe cients, Phys. Rev. Lett. 90 (2003) 113902.
[9] V.I. Kruglov, A.C. Peacock, J.D. Harvey, Exact solutions of the generalized nonlinear Schrödinger equation with distributed coe cients, Phys. Rev. E 71 (2005) 056619.
[10] W.V. Saarloos, P.C. Hohenberg, Front pulses sources and sinks in generalized complex Ginzburg–Landau equations, Physica D 56 (1992) 303–367.
[11] H. Triki, V.I. Kruglov, Chirped self-similar solitary waves in optical fibers governed with self-frequency shift and varying parameters, Chaos Solitons Fractals 143 (2021) 110551.
[12] V.I. Kruglov, A.C. Peacock, J.D. Harvey, Exact self-similar solutions of the generalized nonlinear Schrödinger equation with distributed coe cients, Phys. Rev. Lett. 90 (2003) 113902.
[13] V.I. Kruglov, A.C. Peacock, J.D. Harvey, Exact solutions of the generalized nonlinear Schrödinger equation with distributed coe cients, Phys. Rev. E 71 (2005) 056619.
[14] Q. Zhou, T. Wang, A. Biswas, W. Liu, Nonlinear control of logic structure of all-optical logic devices using soliton interactions, Nonlinear Dyn. 107 (2022) 1215–1222.
[15] Q. Zhou, Y. Zhong, H. Triki, Y. Sun, S. Xu, W. Liu, A. Biswas, Chirped bright and kink solitons in nonlinear optical fibers with weak nonlocality and cubic-quintic-septic nonlinearity, Chin. Phys. Lett. 39 (2022) 044202.
[16] Q. Zhou, Influence of parameters of optical fibers on optical soliton interactions, Chin. Phys. Lett. 39 (2022) 010501.
[17] Y. Fang, G. Wu, Y. Wang, C. Dai, Data-driven femtosecond optical soliton excitations and parameters discovery of the high-order NLSE using the PINN, Nonlinear Dyn. 105 (2021) 603–616.
[18] S. Chen, L. Wang, L. Wen, C. Dai, J. Liu, X. Zhang, Stripe-on-plane-wave phase of a binary dipolar Bose gases with soft-core long-range interactions, Optik 247 (2021) 167932.
[19] J. Zhao, Z. Luan, P. Zhang, C. Dai, A. Biswas, W. Liu, N. Kudryashov, Dark three-soliton for a nonlinear Schrödinger equation in inhomogeneous optical fiber, Optik 220 (2020) 165189.
[20] Y. Zhao, Y. Lei, Y. Xu, S. Xu, H. Triki, A. Biswas, Q. Zhou, Vector spatiotemporal solitons and their memory features in cold Rydberg gases, Chin. Phys. Lett. 39 (2022) 034202.
[21] V.N. Serkin, A. Hasegawa, Novel soliton solutions of the nonlinear Schrödinger equation model, Phys. Rev. Lett. 85 (2000) 4502.
[22] S. Chen, L. Yi, Chirped self-similar solutions of a generalized nonlinear Schrödinger equation model, Phys. Rev. E 71 (2005) 016606.
[23] G.P. Agrawal, C. Headley III, Kink solitons and optical shocks in dispersive nonlinear media, Phys. Rev. A 46 (1992) 1573–1577.
[24] H. Triki, Q. Zhou, A. Biswas, S.L. Xu, A.K. Alzahrani, M.R. Belic, Self-frequency shift effect for chirped self-similar solitons in a tapered graded-indexed waveguide, Opt. Commun. 468 (2020) 125800.
[25] A. Merabti, H. Triki, F. Azzouzi, Q. Zhou, A. Biswas, W. Liu, A.K. Alzahrani, A. EL–Akrmi, Propagation properties of chirped optical

similaritons with dual-power law nonlinearity, Chaos Solitons Fractals 140 (2020) 110158.
[26] W.V. Saarloos, P.C. Hohenberg, Front pulses sources and sinks in generalized complex Ginzburg–Landau equations, Physica D 56 (1992) 303–367.

Detection of QRS by Pan-Tompkins Algorithm applied in the middle variations signal

Siham MOUSSAOUI
Department of Electrical Engineering Systems, Systems and Telecommunications Engineering Laboratory
Boumerdes University, Algeria,
s.moussaoui@univ-boumerdes.dz.

Sidali FELLAG
Department of Electrical Engineering Systems, Systems and Telecommunications Engineering Laboratory
Boumerdes University, Algeria,
s.fellag@univ-boumerdes.dz.

Hocine CHEBI
Faculty of Electrical Engineering, Djillali Liabes University of Sidi Bel Abbes, Laboratory Intelligent Control and Electrical Power System (ICEPS), Sidi Bel Abbes, Algeria. chebi.hocine@yahoo.fr, hocine.chebi@univ-sba.dz, chebi.hocine@univ-boumerdes.dz. https://orcid.org/0000-0003-0940-8072

Yasmina Fadila DIDOUCHE
Department of Chemistry, Faculty of Sciences,
Boumerdes University, Algeria,
y.didouche@ univ-boumerdes.dz.

***Abstract*—The Pan-Tompkins QRS detection algorithm is evaluated using three distinct ECG databases to assess its performance comprehensively. The MIT-BIH Arrhythmia Database serves as a benchmark, allowing for the assessment of the algorithm's baseline accuracy and sensitivity in detecting QRS complexes within a diverse set of arrhythmias, free from noise interference. The MIT-BIH Noise Stress Test (NST) Database tests the algorithm's resilience to various noise types (e.g., baseline wander, muscle artifacts), simulating real-world conditions to measure robustness in noisy environments. Lastly, the Long-Term ST Database evaluates the algorithm's stability and sensitivity to prolonged recordings with ST segment changes, relevant for ischemic or other cardiac episodes, allowing insights into performance amidst clinically significant signal variations. This combined analysis provides a comprehensive performance profile, highlighting accuracy in arrhythmia detection, noise resilience, and long-term stability, essential for assessing the algorithm's applicability across diverse ECG scenarios**

Keywords—ECG, signal variations, QRS detection, Pan-Tompkins algorithm.

I. INTRODUCTION

With nearly 17.8 million deaths a year, cardiovascular diseases are the leading cause of death worldwide. A cardiogram (ECG) is the most commonly used diagnostic technique for cardiovascular disease. Every day, approximately 3 million ECGs are generated worldwide. As wearable technology advances, more ECGs are generated for examination. In order to reduce doctors' workload and analyze ECGs produced by wearable technology, automated diagnosis techniques are necessary. During a cardiac cycle, an electrocardiogram (ECG) visualizes the electrical signals produced. The heart rate rhythm and morphology are described in detail. Arrhythmias, hyperkalemia, and myocardial infarction can be detected by cardiologists using ECGs. The amplitudes and durations of the ECG signals can provide a great deal of medical information. An R-peak identification in ECG signal is essential for calculating heart rate and detecting arrhythmias.

In the Fig.1 shows the ECG signal for a normal-rhythm patient, which includes the P-wave, QRS complex and T-waves. A number of algorithms used for automated ECG analysis depend heavily on QRS complexes and R peaks[1-2]. By identifying the QRS complex and R-wave, the rest of the waves and ECG features can be determined [3] . The QRS changes in shape when the atria or ventricles depolarize or repolarize prematurely or slowly, which leads to a deviation from the typical sinus rhythm on the ECG [4].

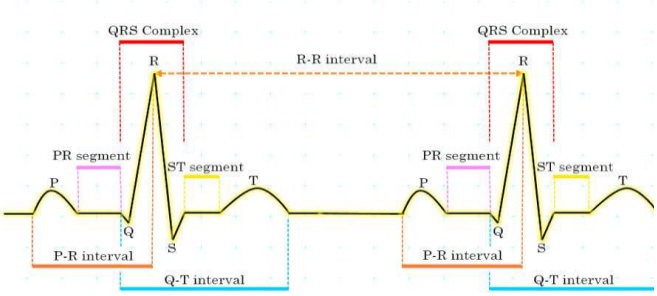


Fig.1. Segments and main characteristics of the ECG signal which is divided into two intervals.

The work reports QRS detection using high-quality clinical ECG signals. It is still unclear whether the Pan-Tompkins algorithm can detect QRS complexes in low-quality and noisy signals, especially in ambulatory ECGs, and whether it can detect clinically significant ST-T segment abnormalities. By using the MIT-BIH Noise Stress Test (NST) and Long-Term ST databases to analyze the Pan-Tompkins QRS detection algorithm, we are able to evaluate its performance from several perspectives.

xxx-w-xxx-xxxx-x/xx/$31.00 ©2024 IEEE

The remainder of this paper is organized as follows: Section 2 explains the addition of noise and artifacts present in an ECG signal; Section 3 describes the databases used in this study for QRS detection. Then, the algorithm implementation and the evaluation performances of the algorithms are presented in Sections 4 and 5. The last section summarizes this work.

II. PAN-TOMPKINS ALGORITHEM

Pan-Tompkins algorithms are commonly used in real-time QRS detection [5].To identify Rpeaks in QRS complexes, this algorithm uses amplitude, slope, and width of an integrated window [5]. There are two stages in the algorithm, namely preprocessing and decision making. During preprocessing, the raw ECG signal is prepared for detection. In preprocessing, noise is removed, the signal is smoothed, and the width and QRS slope are increased. During the decision stage, the thresholds are used to only consider the signal peaks and eliminate the noise peaks, as shown in Fig. 2. The algorithm consists of a:

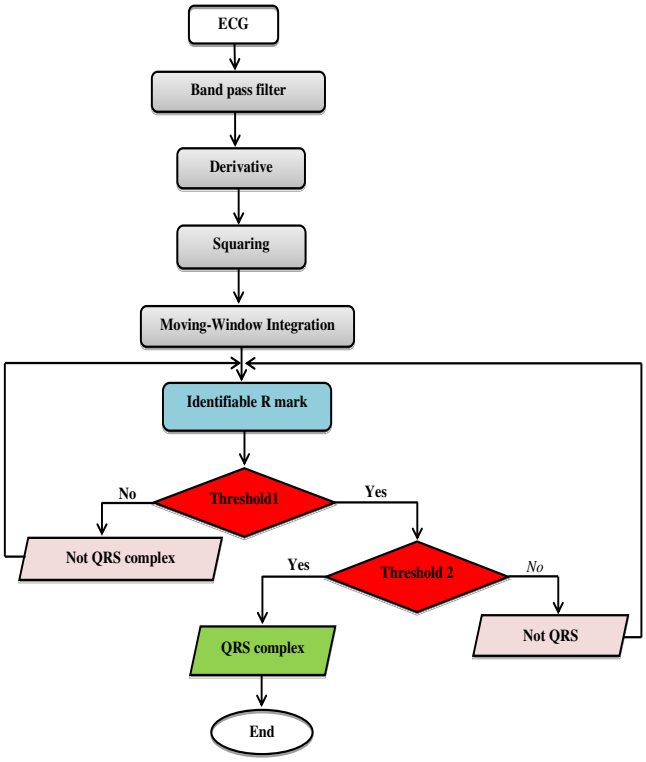


Fig.2. Flowchart represents the proposed method for QRS detection using the Pan & Tompkins algorithm.

A) *Band pass filter*

The signal passes through a cascading high pass and low pass digital filter. Using a band pass filter, muscle noise, 60 Hz interference, baseline wander, and T-wave interference are reduced. A desirable pass band for maximizing QRS energy is approximately 5-15 Hz. It consists of:

*low pass filter :*A second-order low-pass filter has the following transfer function (1), The difference equation of the filter is (2) , where T is the sampling period, the cutoff frequency is about 11 Hz and the gain is 36.The filter processing delay is six samples

$$H(z) = \frac{(1-z^{-6})^2}{(1-z^{-1})^2} \quad (1)$$

$$Y(nT) = 2y(nT - T) - y(nT - 2T) + x(nT) - 2x(nT - 6T) + x(nT - 12T) \quad (2)$$

B) *Derivative*

Once the signal is filtered, it is transformed to extract information about the slope of the QRS complex. This transformation is performed by a five-point derivative operation, using the following transfer function (3), and the difference equation is (4).

$$H(z) = \left(\frac{1}{8T}\right) - (-z^{-2} - 2z^{-1} + 2z^{1} + z^{2}) \quad (3)$$

$$y(nT) = \left(\frac{1}{8T}\right)[-x(nT - 2T) - 2x(nT - T) + 2x(nT + T) + x(nT + 2T)] \quad (4)$$

C) *Squaring Function*

Following the process of differentiation, the signal is squared at each individual point. The formula for this action is (5). This transformation ensures that every data point becomes positive and results in a nonlinear amplification of the derivative, highlighting the increased frequencies.

$$y(nT) = [x(nT)]^2 \quad (5)$$

D) *Moving-Window Integration*

As well as the slope of the R wave, moving-window integration is used to obtain waveform feature information. Calculated from (6), whereNis the number of samples in the integration window.

$$y(nT) = \left(\frac{1}{N}\right)[x(nT - (N-1)T) + x(nT - (N-2)T) + \cdots + x(nT)] \quad (6)$$

1. Decision

During the decision-making phase, it is determined whether the MWI result is a QRS complex or not. Two thresholds are applied to ensure that the right peak was picked. First, the signal is examined using the higher threshold (Threshold 1). In the absence of QRS within a predetermined time window, the lower threshold (Threshold2) is applied, requiring a search back method to locate the QRS. QRS complexes are identified by adjusting the thresholds. Based on the following equations, thresholds are calculated:

$$SPK = 0.125\,PEAK + 0.875\,SPK \quad (7)$$

$$NPK = 0.125PEAK + 0.875NPK \quad (8)$$

$$Thershold1 = NPK + 0.25(SPK - NPK) \quad (9)$$

$$Thershold_2 = 0.5\, Thershold_1 \qquad (10)$$

III. RESULTS AND DISCUSSION

Our work is about QRS region detection to improve the performance of Pan and Tompkins algorithm in noisy environment, with the change of ST segment period. Which influences on R peak detection. The databases used for algorithm validation are given as follows:

The MIT-BIH Arrhythmia Database (MIT) : In this database[6], 48 heartbeat recordings have an 11 bit resolution over 10 mV and are sampled at 360 Hz. From 1975 to 1979, 47 patients were studied by the Arrhythmia Laboratory at Beth Israel Hospital in Boston. For each patient, two channels were taken from modified limb lead II and one from modified lead V1, V2, V4 or V5. According to Mood and Mark [7],To identify the characteristics of heartbeats, Lead A, a modification of Lead II (MLII = modified limb lead II), is commonly used. Supraventricular ectopic beats and Ventricular ectopic beats are typically identified with Lead B, regular leads V1, V2, V5 or V4.

In addition to testing algorithms for QRS detection, the database is also used to detect and classify arrhythmias. Four records (102, 104, 107, and 217) include Beats. Originally, analog tapes could not represent pacemaker artifacts with sufficient fidelity to be recognized by pulse amplitude (or sweep rate) and duration alone, a method commonly used for real-time processing. Database recordings reproduce analog recordings with sufficient fidelity that pacemaker artifact detectors designed for tape analysis can be used [6].

The MIT-BIH Noise Stress Test Database (NST): There are 12 half-hour ECGs and three half-hour noise recordings in the NST database [8]. NST uses two clean ECG records from the MIT BIH Arrhythmia Database (Records 118 and 119) with signal-to-noise ratios of (24, 18, 12, 6, 0 and -6) db. The noise was added after 5 minutes of each file after two-minute segments alternated with clean sections. Using standard ECG recorders and equipment, three noise recordings were collected from physically active volunteers. As in an ambulatory signal, the noise recording contained artifacts such as the baseline wander, the EMG, and the motion artifact.

The Long-Term ST Database: The Long-Term ST [814], Database includes 86 prolonged ECG recordings from 80 subjects, each selected to demonstrate a variety of ST-segment changes, such as ischemic episodes, cardiac axis-related nonischemic episodes, slow ST-level drifts, and combinations of these phenomena. A recording typically lasts between 21 and 24 hours and contains two or three ECG signals digitized at 250 samples per second with 12-bit resolution, covering a range of ±10 mV. In addition to verified ST episodes, signal quality annotations, beat-to-beat QRS annotations, and ST level measurements are provided.

Applying the Moving Window Integration (MWI) method for QRS complex detection in the three ECG databases allows extracting key information about the R-wave slope using an integration window tailored to the characteristics of the QRS complexes. A window of 30 samples (150 ms) was chosen, corresponding to a sampling rate of 200 samples per second, which optimizes the capture of fast QRS slopes while minimizing interference with the slower P and T waves. In the results shown in Fig.3, MWI effectively highlights R-wave slopes in a moderately noisy environment, facilitating clear detection of QRS complexes. In Fig.4, the algorithm detects high amplitudes well, but noise results in false detections, indicating sensitivity to interference. Fig.5 shows similar results, where MWI manages to capture QRS complexes, although background noise can still cause detection errors.

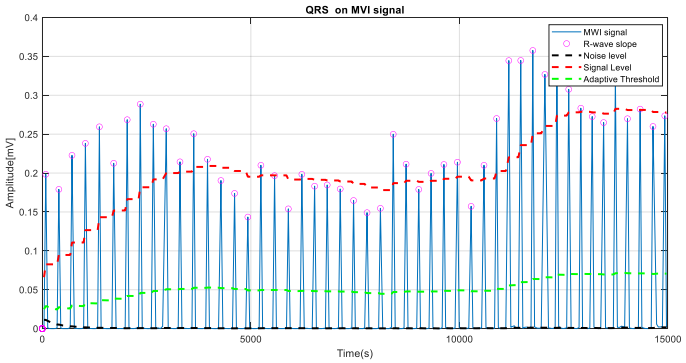


Fig.3. R-wave slope on MWI signal with MIT Database (100).

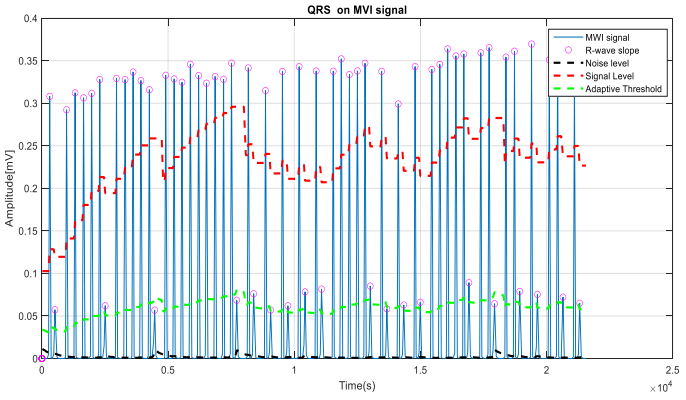


Fig.4. R-wave slope on MWI signal with NST Database (119e00).

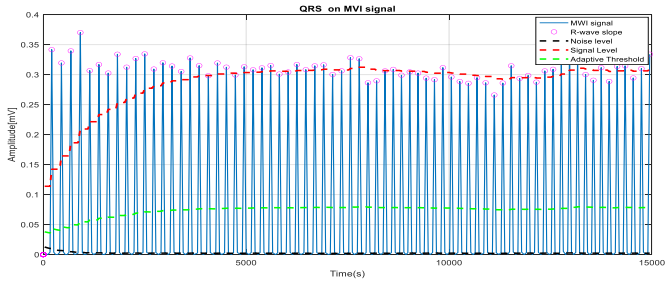


Fig.5. R-wave slope on MWI signal with Long-Term ST Database (s20011m).

Interpretation of QRS complex detection results with the Moving Window Integration (MWI) method includes a crucial decision phase that uses two adaptive thresholds, as illustrated in Fig.3, 4, and 5. In this decision phase, each detected peak is evaluated to determine whether it corresponds to a QRS complex. The thresholds are dynamically adjusted to adapt to signal variations, allowing the differentiation of true QRS complexes from fluctuations due to noise or abnormal ST segments. This step ensures that only relevant peaks are retained, thereby optimizing detection accuracy in the MIT, NST, and Long-Term ST databases despite varying signal conditions. By combining threshold adjustment with a refined decision phase, the process ensures robust and reliable QRS complex detection as shown in Fig.6.

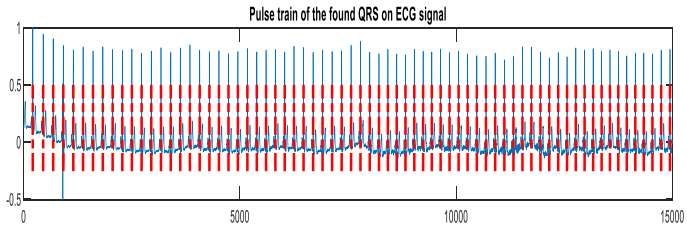


Fig. 6. QRS complex detected in Long-Term ST Database (s20011m).

The results obtained using MATLAB simulations are then presented and compared with each other. To assess the algorithm's ability to detect QRS complexes, a binary classification was used. This is based on three detection states: TP (true positive) which represents the correct detection of a QRS complex by the algorithm when a QRS complex is actually present. FP (false positive) which correspond to the erroneous detections of a QRS complex by the algorithm when no QRS complex is present. FN (false negative), which reflect the absence of detection of a QRS complex by the algorithm when it is actually present. From these states, two main indicators were defined. Sensitivity (Se) measures the probability that a QRS complex is correctly detected, while the positive predictive value (PPV) evaluates the probability that a complex detected by the algorithm is actually a QRS complex. These two indicators are calculated using equations $Se = \frac{TP}{TP+FN}.100\%$ and $PPV = \frac{TP}{TP+FP}.100\%$, respectively:

Evaluation of the algorithm's performance on the three databases revealed a notable ability to detect QRS complexes, with results generally in line with expectations of standard evaluation criteria. Here is a detailed analysis of the results obtained:

Table 1. The results from the MIT-BIH database show an overall satisfactory performance. The sensitivity (Se) and positive predictive value (PPV) values are generally around 98-100%, which demonstrates the efficiency of the algorithm in detecting QRS complexes.

Table 2. The results on the NST database are homogeneous.

Table 3. Performance on this basis is remarkable, with sensitivity and PPV values.

Tab.1 : Performance result with MIT database

Records	Se(%)	PPV(%)	Records	TP	FP	FN	Se(%)	PPV(%)
100	98.66	100	201	90	0	11	98.90	100
101	98.61	100	202	53	0	1	98.14	100
102	100	100	203	100	0	9	91.74	100
103	98.58	100	205	89	0	1	98.88	100
104	90.9	100	207	82	5	15	84.53	94.25
105	98.80	100	208	102	0	9	91.89	100
106	95.74	100	209	93	0	1	98.93	100

Tab.2 : Performance result with NST database

Records	TP	FN	FP	Se(%)	PPV(%)
118e24	73	1	0	98.64	100
118e18	73	1	0	98.64	100
118e12	73	1	0	98.64	100
118e06	73	1	0	98.64	100
118e00	73	1	0	98.64	100
118e_6	73	1	0	98.64	100
119e24	65	6	0	91.54	100
119e18	65	6	0	91.54	100
119e12	65	6	0	91.54	100
119e06	65	6	0	91.54	100
119e00	65	6	0	91.54	100
119e_6	65	6	0	91.54	100

TAB.3: Performance result with long term ST database

Records	TP	FN	FP	Se(%)	PPV(%)
s20011	70	0	0	100	100
s20051	127	0	0	100	100
s20061	93	0	0	100	100
s20071	81	0	0	100	100
s20111	84	0	0	100	100
s20121	60	0	0	100	100
s20131	79	1	0	98.75	100
s20141	84	1	0	98.83	100
s20201	92	1	0	98.92	100
s20211	99	0	0	100	100

IV. Conclusion

The evaluation of the QRS complex detector based on the Pan-Tompkins algorithm, applied to the MIT-BIH, NST, and long-term ST databases, demonstrated a robust ability to detect QRS complexes in various scenarios. The results highlight exceptional performance on the MIT-BIH database, with sensitivities (Se) and positive predictive values (PPV) around 98-100%, although reduced sensitivity was observed in some noisy recordings. The NST database confirmed the resilience of the algorithm to different noise levels (SNR from 24 to -6 dB), with a consistent performance of Se = 98.64% and PPV = 100%, except for record 119, where some decreases were noted. Finally, the long-term ST database validated the stability of the algorithm in prolonged and complex recordings, with perfect scores in several cases and slight variations due to specific ST-T signal anomalies. Overall, these results confirm the effectiveness and reliability of the algorithm in varied environments, while suggesting avenues for improvement to better manage extreme and prolonged signaling conditions.

References

[1] M. Benosman, F. Bereksi-Reguig, et E. G. Salerud, « Strong real-time QRS complex detection », *J. Mech. Med. Biol.*, vol. 17, n° 08, p. 1750111, 2017.

[2] R. Harikumaret S. Shivappriya, « Analysis of QRS detection algorithm for cardiac abnormalities–A review », *Int. J. Soft Comput. Eng.*, vol. 1, n° 5, p. 80‑88, 2011.

[3] B.-U. Kohler, C. Hennig, et R. Orglmeister, « The principles of software QRS detection », *IEEE Eng. Med. Biol. Mag.*, vol. 21, n° 1, p. 42‑57, 2002.

[4] V. Gupta, M. Mittal, V. Mittal, et N. K. Saxena, « BP signal analysis using emerging techniques and its validation using ECG signal », *Sens. Imaging*, vol. 22, n° 1, p. 25, 2021.

[5] J. Pan et W. J. Tompkins, « A real-time QRS detection algorithm », *IEEE Trans. Biomed. Eng.*, n° 3, p. 230‑236, 1985.

[6] A. L. Goldberger *et al.*, « PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals », *circulation*, vol. 101, n° 23, p. e215‑e220, 2000.

[7] G. B. Moody et R. G. Mark, « The impact of the MIT-BIH arrhythmia database », *IEEE Eng. Med. Biol. Mag.*, vol. 20, n° 3, p. 45‑50, 2001.

[8] F. Jager*et al.*, « Long-term ST database: a reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia », *Med. Biol. Eng. Comput.*, vol. 41, p. 172‑182, 2003.

Improving and Numerical Simulation of Highly Stable and Efficient Perovskite Solar Cells FAPbI3-Based Using Fluorinated Spiro-OMeTAD hole transport layers.

Mohammed Bouladame
Department of Science and Technology, Mohamed Khider University,
Laboratory of Metallic and Semiconducting Materials (LMSM),
Biskra, Algeria.
mohammed.bouladame@univ-biskra.dz

Souad Tobbeche
Department of Science and Technology, Mohamed Khider University,
Laboratory of Metallic and Semiconducting Materials (LMSM),
Biskra, Algeria
s.tobbeche@univ-biskra.dz

***Abstract*— In this paper, the investigation reported on research that improved efficiency in perovskite solar cells based on FAPbI3 perovskites. One of those, a Hole-transporting layer (HTL)-structured perovskite solar cell between a formamidinium lead iodide active layer and back contact (gold), presented great promise as light harvesters for thin-film photovoltaics. Its increased stability and high VOC. The three-hole transport layers (HTL) are then tested on the Silvaco TCAD software simulator by replacing the most commonly employed Spiro-OMeTAD with two fluorinated isomeric analogs of the popular hole-transport material (HTM) Spiro-OMeTAD for HTLs and analyzing their effect on device performance. In this work, we improved different aspects affecting the device performance and studied thicknesses, doping concentrations, and carrier mobility for the absorber layer, ETL, and HTL, respectively. The structure with Spiro-mF as HTM, Tio2 as ETL, and FAPbI3 as active layer gave the most optimized parameters, which demonstrated a power conversion efficiency (PCE) of 29.95%. With an open-circuit voltage (Voc) of 1.192 V, short-circuit current density (Jsc) of 28.27 mA·cm−2, and fill factor (FF) at 88.81% using this configuration. The outcomes of this contribution can offer promising future research leads for the Community in the fabrication of stability-effective, high-efficiency perovskite solar cells.**

Keywords*— *Perovskite, PSCs, HTL, doping density, carrier mobility, Silvaco Tcad.

I. INTRODUCTION

Since the milestone proof-of-concept demonstration of perovskite solar cells (PSCs) in 2009, this technology has attracted much attention as an effective strategy for transitioning to renewable energy sources — a key aspect of mitigating climate change and reducing fossil fuel consumption worldwide. The non-linear use of solar energy available to almost all households arises as a promising alternative source of household energy, which can sustain more and more global needs for power across the globe with positive environmental ratings. The commercial industry surrounding perovskite solar cells is significant due to their promise of high efficiencies at low manufacturing costs that could be broadly applied in a variety of photovoltaic technologies. PSCs, known for their low-cost solution processibility and excellent optoelectronic properties (e.g., tunable bandgap, high absorption coefficient, low recombination rate, and high mobility of charge carriers), have been researched rapidly over the last few years. In less than a decade, the power conversion efficiency (PCE) of single-junction PSCs has increased from 3% to a certified value of 25.5%, representing the highest for thin-film photovoltaics. Additionally, Regular (n–i–p) structures are used in the PSCs and have the highest efficiency among all kinds of devices. As a result, PSCs are highly expected as the next-generation photovoltaic technology. The positive role of compositional engineering in realizing highly eco-friendly and stable PSCs. A combination of both bulk and computational research has been carried out to deposit methylammonium lead triiodide MAPbI 3 with formamidinium lead triiodide FAPbI 3 as an example of combining layers together5,7. As compared to MAPbI3, FAPbI3 is thermally more stable and has a bandgap closer to the Shockley–Queisser limit6, thus making it a good candidate for highly efficient as well as stable single-junction PSCs.

However, attempts to achieve high power conversion efficiencies (PCEs) >23% in perovskite solar cells (PSCs) have been confined to formamidinium-rich lead iodide (FAPbI3)-based compositions owing to their better bandgap-compared with methylammonium based or mixed-halide perovskites. A power conversion efficiency (PCE) >25% with Jsc >26 mA/cm2 for the mesostructure PSCs by fully exploiting the wide absorption spectrum of FAPbI3 was certified. Jeong et al. synthesized hydrophobic fluorinated analogs of 2,2′,7,7′-tetrakis[N, N-di(4-methoxyphenyl)amino]-9,9′-spirobifluorene (Spiro-OMeTAD) as hole-transporting materials leading to a favorable shifting of the electronic state for hole extraction in perovskite solar cells. One device employed an already certified power conversion efficiency of up to 24.8%; a similar article quoted the open-circuit voltage of this device as being in good agreement with the Shockley-Queisser limit [10]. During continuous operation under 50% relative humidity, more than 87% of the original power conversion efficiency (PCE) had been retained for over 500 h. To answer this question, it is clear that the PV performance of PSCs is governed by many different aspects including device structure, perovskite composition, charge transport layers

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

(CTLs), and electrode materials. CTLs also play a direct role in the performance and stability of the utility as suggested by the literature. ETL and HTL materials, which are organic as well as inorganic. Before selecting the CTLs, several factors have to be taken into account which are described as follows: a) As much as possible, CTLs should be highly transparent to ensure that there is little optical loss before the light arrives at the active perovskite layer. b) CTLs ought to have an energy band structure that allows one charge carrier while blocking the other. c) The CTLs must show excellent charge carrier mobility to recombine or extract the charge carriers from the perovskite layer into their corresponding electrode. d) The CTL should be chemically resistant to both the perovskite layer and the electrodes. There have been significant advancements in electron transport material (ETM) technologies over the last few years, and n-type metal oxides represented by TiO2 or SnO2 are currently the most commonly used materials for ETL. In contrast, the selection of hole transport material (HTM) is more complex and challenging than ETMs as well as accurately modeling the physics. The Hole Transport Material (HTM) in Organic-Inorganic Perovskite Solar Cells (PSCs) is essential to obtain high power conversion efficiency and long-term stability. While there are some published review articles on the topic of spiro-OMeTAD as HTM for PSCs, they tend to be either limited in scope or a bit dated. Consequently, in this review article, we also provide a critical examination of the doping materials and interface modifiers as well as recently introduced spiro-based strategies purported for establishing enhanced stability that has not been previously reported in the open literature to our knowledge. Thus, this review is proposed to fill the research void present in existing literature with the expectation that it will lead to a myriad of novel paradigms in spiro-based PSCs.

Here, we present a review that gives a detailed introduction to spiro-OMeTAD as HTM for PSCs, covering its background, evolution, and recent development. A unique approach, using Spiro-OMeTAD with varied degrees of fluorine doping as HTM, was taken to accomplish the goal. With fluorination, mobility, oxidation susceptibility, and ionization energy change. The Photoelectron yield spectroscopy data shown herein demonstrate that the HTM ionization energy increases concerning increasing fluorination level, which is equivalent to a downshift of the valence band maximum (VBM) position, from −4.97 eV (Spiro-OMeTAD reference) up to -5.19 eV relative to the vacuum level. In addition, the photoluminescence results also demonstrated a more efficient valence band offset (VBO) at room temperature for charge carrier separation. For all solar cells, the carriers reside at these interfaces and have to move through them to be collected by electrodes; this is a fact. A study of the experimental photovoltaic performance versus their numerical simulations using Silvaco Tcad for highly efficient single-junction solar cells based on formamidinium perovskites that can have a suitable bandgap (of 1.48 eV) to harvest the solar spectrum is described in this work. The responses to these questions were sought as they investigated the effects of changes in the hole transport material (HTM) type and thickness, assembly doping density, and carrier mobility state of the nanostructured absorber layer, ETL, and HTL on light harvesting efficiency of Perovskite solar cells.

II. DEVICE STRUCTURE AND SIMULATION METHOD

A. DEVICE STRUCTURE

First, An efficient PSC includes a perovskite absorber between positive (p) and negative (n) charge extraction semiconducting materials, known as p-type and n-type layers in this review. PSCs are commonly classified into three types: mesoporous n-i-p, planar n-i-p, and planar p-i-n (inverted planar) PSCs. PSCs utilizing n-i-p architecture (both mesoporous and planar) have achieved efficiencies of up to 23.44%, nearing 25.5% of p-i-n PSCs. However, non-inverted planar perovskite solar cells (n-i-p PSCs) are attracting attention for their simple manufacturing process, ability to be used on flexible substrates, flexibility in energy-band engineering, and potential for creating single-junction solar cells. The perovskite material's capability to enhance charge transport and optimize device structure independently has led to notable advancements in device performance and stability. When a single-junction solar cell is exposed to radiation from the sun (air mass (AM) 1.5G) and the surrounding environment (temperature T=300K). The simulated cell is set up with the following configuration:

Glass substrate/FTO/TiO2/FAPbI3/Spiro-OMeTAD/Au

The physical structure of the components of the Perovskite solar cell is shown in ***Fig.1***. Under a top-down proximity approach:

a. A layer of Fluorine-doped tin oxide (FTO) is used as a transparent conductive oxide in this solar cell to prevent contacts from leaking to charge carrier layers. FTO has a work function of 4.4 eV. The layer is 200 nm thick.

b. Titanium dioxide (TiO2) is effectively used as an electron transport layer (ETL) in photovoltaic devices, not least because of its non-toxic, stable, and affordable properties. A thickness of 150 nm is common for this simulation, as it helps to avoid chemical species migration and improve the efficiency of solar cells.

c. The main layer of the solar cell is the Perovskite layer, made up of a mix of organic-inorganic halides (PSCS), such as CH(NH2)2PbI3 Formamidinium lead triiodide (FAPbI3). FAPbI3 serves as the absorbent layer, allowing for the absorption of visible/ near-infrared light and enhancing thermal stability, with a thickness of 750 nm.

d. The most common material used as a hole transducer in Perovskite solar cells is the Spiro-OMeTAD polymer composition. Compared with the control device fabricated with doped Spiro-OMeTAD, both fluorinated HTM-based PSCs (Spiro-mF and Spiro-oF). The properties of this composition make it suitable for photovoltaic structures. It has the appropriate glass transition temperature, solubility, ionization potential, and transparency in the visible spectra range. The layer thickness is 270 nm. In contrast, Spiro-oF and Spiro-mF are included in the sub-model of the SCAPS 1D software package for the hole transporting layer in the perovskite solar cell setup. The absorption coefficient (alpha) is determined by the equation A/h υ √ (h υ - Eg), where A is a constant known as the band tailing parameter A=105. This assumption has been employed in numerous studies of perovskite-based solar cells. Eg represents the energy of the optical band gap, which varies based on the material's crystalline or amorphous nature. It is linked to the photon energy (h*neo) law model, which is associated with The refractive index (n) and

extinction coefficient (k) calculated using n=sqrt(epsilon), Alpha(α) = 4πk/lambda(λ).

e. In this study, we employed material like Gold on the back contact layer to enhance the efficiency of Perovskite solar cells, with a work function of 5.1 eV.

The refractive index and extinction coefficient (n, k) for these layers are extracted from FTO(40), TiO2 (41), FAPbI3 (42), and Spiro-OMeTAD (43) according to Fig. 2. The roll-off in efficiency for simulation and experimental results at lower illumination intensities results from the non-radiative recombination processes, including interfacial and Shockley–Read–Hall (SRH) recombination.

B. Simulation method

The simulation was carried out with the SILVACO TCAD environment using the Atlas module. It is a very useful two-dimensional component modeling simulator [25] with the ability to predict the electrical characteristics of most semiconductors. In addition to external electrical properties, the software provides information on internal distributions of variables such as current lines, electric fields, and potentials. All the properties of the materials involved in this solar cell are defined. The cell involves a network of thin semiconductor layers of FTO/TiO2/FAPbI3/Spiro-OMeTAD/Au simulated by TCAD Silvaco-Atlas. The complete structure of the device is schematically shown in Fig 1. In this study, the gold (Au) element is used here as the back contact, FAPbI3 as the absorbent layer, TiO2 as the buffer layer, undoped Spiro-OMeTAD: p as the hole transporting layer, and FTO as TCO. Good choice of mesh, models, and methods plays an important role in the accuracy of the simulation. All optical parameters used were extracted from previous reports [26, 27]. With a power density of 1000 W/m2(AM 1.5 illumination), is used as the light source in the simulation. This is achieved by numerically solving the Poisson equation with the electron and hole continuity equations in a finite number of points forming the mesh of the structure defined by the program and the drift-diffusion model:

Poisson's equation:

$$\frac{\partial^2\varphi}{\partial x^2} = \frac{q}{\varepsilon}(n - p) \qquad (1)$$

The continuity equations for electrons and holes:

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial x} + G - R \quad , \frac{\partial p}{\partial t} = -\frac{1}{q}\frac{\partial J_p}{\partial x} + G - R \qquad (2)$$

Jn and Jp are explained by the drift-diffusion model:

$$J_n = qD_n\frac{\partial n}{\partial x} - q\mu_n n\frac{\partial \varphi}{\partial x}, \; J_P = -qD_p\frac{\partial p}{\partial x} - q\mu_p p\frac{\partial \varphi}{\partial x} \qquad (3)$$

φ:Electric potential, ε: the permittivity, q: charge of an electron, n, p: the electron and hole concentrations, J_n, J_P: The electron and hole current densities, G, R: generation (recombination) rates for the electrons and holes, D_n, D_p: Electron and Hole constant diffusion, μ_n, μ_p: Electron and Hole mobility's.

Device simulation allows us to understand the basic physical processes in a solar cell, helping us confirm the best conditions for optimal performance. It is capable of simulating perovskite solar cell structures and calculating key characteristics including external quantum efficiency(QE), Current Density-Voltage (J–V) characteristics open-circuit voltages (Voc), short-circuit currents (Jsc), fill factor (FF), and power conversion efficiency (PCE).

For SILVACO TCAD simulations, the input parameters are taken from the literature and the results show that the simulated technological and physical parameter matches their real values excellently. Also, the Photovoltaic parameters of the experimental solar cell were compared with those of the simulation in Table 1. The interface defects at ETL/absorber and absorber/HTL interfaces are considered neutral and single. All the simulations are performed at a working temperature of 300 K using a series resistance of 1 W and under standard AM1.5G illumination.

III. RESULTS AND DISCUSSION

INVESTIGATING THE PHOTOVOLTAIC PROPERTIES OF PSCS DURING MANUFACTURING REQUIRES BOTH TIME AND FINANCIAL RESOURCES. - ONE OF THE ESSENTIAL ELEMENTS IN STUDYING THEIR PHOTOVOLTAIC PROPERTIES IS THE NUMERICAL DESIGN AND MODELING OF PEROVSKITE SOLAR CELL ARCHITECTURE. ANY SOLAR STRUCTURE CAN BE DESIGNED UTILIZING DIFFERENT INPUT MATERIAL PARAMETERS AND THE MODELING PROCESS. THEY PERMIT THE EXAMINATION OF THE PSC APPLICATION'S PERFORMANCE THROUGH ADJUSTMENTS TO THE MATERIAL'S PROPERTIES. WE CAN EASILY INTERPRET THE BEHAVIOR OF THE PSC DEVICE FROM THE SIMULATION RESULTS, AIDING IN OUR UNDERSTANDING OF THE SPECIFIED PSC RESTRICTIONS. THE SOFTWARE'S ACCURACY IS DEMONSTRATED IN TABLE 1, AS THE SIMULATED RESULTS FROM SILVACO TCAD CLOSELY ALIGN WITH THE EXPERIMENTAL RESULTS. THE FAPBI3-BASED SOLAR CELL STRUCTURES ARE DESIGNED AS FOLLOWS: FTO/TIO2/FAPBI3/SPIRO-OMETAD/AU. THESE STRUCTURES ARE ILLUSTRATED IN FIG 1. THE J-V CURVES OF N-I-P PLANAR PEROVSKITE SOLAR CELLS WITH A FAPBI3 ABSORBER LAYER ARE DISPLAYED IN FIG 3. IN THE SIMULATION STUDY. THE PHOTOVOLTAIC PARAMETERS OF THE SOLAR CELLS ARE LISTED IN TABLE 1. THE DEVICE UTILIZING THE PURE FAPBI3 PEROVSKITE LAYER DEMONSTRATES AN OPEN-CIRCUIT VOLTAGE (VOC) OF 1.150 V, A SHORT-CIRCUIT CURRENT (JSC) OF 26.05 MA·CM−2, A FILL FACTOR (FF) OF 78.19%, AND A POWER CONVERSION EFFICIENCY (PCE) OF 23.44%. THE EQE SPECTRUM OF THE DEVICES WAS MEASURED FROM 200 TO 900 NM OVER THE AM 1.5 PHOTON FLUX, AS SHOWN IN FIG 4.

By measuring the current response at a zero voltage and the normalization of the current to the flux density of the incident light, it is possible to calculate the quantum efficiency of the flux density of the total monochromatic optical light. At the peak of the quantum efficiency, the wavelength varies within the spectrum of the solar cells. Also, one can express the relationship between the short-circuit current density and QE. The choice of buffer layers can also be characterized by the plot of QE, which is shown

in Fig. 4. The results suggest that perovskite solar cells with Spiro-OMeTAD and TiO2 show an increased spectral response that reaches a maximum of 95%, from the wavelength of 350 nm to 800 nm.

1. **Selection of appropriate HTL for PSC**

Various Spiro-based materials have been suggested as the hole-transport layer (HTL) for a perovskite solar cell (PSC) based on FAPbI3 in this study. The simulation process is first carried out for various HTLs, including Spiro-OMeTAD, Spiro-mF, and Spiro-oF. The thicknesses of the active layers for electron transport, hole transport, and absorber materials are assumed to be 150 nm, 270 nm, and 750 nm, respectively. Fig 2 displays the energy band diagram of the PSC and the alignment of energy bands when using different Spiro-based materials as the HTL. Energy band alignment controls the movement of photo-generated charge carriers, influencing the performance of the PSC(22). From Table 2, we can observe that Spiro-mF and Spiro-oF have the same Voc and Jsc values. Spiro-mF, however, exhibits a higher fill factor and increased efficiency of 24.82%. The greater efficiency is evident in Fig 5 and 6. The generated responses show generation rates across the device and the distribution of the electric field at the FAPbI3/HTLs interface. The generation rates for Spiro-mF and Spiro-oF are almost the same. The lower production rates of Spiro-OMeTAD result in decreased efficiency. Despite its low power output, the Spiro-mF possesses the highest electric field, which boosts its performance. The electric fields of Spiro-mF and Spiro-oF are essentially the same, leading to decreased recombination at the contact. Spiro-OMeTAD has a lower electric field, enabling a minority of electrons to migrate toward the HTL, resulting in recombination and decreased efficiency. Fig 5 and 6 show the current-voltage (I-V) and quantum efficiency (QE) curves of different hole transport layers (HTLs). The graphs show that the QE responses of each HTL are similar. In subsequent research, FF Spiro-mF was chosen as the HTL because of its effectiveness.

2. **Effect of absorber layer thickness**

The thickness absorber material has a crucial effect on the efficiency of the PSC device. Light absorption and charge transfer are accomplished by the absorber layer. As a result, optimizing the absorber material is critical for improving PSC performance. Fig 7 shows the thickness of absorber material varies in the range between 0.1 and 1.4 µm. The rise in charge carrier recombination owing to the escalation in dark saturation current leads to a slight decrement in the Voc (green line). Efficiency and Jsc increase until 1.2 µm and then get saturated. As absorber material thickness reduces, the electrons and holes combine quickly, thereby increasing the rate of recombination. The rise in absorber thickness gives ample space for an increase in the number of photons to be absorbed, thus increasing the Jsc, (blue line). Further, the FF also rises with increasing thickness. We observe a PSC efficiency of 25.60% at the optimum absorber thickness (1.2 µm).

3. **Effect of absorber layer Doping Density**

The acceptor concentration of absorber material varies between 1012 and 1018 per cm-3. It can be noticed from Fig. 8 that when the acceptor concentration of FAPbI3 increases efficiency, Voc, and FF increase. However, Jsc remains constant and decreases when doping concentration goes beyond 1× 1015 per cm-3. This is because high doping concentration decreases the depletion region and charge carrier mobility which leads to an increase in holes and electrons recombining and decreasing the chance of capturing electrons produced by photons. Voc increases on increasing acceptor concentration. The device saturation current increases lead to an increase in Voc. The efficiency increases from 22 to 27% as the absorber material's acceptor concentration varies between 1014 and 1018 per cm-3. The optimum value of acceptor concentration for the FAPbI3 layer is taken as~ 10^{17} per cm3. Columbia traps will be created by extremely high doping, which will increase recombination (57).

4. **Effect of carrier mobility absorber layer**

In this step, the relation of the mobility with the doping concentration of the absorber layer is investigated first based on the parameters of the initial cell in Table 1. Fig9(a-b-c-d) illustrates the effects of mobility on the device performance under various absorber doping concentration values (NA=10^{12}-5.10^{17} cm^{-3}). It can be seen that the optimal value of mobility (Figs. 9(a)) changes with the absorber doping concentration (µ=2.10-3-20 cm2 /V·s).] Fig 9 illustrates the effects of mobility on the cell performance at various doping concentrations of the absorber layer based on the optimized absorber thickness of 1200 nm obtained from Fig. 7. The *J*sc and FF both increase with mobility increasing at different doping concentrations, especially the *J*sc can be significantly augmented (Fig. 9(c)). The obvious influence of mobility on *J*sc can be observed when NA is higher than 1 × 1016 cm−3. With a certain mobility, say, µ = 5 cm2 /V·s, *J*sc first increases with the augment of doping concentration and then saturated significantly with NA further increasing to 5×1016 cm−3. This is because both *J*sc and recombination rate increase with the augment of doping concentration. Mobility can surmount the contradiction between high *J*sc and high recombination rate caused by the high doping concentration. Recombination can be weakened through good transportation performance of perovskite material with high mobility(67).The optimal value of mobility gradually decreases with NA increasing, thus lowering the negative effect of the doping concentration increment (see Fig. 9(a)). Moderate doping concentration (5×1015 cm−3 ≤ NA < 5×1017 cm−3) is beneficial to the improvement of efficiency for µ > 10 cm2 /V·s. This analysis presents that monotonically increasing the mobility of the absorber layer is not beneficial to the performance of the PSCs. Taking into account the influence of the thickness and doping concentration, an optimal value of mobility exists. The corresponding characteristic curve, *J*sc rises significantly from the initial 31.59 mA/cm2 to 33.30 mA/cm2 . The Voc of 0.90 V, FF of 82.38%, and PCE of 24.78% are obtained at this preliminary optimization of the absorption layer. The transportation parameter's matching between the absorber and the charge transport layer is important to avoid the excessive accumulation of the carriers at their interface during the photon-generated carriers transporting to the corresponding electrodes. Hence, after the absorber layer optimization, the carrier mobility and other parameters, such as doping concentration and thickness, of HTM and ETM should also be optimized.

5. **Effect of electron transport layer (ETL) thickness**

Furthermore, we will explore the impact of variations in the thickness of the TiO2 electron transport layer, serving as the window layer, on crucial features of photovoltaic devices. Meanwhile, the thickness of the FAPbI3 layer, carrier concentrations, and carrier mobility will be held constant at 1.2 µm, 1017 cm-3, and 5 cm2 /V·s, respectively.

Fig. 10 presents a contour map reflecting the simulated fundamental characteristics of a solar cell, illustrating how these properties change as the thickness of TiO2 (which ranges from 0.01 to 0.21 µm on the x-axis) and illustrate the variations in open-circuit voltage (Voc), the short-circuit current (Jsc), the fill factor (FF), the efficiency (PCE) at different points (which ranges on the y-axis) change. First the variations in open-circuit voltage (Voc) at different points. Specifically, as thickness increases from 0.01 to 0.21 µm, it drops from roughly 1.195 V to roughly 1.129 V, with the maximum Voc observed at approximately 0.01 µm thickness. The Jsc values follow a similar pattern to the Voc, decreasing from around 28.02 mA/cm2 to approximately 27.59 mA/cm2 as thickness increases from 0.01 to 0.21 µm. The peak Jsc is observed at approximately 0.01 µm in thickness.. the FF values exhibit multiple zones where the FF decreases from 82.44 to 77.11% as the thickness increases from 0.01 to 0.21 µm. An average FF value of approximately 80.79%, which attains a higher performance of around 82.24%. The PCE values exhibit a similar trend to the Voc, where PCE decreases from around 27.61% to about 24.04% as the thickness increases from 0.01 to 0.21 µm, an average value of approximately 26.18% is observed. The obtained results can be understood in the following manner: Moreover, the reduction in the Voc value is associated with an increase in thickness. This augmentation promotes the generation of electron–hole pairs across the junction by extending the charge collection length(71). Additionally, bandgap alignment plays a crucial role in enhancing the Voc value. A similar rationale can be applied to explain the behavior of Jsc values. This is attributed to an increase in the recombination process and leakage current, particularly noticeable as the thickness of TiO2 increases(65,66). The behavior of the fill factor (FF) can be explained by its inverse relationship with Voc values. The efficiency (η) values are predominantly influenced by the combined effects of Voc, Jsc, and FF values(68). Our findings suggest that maintaining the optimum thickness of TiO2 is 0.03µm, decreasing with thickness, Under these conditions, as shown in Fig. 10, the thickness 0.03 µm, the open-circuit voltage (Voc) of 1.192 V, and the short-circuit current (Jsc) of 27.92 mA/cm2 the fill factor (FF) of 82.24%, and the efficiency (PCE) of 27.38%.

6. **Effect of ETL doping density**

Fig. 11 depicts a contour plot demonstrating variations in critical solar cell characteristics, where the carrier concentration ranges from 1015 to 1019 cm−3 on the x-axis and the device photovoltaic parameters(Voc, Jsc, FF, and PCE), are the dependent variable.

The study of Fig. 11 demonstrates that the open-circuit voltage (Voc) varies across distinct regions. More precisely, with Voc increasing from roughly 1.168 V to roughly 1.199 V carrier concentration rose from 1015 to 1019 cm−3, while the highest Voc reached a carrier concentration of approximately 5.1018 cm−3. When carrier concentrations are more than 1018 cm−3, the Voc value is saturated for each given carrier concentration value. According to Figure, the Jsc value remained constant at around 28.02 mA/cm2 for carrier concentrations more than 1015 cm−3 and less than 1018 cm-3, with Jsc decreasing from approximately 28.02 mA/cm2 to approximately 28.00 mA/cm2 as carrier concentration increases from 1018 to 1019 cm-3 and the highest Jsc attained at carrier concentration of approximately 1018 cm-3. The Jsc value improves for each carrier concentration value when the carrier concentration is less than 1019 cm−3. The FF values follow the same pattern as the Voc, with FF increasing from approximately 73.48% to approximately 82.71% as carrier concentration increases from 1015 to 1019 cm-3 and the highest FF attained at carrier concentration of approximately 1019 cm-3. In contrast, the FF reached its maximum value for levels of carrier concentration greater than 1019 cm−3. The PCE values have the same form as the Voc and FF, which increase from about 24.04% to about 27.76% as the carrier concentration increases from 1015 to 1019 cm-3, while the highest achieved at carrier concentration around 1019 cm-3. The value improves for every individual carrier concentration value if the carrier concentration is below 1018 cm−3. The data clearly show that increasing the donor concentrations of the TiO2 increases the Voc value as the carrier concentration increases. This influences the charge carrier generation rate and minority carrier diffusion length. The flow of electrons in the TiO2 layer and holes in the Spiro-mF layer influences the photogenerated current significantly (65). The full built-in voltage (Vbi) and the depletion width (Wd), primarily determined on the p-Spiro thin film side, are directly associated with this current64. For each given carrier concentration below 1020 cm−3, the Voc value drops as the TiO2 carrier concentration increases. This is due to the longer lifetime and larger charge collection channel, which results in the enhanced photogenerated current while concurrently decreasing leakage current and recombination rates as PSC thickness increases 71. These observations are corroborated by the fact that raising the carrier concentration above 1018 cm−3 results in a drop in Jsc value. Furthermore, the decrease in Jsc is due to the window layer predominantly functioning as a conducting channel for electrons rather than contributing considerably to solar spectrum absorption73. The efficiency (PCE) values provide a complete view of performance through the combination of Voc, Jsc, and FF, with its value predominantly following Voc and FF trends. As shown in Fig. 11. comparison, the optimal parameters for TiO2 are a thickness of 0.03 µm and a carrier concentration of less than 1019 cm−3. With a Voc value of 0.79 V, a Jsc value of 41.2 mA/cm2, and an FF value of 85.8%, the maximum efficiency was attained under these conditions. Our results show that perovskite solar cell devices based on The best values were obtained with donner densities of 5×1018 cm−3, and comprised a Jsc of 28.00 mA/cm2, a Voc of 1.199 V, an FF of 82.69%, and an efficiency (PCE) of 27.77%.

7. **Effect of electron mobility of ETL layer**

The importance of the high doping of the TiO2 ETM to match the HTM with high mobility was also investigated systematically in Ref [46]. The critical value of electron

mobility (μn), i.e., when the charge transportation is easier, depends on the doping concentration of a layer, the electron mobility was varied from 10-4 cm2 /Vs to 100 cm2 /Vs with different carrier concentrations (1015-5×1018 cm-3), The performance is illustrated in Fig 12. (Fig.12 a and b) It can be seen that The JSC and Voc were constant with increasing electron mobility and carrier concentration. After increasing the concentration by more than 1015 cm−3 and improves carrier mobility in the range 10− 4 to 1 cm2 /vs leading to the conclusion that higher mobility contributes towards the improvement of carrier transportation and suppression of recombination. (Fig.12 c and d) the FF and PCE were increased. After increasing the carrier concentration and electron mobility by more than 1017 cm−3 and 1 cm2 /Vs, respectively. The FF and PCE became stable, so they became unaltered by either carrier concentration or carrier mobility. The higher values of carrier mobility beyond 1 cm2 /vs do not affect Jsc as they lead to higher carrier diffusion length, which is compatible with that in Ref [54]. Interestingly, increasing the mobility decreases the resistivity of the ETL such that the series resistance decreases. Our results show that perovskite solar cell devices based on The best values were obtained with donor densities and electron mobility of 5×1018 cm−3 and 1 cm2 /Vs, respectively, comprised a Jsc of 28.02 mA/cm2, a Voc of 1.195 V, an FF of 82.43%, and an efficiency (PCE) of 27.61%.

8. ***Effect of hole transport layer (HTL) thickness***

Proper thickness of HTL plays a crucial role in shaping the high performance of the PSC. The best HTL is the one that reduces electrical resistance and recombination setbacks, which further aids in boosting the PSC's performance. The thickness of Spiro-mF material in the designed PSC structure is varied in the range of 0.05–0.7 μm. The thickness of Spiro-mF shows an insignificant effect on the PSC's performance, as shown in Fig. 13. However, the Jsc and Voc increased until the value 0.25 μm became stable. When Increasing the thickness of Spiro-mF leads to a reduction in both FF and PCE values. Spiro-mF plays an active role in transporting the hole to the counter electrode owing to the making of an energy cliff near the interface's HTL/absorber. Thereby, the thickness of HTL is taken to be 25 nm for the optimized device structure, comprised of a Jsc of 28.02 mA/cm2, a Voc of 1.195 V, an FF of 83.24%, and an efficiency (PCE) of 27.88%.

9. **Effect of HTL layer doping density**

Hole Transport Layer doping density controls PSC performance by changing the electric field intensity at the HTL/absorber interface. An increase in an electric field leads to a greater separation of electron–hole pairs and thus an increase in efficiency. Here, the acceptor density of HTL varies in between the range of 1015–1019 per cm3. Fig. 14 illustrates that increasing the acceptor concentration of Spiro-mF from 1× 1015 per cm3 to 1×1019 per cm3 has no significant effect on the performance of PSC. According to Fig. 14, the Voc value remained constant at around 1.195 V, and a small change in Jsc average of 28.06 mA/cm2 for carrier concentrations more than 1015 cm−3. However, we observed that there is a significant change in efficiency from 25.33% to 29.71% and FF from 75.68 to 88.47%. As we increase the doping concentration, charge carrier separation is increased and then the perovskite solar cell performances are also increased. The optimized result for the acceptor concentration of Spiro-mF is considered 1× 1018 per cm3. Our modeling outcomes indicate that the ideal values for Spiro-mF acceptor densities of 1018 cm-3, and comprised a Jsc of 28.10 mA/cm2, a Voc of 1.192 V, an FF of 88.44%, and an efficiency (PCE) of 29.65%.

10. **Effect of hole mobility of HTL layer**

Besides, we have investigated the hole transport properties effect like other layers. While maintaining the previous optimum properties, of the most operating HTM such as the Spiro- layer, on the device efficiency. The effects of hole mobility on the performance of perovskite solar cells under various doping density values, when the hole mobility of HTM increases from 10−4 cm2 /V·s to 20 cm2 /V·s, as shown in Fig.15. (Fig.15 d) the efficiency of perovskite solar cells increases gradually, little change of the efficiency can be seen for μp > 10-2 cm2 /V·s because the carriers' diffusion length is longer than the thickness of the HTM. When doping concentration is smaller than 1 × 1017 cm−3, PSCs' efficiency augments gradually with mobility increasing and reaches a saturation level at an HTM mobility of 20 cm2 /V·s. particularly noticeable as the doping concentration is 1 × 1018 cm−3, the effect of the mobility is very weak on the cell's efficiency, in our simulations(Fig.15 c), increasing the mobility from 1·10− 4 up to 1·10− 1 cm2 /V·s, it results in an improvement of the FF from 75.56 to 88.81. the strong influence of the HTL conductivity degradation in the device FF[64], while JSC and VOC remain unchanged shown in (Fig.15 a-b), This shows that the electron-hole pairs can easily transport at the interface and be collected by the electrodes. Moreover, the efficiency reaches a maximum saturation level of ∼30%, at hole mobility of 1 cm−2 v−1 s −1 with carrier concentration value 1018 cm-3. Our modeling outcomes indicate that the ideal values for Spiro-mF hole mobility in carrier concentration of 1 cm-2v-1s-1 and 1018 cm-3, respectively, and comprised a Jsc of 28.27 mA/cm2, a Voc of 1.192 V, an FF of 88.81%, and an efficiency (PCE) of 29.95%.

11. **Optimized result of I–V and EQE curve**

Table 3 presents a comparison of the optimized and experimental PSCs. It illustrates the design structure's capability to improve the PSC. This is possible by employing efficient CTL materials and parameters of each layer. Table 4 compares the optimized solar cell against the state-of-the-art solar cells based on FAPbI3 absorber material. The comparison clearly shows that the suggested solar cell is capable of producing superior outcomes, making it an appealing alternative.

The current density–voltage (*J*–V) characteristic curve of the cell with initial parameters is shown in Fig. 16 (curve green). Short-circuit current density (*J*sc) of 26.35 mA/cm2, open circuit voltage (Voc) of 1.164 V, fill factor (FF) of 80.90%, and power conversion efficiency (PCE) of 24.82% are obtained, and the same observation for quantum efficiency (EQE-λ) is shown in Fig. 17.

Based on the previous optimized structure, we further investigate the relation between carrier mobility and cell performance. Taking into account the influence of each layer on the performance of perovskite solar cells, the optimization process consists of three steps. The parameters of thickness, carrier mobility, and doping concentration are

optimized in the absorber layer, ETM, and HTM layers in sequence in the first three steps. Experimental research has shown that the mobility μ is independent of the charge carrier concentration of the perovskite absorber layer at lower concentrations (10^{16} cm^{-3}–10^{18} cm^{-3})[65-66].

TABLE I. BASIC INPUT PARAMETERS OF THE MATERIALS USED IN THE PSC.

Parameters	**FTO [49-51]**	**TiO_2 [45,24]**	**$FAPbI_3$ [24,48,49]**	**Spiro-OMeTAD [24,47,49]**	**Spiro-mF [24,54,56]**	**Spiro-mF [24,54,56]**
Thickness (nm)	200	50	750 *	270 *	270	270 *
Band gap Eg(eV)	3.5	.26	1.48*	2.94*	2.96*	2..96*
Electron affinity χ(eV)	4.0	.2	3.92*	2.03*	2.23*	2.23*
Dielectric permittivity ε_r	9	9	6.6	3.0	3.0	.0
CB effective density of states Nc(cm^{-3})	$2{\cdot}2.10^{18}$	$1.0{\cdot}10^{21}$	$1.2{\cdot}10^{19}$	$2.2{\cdot}10^{18}$	$1.0{\cdot}10^{19}$	$1.0{\cdot}10^{19}$
VB effective density of states Nv(cm^{-3})	$2{\cdot}2.10^{18}$	$2.0{\cdot}10^{20}$	$2.9{\cdot}10^{18}$	$1.8{\cdot}10^{19}$	$1.0{\cdot}10^{19}$	$1.0{\cdot}10^{19}$
Shallow donor density N_D (cm^{-3})	10^{15}	$\cdot10^{16}$	$.3{\cdot}10^{16}$	0	0	0
Shallow acceptor density N_A (cm^{-3})	0	0	$.3{\cdot}10^{16}$	$1.3{\cdot}10^{18}$	10^{16}	10^{16}
Electron mobility μ_n(cm2/Vs)	20	0	.7	6.59.10-3*	7.47.10-3*	47.10-3*
Hole mobility μ_p(cm2/Vs)	10	0	.8	6.59.10-3*	7.47.10-3*	47.10-3*
Electron lifetime τn (S)	10−7	10−7	10−6	10−7	10−7	10−7
Hole lifetime τp (S)	10−7	10−7	10−6	10−7	10−7	10−7

*** Experimentally determined**

TABLE II. BASIC INPUT PARAMETERS OF THE MATERIALS USED IN THE PSC.

	CELL CONFIGURATION (HTL'S)	**PCE (%)**	**VOC(V)**	**JSC (MA/CM^2)**	**FF (%)**
EXPERIMENTAL [10]	FTO/TIO2/FAPBI3/SPIRO-OMETAD/AU	23.44	1.152	26.04	78.13
	FTO/TIO2/FAPBI3/SPIRO-MF/AU	24.82	1.164	26.35	80.90
	FTO/TIO2/FAPBI3/SPIRO-OF/AU	24.50	1.161	26.34	80.15
SIMULATION	FTO/TIO2/FAPBI3/SPIRO-OMETAD/AU	23.44	1.150	26.05	78.19
	FTO/TIO2/FAPBI3/SPIRO-MF/AU	24.82	1.164	26.35	80.90
	FTO/TIO2/FAPBI3/SPIRO-OF/AU	24.54	1.162	26.34	80.15

TABLE III. **COMPARISON BETWEEN OPTIMIZED PEROVSKITE SOLAR CELL AND CONVENTIONAL STRUCTURES.**

DESIGN PARAMETERS	**CONVENTIONAL PEROVSKITE SOLAR CELL**	**OPTIMIZED PEROVSKITE SOLAR CELL**
ETL MATERIAL	TIO2	TIO2
DOPPING CONCENTRATION (CM− 2)	5×10^{16}	5×10^{18}
THICKNESS (NM)	68	50
ELECTRON MOBILITY μ_N(CM2/VS)	20	1
HTL MATERIAL	SPIRO-OMETAD	SPIRO-MF
DOPPING CONCENTRATION (CM− 2)	10^{16}	10^{18}
THICKNESS (NM)	270	250
ELECTRON MOBILITY μ_P(CM2/VS)	6.59×10-3	1

PEROVSKITE MATERIAL	FAPBI3	FAPBI3
DOPPING CONCENTRATION (CM− 2)	1.3×10^{16}	10^{17}
THICKNESS (NM)	750	1200
ELECTRON MOBILITY μ(CM2/VS)	1.8	5
Output parameters		
*J*SC (MA/CM2)	26.04	28.27
V_{OC} (V)	1.152	1.192
FF (%)	78.13	88.81
PCE (%)	23.44	29.95

TABLE IV. **BENCHMARKING TABLE.**

DEVICE STRUCTURE (EXPERIMENTAL	**PCE (%)**	**REF**
TIO2/FAPBI3/PTAA/Au	21.6	[58]
SNO2/FAPBI3/SPIRO-OMETAD/AU	16.68	[59]
SNO2/CS0.1(GABA0.075-FA0.925)0.9PBI3/SPIRO-OMETAD/AU	23.71	[59]
TIO2/FAPBI3/MEO-2PACZ	22.13	[60]
TIO2/FAPBI3/SPIRO-OMETAD/AU	22.25	[61]
ZN0.8MG0.2O/(FAPBI3)0.85/(MAPBBR3)0.15/CU2O	22.17	[62]
TIO2/CS2BICUI6/SPIRO-OMETAD/AU	22.02	[63]
TIO2/FAPBI3/SPIRO-MF/AU	29.95	**THIS WORK**

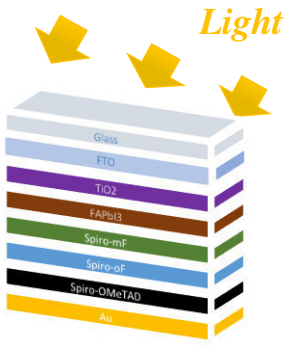


Fig 1. **Planar device structure**

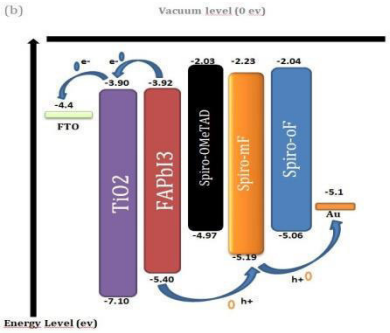


Fig. 2. **Energy band diagram of FTO/ ETM/ Perovskite/HTM's/Au structure.**

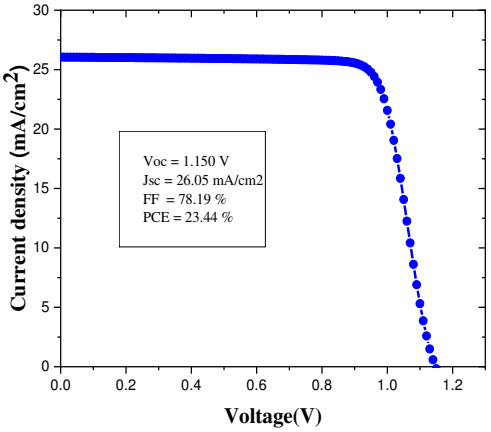


Fig. 3. **J-V characteristics of the n-i-p PSC.**

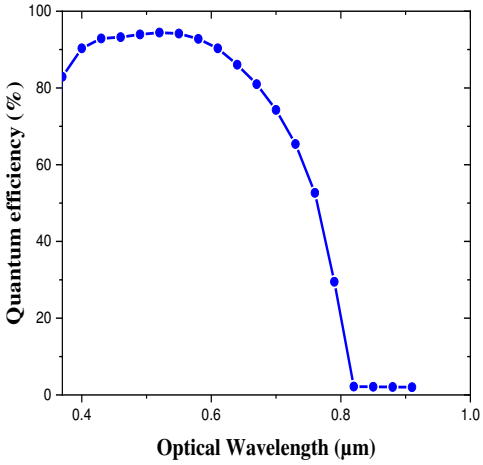


Fig. 4. **The quantum efficiency of the n-i-p PSC.**

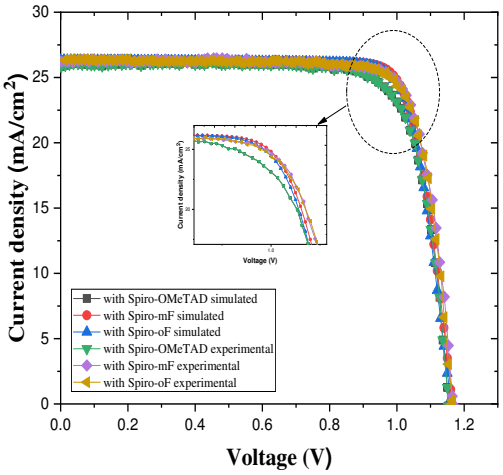


Fig. 5. **Effect of different HTL materials on J-V characteristic**

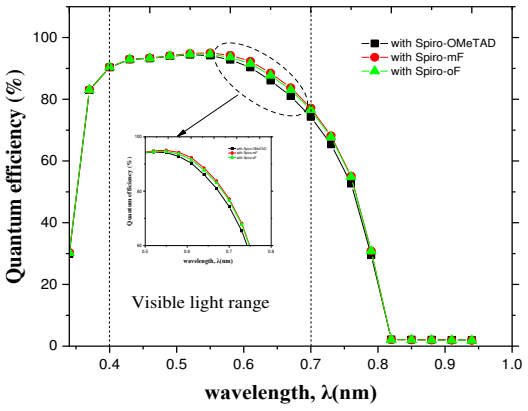


Fig. 6. **Effect of the different HTL materials on quantum efficiency**

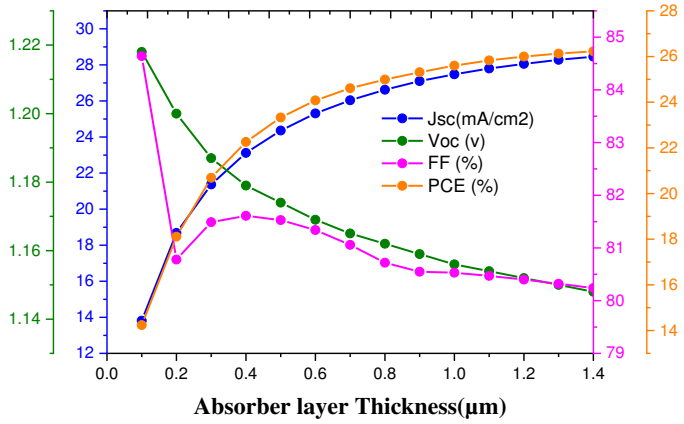


Fig. 7. **Output parameters of PSC by varying active layer thickness**

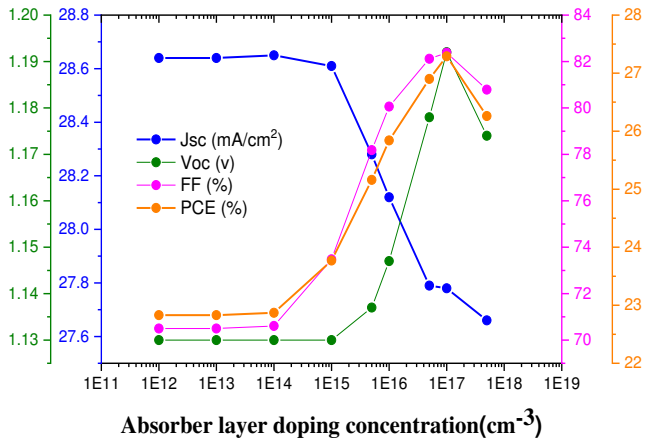
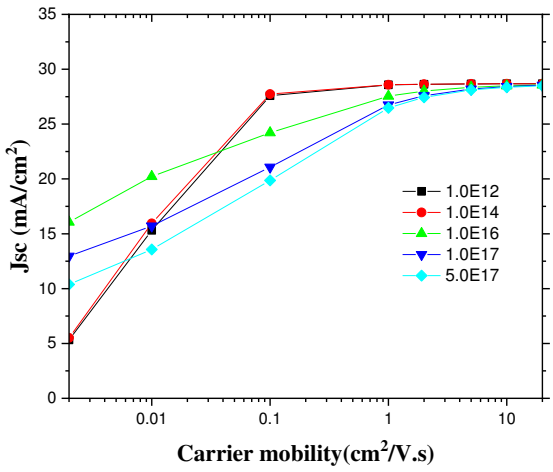
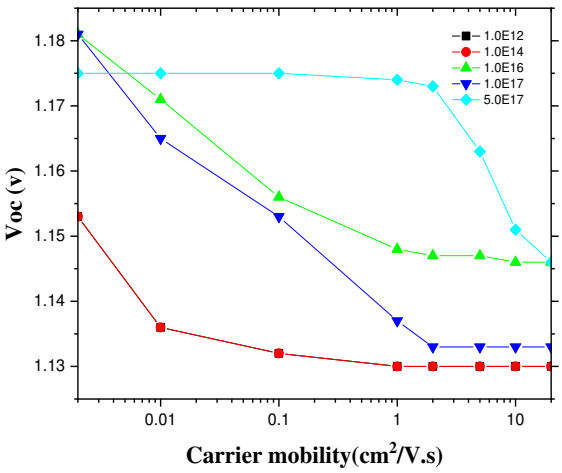


Fig. 8. **Output parameters of PSC by changing acceptor concentration (N_A) of the active layer.**

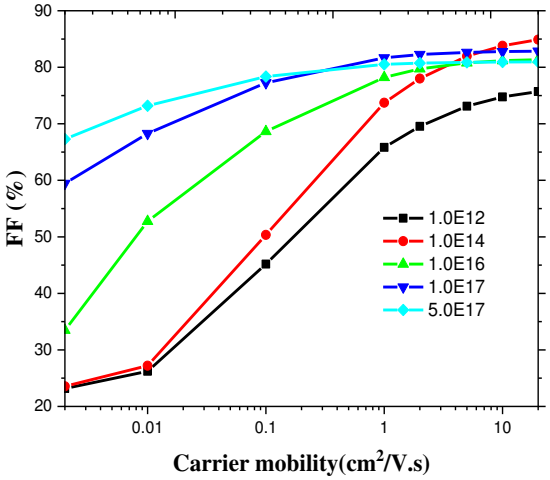
a)



b)



c)



d)

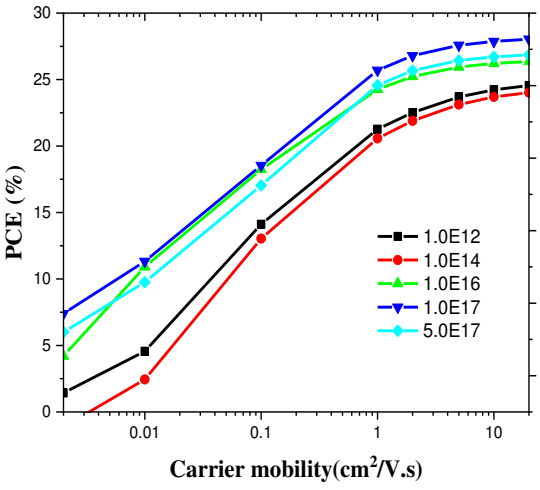


Fig. 9 (a-b-c-d). **Output parameters of PSC by varying carrier mobility (µ) of the active layer**

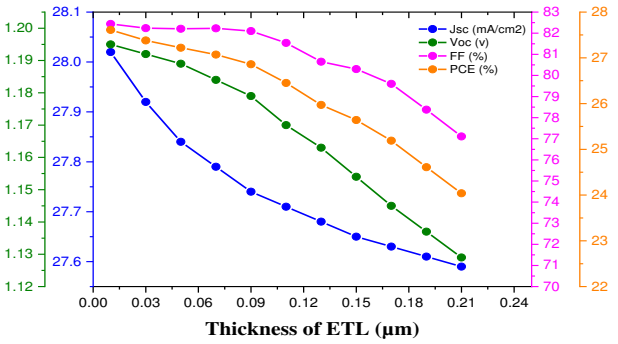
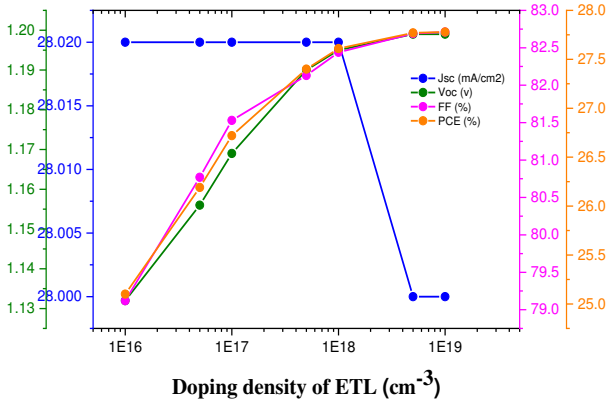
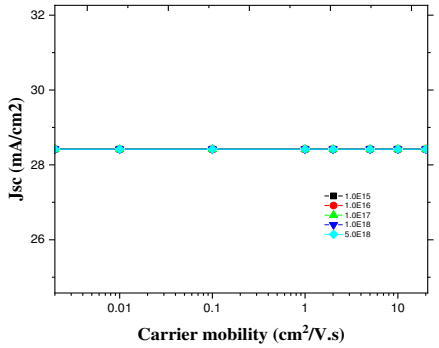


Fig.10. **Impact of thickness of ETL on device photovoltaic parameters.**

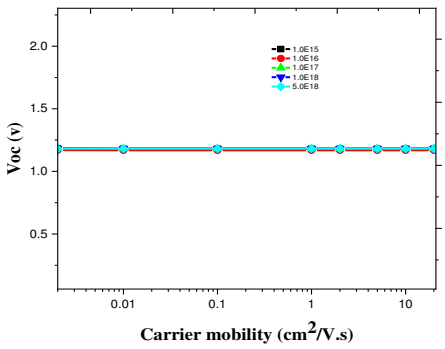


FFig.11. **Impact of doping density of ETL layer on device photovoltaic parameters.**

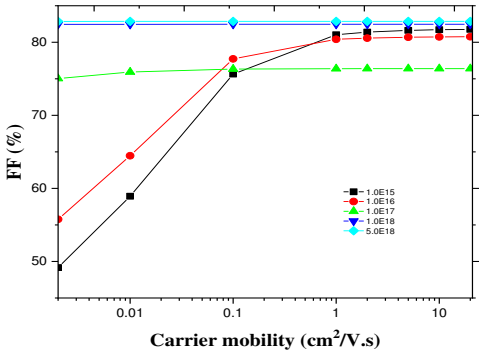
a)



b)



c)



d)

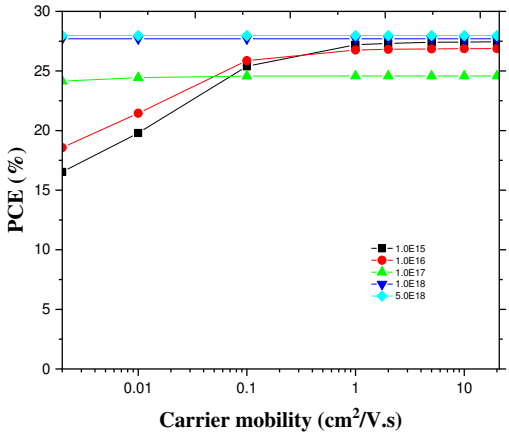


Fig.12. **Impact of electron mobility of ETL layer on device photovoltaic parameters.**

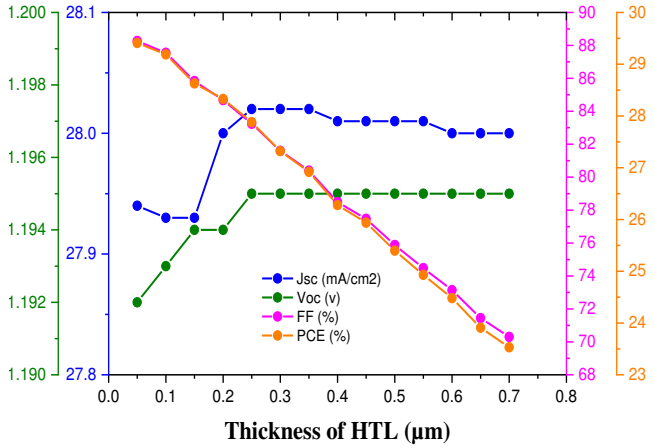


Fig.13. **Impact of different thickness of HTL layer on device photovoltaic parameters.**

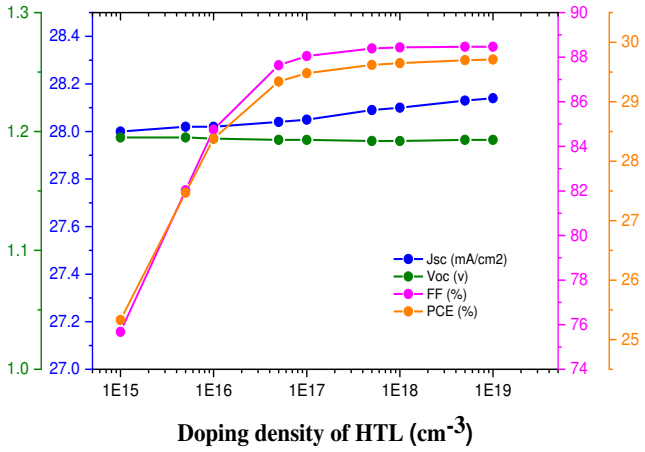
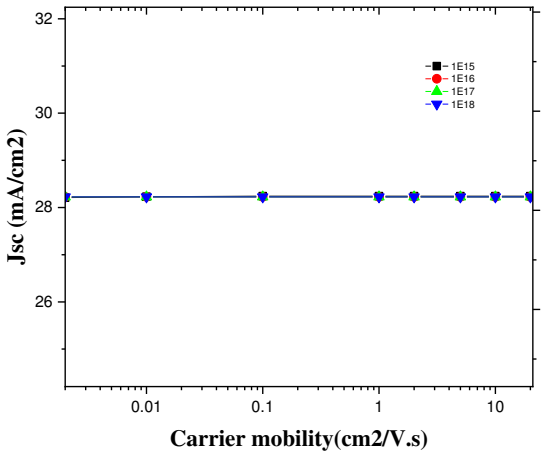
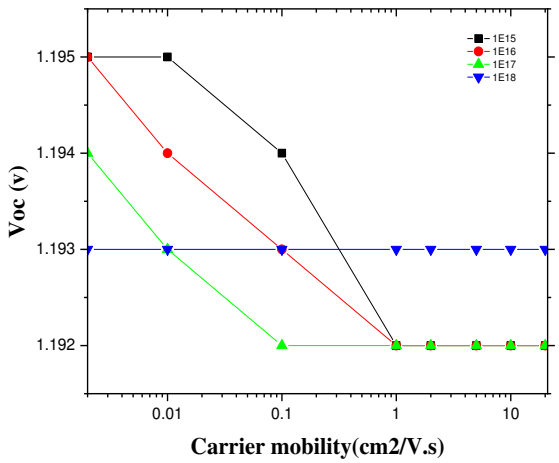


Fig.14. **Impact of doping density of HTL layer on device photovoltaic parameters.**

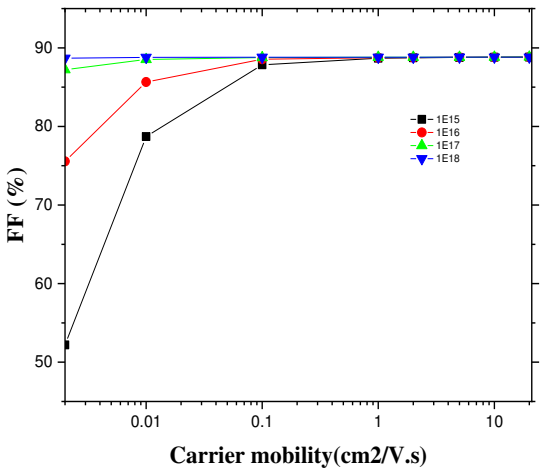
a)



b)



c)



d)

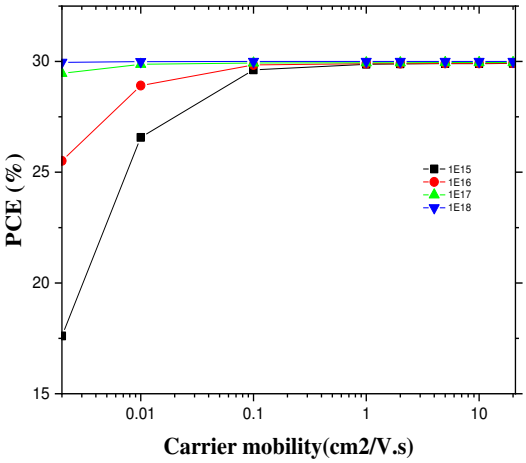


Fig.15. **Impact of hole mobility of HTL layer on device photovoltaic parameters**

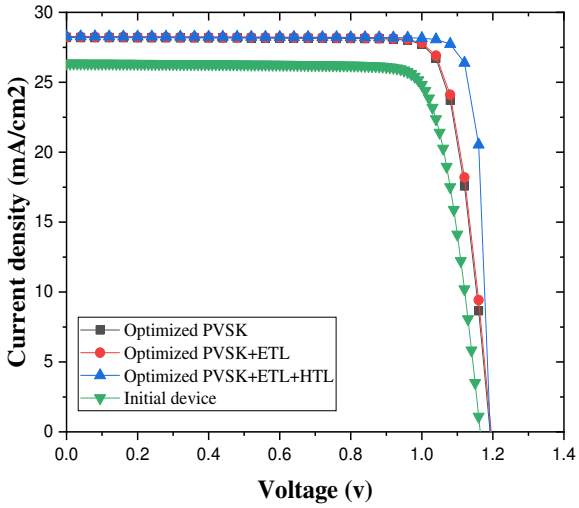


Fig.16. **J-V characteristics comparison of initial and optimized devices.**

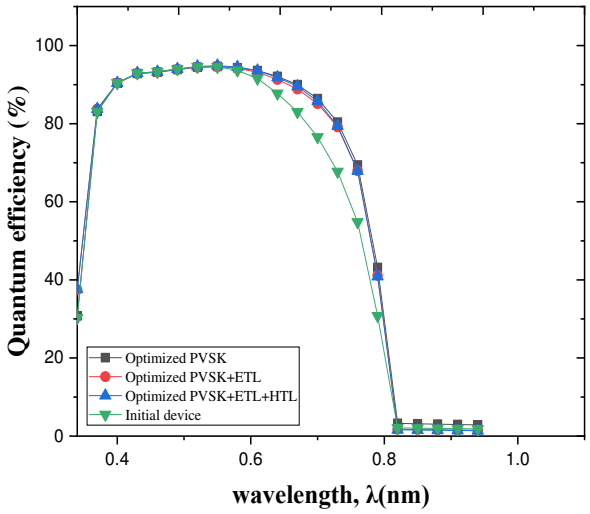


Fig.17. **The quantum efficiency comparison of initial and optimized devices.**

CONCLUSION

This paper uses a planar configuration of FAPbI3-based PSCs with TiO2 as an ETL layer. We employ two extracts of Spiro-OMeTAD HTMs that were studied using the Silvaco Tcad simulation program. It was found that PSC with the FTO/TiO2/FAPbI3/ Spiro-mF/Au configuration attained an efficiency of 24.82%. Thickness, doping density, and carrier mobility of the absorber, ETL, and HTL were varied to examine their influence on Voc, Jsc, fill factor, and efficiency. The effects of the carrier mobility and doping density on the PSC's performance are studied systematically with the device simulation. A similar parabolic relationship between the mobility of the absorber and the efficiency of PSC to that for the perovskite solar cell is observed and there also exists an optimal mobility of the absorber. A doping density of 1017 cm-3 is also beneficial to the PSCs' performance when the absorber is thick and has high crystalline quality. The improved carrier mobility of the absorber, ETL, and the HTL appropriately is favorable to the reduction of the recombination rate and the improvement of the carrier transportation, which can avoid the joint reduction of *J*sc, and Voc when the absorber layer thickness increases and these can improve the cell's efficiency. The doping of the carrier transport layer has a significant role in improving cell performance. Finally, we obtain the *J*sc of 28.27 mA/cm2, Voc of 1.192 V, FF of 88.81%, and PCE of 29.95% in our cell structure with an absorber thickness of 1200 nm, a high mobility of 5 cm2 /V·s, and doping density of 1017 cm-3. Using the hot casting method to realize high-quality perovskite films may be a promising method to learn the high-efficiency PSCs with thick absorbers and high stability.

REFERENCES

[1] Jiajia Suo , Bowen Yang , Jaeki Jeong ,Tiankai Zhang , Selina Olthof , Feng Gao , Michael Gr¨ atzel, Gerrit Boschloo , Anders Hagfeldt. Interfacial engineering from material to solvent: A mechanistic understanding on stabilizing α-formamidinium lead triiodide perovskite photovoltaics. Nano Energy 94 (2022) 106924.

[2] Shaikh, S.F.; Kwon, H.C.; Yang, W.; Mane, R.S.; Moon, J. Performance enhancement of mesoporous TiO2 -based perovskite solar cells by ZnS ultrathin-interfacial modification layer. J. Alloys Compd. 2018, 738, 405–414. [CrossRef]

[3] Eperon, G.E.; Burlakov, V.M.; Docampo, P.; Goriely, A.; Snaith, H.J. Morphological control for high performance, solutionprocessed planar heterojunction perovskite solar cells. Adv. Funct. Mater. 2014, 24, 151–157. [CrossRef]

[4] Liu, M.; Johnston, M.B.; Snaith, H.J. Efficient planar heterojunction perovskite solar cells by vapour deposition. Nature 2013, 501, 395–398. [CrossRef]

[5] Chang, S.H., Tseng, P.-C., Chiang, S.-E., Wu, J.-R., Chen, Y.-T., Chen, C.-J., Yuan, C.-T., Chen, S.-H., 2020. Structural, optical and excitonic properties of MAxCs1-xPb(IxBr 1–x)3 alloy thin films and their application in solar cells. Sol. Energy Mater. Sol. Cells. 210, 110478. https://doi.org/10.1016/j.solmat.2020.110478.

[6] Yaxin Wang, Xin Zhang, Zejiao Shi, Lixiu Zhang, Anran Yu, Yiqiang Zhan, and Liming Ding., 2022. Stabilizing α-phase FAPbI3 solar cells. Journal of Semiconductors 43, 040202. doi: 10.1088/1674-4926/43/4/040202.

[7] Jaeki Jeong, Minjin Kim, Jongdeuk Seo, Haizhou Lu, Paramvir Ahlawat, Aditya Mishra, Yingguo Yang , Michael A. Hope , Felix T. Eickemeyer , Maengsuk Kim , Yung Jin Yoon , In Woo Choi , Barbara Primera Darwich , Seung Ju Choi , Yimhyun Jo , Jun Hee Lee , Bright Walker , Shaik M. Zakeeruddin, Lyndon Emsley , Ursula Rothlisberger , Anders Hagfeldt, Dong Suk Kim, Michael Grätzel & Jin Young Kim. Pseudo-halide anion engineering for α-FAPbI3 perovskite solar cells. Nature | Vol 592, (2021).

[8] Mehrabian, M.; Norouzi, A.E.; Norouzzadeh, P. Numerical simulation of bilayer perovskite quantum dot solar cell with 18.55% efficiency. Opt. Quant. Electron. 2022, 54, 439. [CrossRef]

[9] Niu, T.; Chao, L.; Dong, X.; Li, F.; Chen, Y. Phase-pure α-FAPbI3 for perovskite solar cells. J. Phys. Chem. Lett. 2022, 13, 1845–1854. [CrossRef] [PubMed]

[10] Mingyu Jeong, In Woo Choi, Eun Min Go, Yongjoon Cho , Minjin Kim, Byongkyu Lee , Seonghun Jeong , Yimhyun Jo , Hye Won Choi , Jiyun Lee , Jin-Hyuk Bae, Sang Kyu Kwak, Dong Suk Kim, Changduk Yang. Stable perovskite solar cells with efficiency exceeding 24.8% and 0.3-V voltage loss, Science 369, 1615–1620 (2020).

[11] H.-J. Kim, H.-S. Kim, N.-G. Park, Adv. Energy Sust. Res. 2021, 2, 2000051.

[12] P. P. Boix, K. Nonomura, N. Mathews, S. G. Mhaisalkar, Mater Today. 2014, 17, 16.

[13] K. Deng, Q. Chen, Liang Li, Adv. Funct. Mater. 2020, 30, 2004209.

[14] S. Ravishankar, S. Gharibzadeh, C. R. Carmona, G. Grancini, Y. Lee, M. Ralaiarisoa, A. M. Asiri, N. Koch, J. Bisquert, M. K. Nazeeruddin. Influence of Charge Transport Layers on Open-Circuit Voltage and Hysteresis in Perovskite Solar Cells, Joule 2, 788–798,(2018).

[15] K. Deng, Q. Chen, Liang Li. Modification Engineering in SnO2 Electron Transport Layer toward Perovskite Solar Cells: Efficiency and Stability, Adv. Funct. Mater. 2020, 30, 2004209.

[16] S. Sun, T. Buonassisi, J.-P. Correa-Baena, State-of-the-art electron-selective contacts in perovskite solar cells, Adv. Mater. Interfac. 5 (2018) 1800408, https:// doi.org/10.1002/admi.201800408.

[17] M. Bescond, Semicond. Quantum transport in semiconductor nanowires, Nanowires 2015, 173.

[18] Laxmi Nakka, Yuanhang Cheng, Armin Gerhard Aberle, and Fen Lin. «Analytical Review of Spiro-OMeTAD Hole Transport Materials: Paths Toward Stable and Efficient Perovskite Solar Cells, Adv. Energy Sustainability Res. 2022, 2200045.

[19] Vinutha K.V, Naveen Kumar K. B, Tejas M K, Jai Kumar B. Natural Dye Sensitized Solar Cells Using Anthocyanin Pigment Of Strawberry As Sensitizers, Imperial Journal of Interdisciplinary Research (IJIR) Vol-2, Issue-10, 2016.

[20] Hiroyuki Fujiwara, Masato Kato, Masato Tamakoshi, Tetsuhiko Miyadera, Masayuki Chikamatsu, Optical Characteristics and Operational Principles of Hybrid Perovskite Solar Cells. Phys. Status Solidi A 2018, 1700730.

[21] Miha Filipič, Philipp Löper, Bjoern Niesen, Stefaan De Wolf, Janez Krč, Christophe Ballif, and Marko Topič. CH3NH3PbI3 perovskite / silicon tandem solar cells: characterization based optical simulations, | Vol. 23, No. 7 (2015)| DOI:10.1364/OE.23.00A263 | OPTICS EXPRESS A263

[22] Hossain, M. K.; Arnab, A. A.; Das, R. C.; Hossain, K. M.; Rubel, M. H. K.; Rahman, M. F.; Bencherif, H.; Emetere, M. E.; Mohammed, M. K. A.; Pandey, R. Combined DFT, SCAPS-1D, and WxAMPS Frameworks for Design Optimization of Efficient Cs2BiAgI6-Based Perovskite Solar Cells with Different Charge Transport Layers. RSC Adv. 2022, 12 (54), 34850−34873.

[23] P. B. Johnson and R. W. Christy. Optical Constants of the Noble Metals, PHYSICA L RE VIEW B VOLUME 6, NUMBER 12 (4370-4379) 1972.

[24] M. He, D. Zheng, M. Wang, C. Lin and Z. Lin, J. High Efficiency Perovskite Solar Cells: From Complex Nanostructure to Planar Heterojunction , Mater. Chem. A, 2014, DOI https://doi.org/10.1039/C3TA14160H.

[25] Teddy Salim, Shuangyong Sun, Yuichiro Abe, Anurag Krishna, Andrew C. Grimsdale and Yeng Ming Lam. Perovskite-based solar cells: impact of morphology and device architecture on device performance, J. Mater. Chem. A, 2015, 3, 8943–8969.

[26] Mohamed Alla , Vishesh Manjunath , Najwa Chawki , Diwakar Singh , Subhash C. Yadav , Mustapha Rouchdi , Fares Boubker. Optimized CH3NH3PbI3-XClX based perovskite solar cell with theoretical efficiency exceeding 30%, Optical Materials 124 (2022) 112044.

[27] Nam Joong Jeon, Hyejin Na, Eui Hyuk Jung , Tae-Youl Yang , Yong Guk Lee , Geunjin Kim , Hee-Won Shin , Sang Il Seok , Jaemin Lee and Jangwon Seo. A fluorene-terminated hole-transporting material for highly efficient and stable perovskite solar cells, Nature Energy | VOL 3 (2018) 682–689.

[28] Taesu Kim1,2, Jae-Han Kim2,3, Tae Eui Kang1,2, Changyeon Lee1,2, Hyunbum Kang1,2, Minkwan Shin4, Cheng Wang5, Biwu Ma6, Unyong Jeong4, Taek-Soo Kim2,3 & Bumjoon J. Kim. Flexible, highly efficient all-polymer solar cells, NATURE COMMUNICATIONS | 6:8547 (2015)| DOI: 10.1038.

[29] Laura Calil, Samrana Kazim, Michael Gr-tzel, and Shahzada Ahmad. Hole-Transport Materials for Perovskite Solar Cells, Angew. Chem. Int. Ed. 2016, 55, 14522 – 14545.

[30] S. Karthick, et al. Solar Energy 205 (2020) 349–357, Experimental and SCAPS simulated formamidinium perovskite solar cells: A comparison of device performance.

[31] S. Karthicka , S. Velumania , J. Bouclé. Experimental and SCAPS simulated formamidinium perovskite solar cells: A comparison of device performance, Solar Energy 205 (2020) 349–357.

[32] K. B. Nine, M. N. H. Shazon, S. A. Mahmood, Performance evaluation and comparative analysis of a highly efficient FAPbI 3-based perovskite solar cell. Journal of the Optical Society of America B pp. 2996-3004 (2020).

[33] Faiza Azri, Afak Meftah, Nouredine Sengouga, Amjad Meftah. Electron and hole transport layers optimization by numerical simulation of a perovskite solar cell, Solar Energy 181 (2019) 372–378.

[34] S. Karthicka, S. Velumania, J. Bouclé. Experimental and SCAPS simulated formamidinium perovskite solar cells: A comparison of device performance, Solar Energy 205 (2020) 349–357.

[35] Sumbel Ijaz , Ehsan Raza , Zubair Ahmad , Muhammad Zubair , Muhammad Qasim Mehmood , Haris Mehmood , Yehia Massoud , M. Muqeet Rehman. Numerical simulation to optimize the efficiency of HTM-free perovskite solar cells by ETM engineering, Solar Energy 250 (2023) 108–118.

[36] M. Born, E. Wolf, Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light", Elsevier, 2013.

[37] Lin, L.; Jiang, L.; Li, P.; Xiong, H.; Kang, Z.; Fan, B.; Qiu, Y. Simulated development and optimized performance of CsPbI3 based all-inorganic perovskite solar cells. Sol. Energy 2020, 198, 454–460.

[38] Priyanka Roy, Yassine Raoui, Ayush Khare. Design and simulation of efficient tin based perovskite solar cells through optimization of selective layers: Theoretical insights, Optical Materials 125 (2022) 112057.

[39] Laxmi Nakka, Yuanhang Cheng, Armin Gerhard Aberle, and Fen Lin. Analytical Review of Spiro-OMeTAD Hole Transport Materials: Paths Toward Stable and Efficient Perovskite Solar Cells, Adv. Energy Sustainability Res. 2022, 2200045.

[40] N. J. Jeon et al., Nat. Energy 3, 682–689 (2018).

[41] Lin, L.; Jiang, L.; Li, P.; Fan, B.; Qiu, Y. A modeled perovskite solar cell structure with a Cu2O hole-transporting layer enabling over 20% efficiency by low-cost low-temperature processing. J. Phys. Chem. Solids 2019, 124, 205–211.

[42] Devi, N.; Parrey, K.A.; Aziz, A.; Datta, S. Numerical simulations of perovskite thin-film solar cells using a CdS hole blocking layer. J. Vac. Sci. Technol. B 2018, 36, 04G105.

[43] Duan, Q.; Ji, J.; Hong, X.; Fu, Y.; Wang, C.; Zhou, K.; Wang, Z.-Y. Design of hole-transport-material free CH3NH3PbI3/CsSnI3 all-perovskite heterojunction efficient solar cells by device simulation. Sol. Energy 2020, 201, 555–560.

[44] Ehsan Raza, Zubair Ahmad , Muhammad Asif, Fakhra Aziz, Kashif Riaz, Muhammad Qasim Mehmood, Jolly Bhadra, Noora J. Al-Thani. Numerical modeling and performance optimization of carbon-based hole transport layer free perovskite solar cells, Optical Materials 125 (2022) 112075.

[45] M. Khalid Hossain, G. F. Ishraque Toki, Intekhab Alam, Rahul Pandey, D. P. Samajdar, Md. Ferdous Rahman, Md. Rasidul Islam, M. H. K. Rubel, H. Bencherif, Jaya Madan and Mustafa K. A. Mohammed. Numerical simulation and optimization of a CsPbI3-based perovskite solar cell to enhance the power conversion efficiency, New J. Chem., 2023, 47, 4801–4817.

[46] Yi-Fan Gu, Hui-Jing Du , Nan-Nan Li, Lei Yang, and Chun-Yu Zhou. Effect of carrier mobility on performance of perovskite solar cells, Chin. Phys. B Vol. 28, No. 4 (2019) 048802.

[47] A. Tara, V. Bharti, S. Sharma and R. Gupta, Opt. Mater., 2021, 119, 111362.

[48] M. Atowar Rahman. Enhancing the photovoltaic performance of Cd-free Cu_2ZnSnS_4 heterojunction solar cells using SnS HTL and TiO_2 ETL, Solar Energy ,V 215,2021,64-76/ https://doi.org/10.1016/j.solener.2020.12.020.

[49] Y. Raoui, H. Ez-Zahraouy, N. Tahiri, O. El Bounagui, S. Ahmad, S. Kazim, Performance analysis of MAPbI3 based perovskite solar cells employing diverse charge selective contacts: simulation study, Sol. Energy 193 (2019) 948–955.

[50] M. M. Salah, M. Abouelatta, A. Shaker, K. M. Hassan and A. Saeed, Semicond. Sci. Technol., 2019, 34, 115009.

[51] Liu D, Li S, Zhang P, Wang Y, Zhang R, Sarvari H, Wang F, Wu J, Wang Z and Chen Z D 2017 Nano Energy 31 462.

[52] Ompong D and Singh J 2016 Charge carrier mobility dependent open-circuit voltage in organic and hybrid solar cellsJ. Front. Nanosci. Nanotechnol. 21 43–7.

[53] Ebenhoch B, Thomson S A, Genevičius K, Juška G and Samuel I D 2015 Charge carrier mobility of the organic photovoltaic materials PTB7 and PC71BM and its influence on device performance J. Organic Electronics. 22 62–8.

[54] Burgelman, M., Nollet, P., Degrave, S., 2000. Modelling polycrystalline semiconductor solar cells. Thin Solid Films. https://doi.org/10.1016/S0040-6090(99)00825-1.

[55] A. Bag, R. Radhakrishnan, R. Nekovei, R. Jeyakumar, E ff ect of absorber layer , hole transport layer thicknesses , and its doping density on the performance of perovskite solar cells by device simulation, Sol. Energy 196 (2020) 177–182, https://doi.org/10.1016/j.solener.2019.12.014.

[56] Atanu Baga , R. Radhakrishnanb , Reza Nekoveic , R. Jeyakumar. Effect of absorber layer, hole transport layer thicknesses, and its doping density on the performance of perovskite solar cells by device simulation, Solar Energy 196 (2020) 177–182.

[57] Nguyen, W.H., Bailie, C.D., Unger, E.L., McGehee, M.D., 2014. Enhancing the hole-conductivity of spiro-OMeTAD without oxygen or lithium salts by using spiro(TFSI)2 in perovskite and dye sensitized solar cells. J. Am. Chem. Soc. 136, 10996–11001.

[58] W.S. Yang, B.W. Park, E.H. Jung, N.J. Jeon, Y.C. Kim, D.U. Lee, S.S. Shin, J.S.K. Kim, S.I. Seok, Iodide management in formamidinium-lead-halide–based perovskite layers for efficient solar cells, Science 356 (6345) (2017) 1376–1379, https://doi.org/10.1126/science.aan2301.

[59] Y. Hu, Y. Shan, Z. Yu, H. Sui, T. Qiu, S. Zhang, F. Xu, "Incorporation of γ-aminobutyric acid and cesium cations to formamidinium lead halide perovskites for highly efficient solar

cells". J. Energy Chem., vol. 64, pp. 561-567. doi: 10.1016/j.jechem.2021.05.025.

[60] D. Zhang, H. Zhang, H. Guo, F. Ye, S. Liu, Y. Wu, Stable α-FAPbI3 in inverted perovskite solar cells with efficiency exceeding 22% via a self-passivation strategy, Adv. Funct. Mater. (2022), https://doi.org/10.1002/adfm.202200174, 2200174.

[61] E. Akman, A.E. Shalan, F. Sadegh, S. Akin, Moisture-resistant FAPbI3 perovskite solar cell with 22.25% power conversion efficiency through pentafluorobenzyl phosphonic acid passivation, ChemSusChem 14 (4) (2021) 1176–1183, https://doi.org/10.1002/cssc.202002707.

[62] H. Bencheri, F. Meddour, M.H. Elshorbagy, M. Khalid Hossain, A. Cuadrado, M.A. Abdi, T. Bendib , S. Kouda, J. Alda, Performance enhancement of (FAPbI3)1-x(MAPbBr3)x perovskite solar cell with an optimized design, Micro and Nanostructures 171 (2022) 207403.

[63] Ajay Kumar, Neha Gupta, Aditya Jain , Amit Kumar Goyal , Yehia Massoud. Numerical assessment and optimization of highly efficient lead-free hybrid double perovskite solar cell, Results in Optics 11 (2023) 100387.

[64] Mari Carmen Lopez-Gonz ´ alez, Gonzalo del Pozo , Bel´en Arredondo, Silvia Delgado, Diego Martín-Martín, Marina García-Pardo, Beatriz Romero. Temperature behaviour of mixed-cation mixed-halide perovskite solar cells. Analysis of recombination mechanisms and ion migration, Organic Electronics 120 (2023) 106843.

[65] Manser J S, Christians J A and Kamat P V 2016 Chem. Rev. 116 12956.

[66] He Y and Galli G 2017 Chem. Mater. 29 682.

[67] G. O. Odunmbaku, S. Chen, B. Guo, Y. Zhou, N. A. N. Ouedraogo, Y. Zheng, J. Li, M. Li, K. Sun, Recombination Pathways in Perovskite Solar Cells. *Adv. Mater. Interfaces* 2022, 9, 2102137. https://doi.org/10.1002/admi.202102137

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

1st National Conference of Advanced Systems in Electrical Engineering (NCASEE'24). December 9th, 2024, Boumerdes, Algeria.

TCAD simulation of AlN/GaN MOS-HEMT using SiO_2 as gate insulator

Alem Raida
Laboratoire de Physique Théorique et Interactions Rayonnement-Matière (LPTHIRM), Department of Physics, Faculty of Sciences, University of Blida 1
route SOUMAA, BP 270, 09000 Blida, Algeria
alemraida9 @gmail.com

Soltani Ali
Laboratoire de Physique Théorique et Interactions Rayonnement-Matière (LPTHIRM), Department of Physics, Faculty of Sciences, University of Blida 1
route SOUMAA, BP 270, 09000 Blida, Algeria

Mattalah Maghnia
Laboratoire de Physique Théorique et Interactions Rayonnement-Matière (LPTHIRM), Department of Physics, Faculty of Sciences, University of Blida 1
route SOUMAA, BP 270, 09000 Blida, Algeria

Benzekkour Nadjet
Laboratoire de Physique Théorique et Interactions Rayonnement-Matière (LPTHIRM), Department of Physics, Faculty of Sciences, University of Blida 1
route SOUMAA, BP 270, 09000 Blida, Algeria

***Abstract*— In this work, a simulation of a normally-on AlGaN/GaN Metal–Oxide–Semiconductor High Electron Mobility Transistor (MOS-HEMT) on silicon substrate is investigated by Silvaco TCAD using a thin AlN barrier layer with 7 nm of SiO_2 as gate insulator. The utilization of AlN/GaN heterostructures is specifically examined for its potential to shift the threshold voltage toward more negative values and a comparison of device performances of two MOS-HEMT structures using thin AlN and AlGaN barrier layer is reported. The simulated normally-on device exhibits a pinch-off voltage of -6.7 V and a maximum IDS current density of 1000 mA/mm at Vg= -3V.**

Keywords—MOS-HEMT, TCAD simulation, normally-on, AlGaN, AlN .

I. INTRODUCTION

Technology Computer Aided Design (TCAD) simulations have emerged as a powerful tool for exploring and optimizing GaN-based HEMTs (High Electron Mobility Transistor). These devices benefit from the exceptional material properties of gallium nitride (GaN), including a wide bandgap, excellent electron mobility, high breakdown voltage, and suitability for high-frequency operation. Such properties make GaN-based HEMTs ideal for advanced power electronics and RF applications [1,2]. However, the inherent challenges associated with their design and performance demand more in-depth research efforts to better control and fine-tune the behavior of these devices.

In recent years, various structural modifications have been proposed to enhance the performance of GaN-based HEMTs. Among these approaches, reducing the AlGaN barrier thickness with increasing the aluminum concentration has shown promise in improving device characteristics [3,4]. Additionally, the use of ultrathin AlN/GaN barrier heterostructures which provide higher carrier density and improved scalability, has opened new avenues for millimeter wave power applications [5,6].

In this work, based on experimental published results, a simulation of a normally-ON AlGaN/GaN Metal–Oxide–Semiconductor High Electron Mobility Transistor (MOS-HEMT) on silicon substrate is investigated by Silvaco TCAD using an ultra-thin $Al_{0.45}Ga_{0.55}N$ barrier layer with 7 nm of SiO_2 as oxide gate insulator. The utilization of AlN/GaN heterostructures is specifically examined for its potential to shift the threshold voltage toward more negative values and a comparison of device performances of two MOS-HEMT structures using thin AlN and AlGaN barrier layer is reported..

II. TCAD SIMULATION METHOD

Using Silvaco Atlas TCAD (Technology Computer-Aided Design), a simulated AlGaN/GaN MOS-HEMT is presented in Fig.1 and obtained from the literature [3].

the AlGaN/GaN structure is grown on a silicon (111) substrate with 1µm Al_0.08 Ga_0.92 N back barrier layer and 2nm AlN (Aluminum Nitride) nucleation layer. The epitaxial structure consists of 150 nm undoped GaN channel layer and thin 4 nm $Al_{0.08}Ga_{0.92}N$ barrier layer, with 7 nm of SiO_2 as gate insulator layer. The dimensions of the device structure are 2.5 µm, 2 µm and 3.5 µm for source to gate distance, gate length and gate to drain distance, respectively.

A Ni/Au metal was used as gate contact. The structure has been passivated with 50 nm of silicon nitride.

The source and drain contacts are ohmic contacts, while the gate is made of a Schottky materials. The model is based on a drift-diffusion transport model and fermi statistics, a doping dependency model, a high velocity saturation model, and SRH model for generation-recombination are used in the simulation. Piezoelectric and spontaneous polarization were included to consider the formation of the 2-D electron gas (2DEGs) at the AlGaN/GaN interface. The AlGaN barrier was changed with thin AlN barrier to investigate the impact of thin AlN barrier layer on electrical performance of MOS-HEMT transistor.

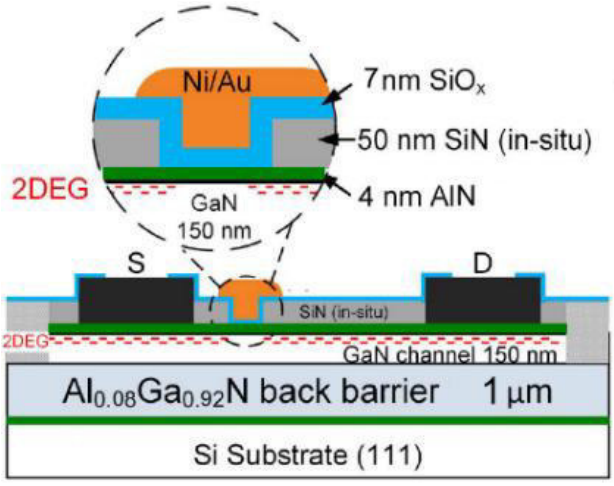


Fig.1. Schematics of MOS-HEMT structure with 4 nm AlN barrier.

The mesh is significantly refined at critical regions, particularly around the gate edges, and especially for the Gate-Drain, and all semiconductor interfaces, to ensure an efficient simulation.

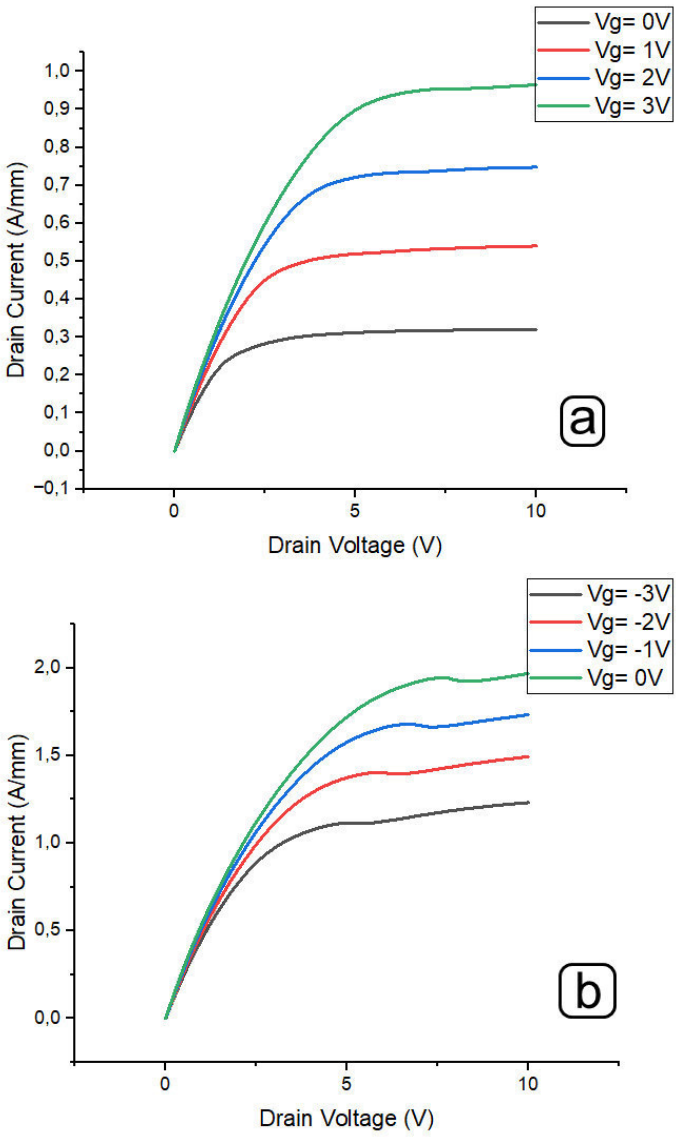


Fig.2. IDS –VDS characteristic calculated with Silvaco TCAD on MOS-HEMTs (a) with 4 nm AlGaN barrier (reference) and (b) with 4 nm AlN Barrier.

III. Results and Discussions

Figure 2 shows the IDS –VDS characteristics for the simulated AlGaN/GaN MOS-HEMTs with 4 nm AlGaN barrier thickness (Fig. 2(a)), demonstrates a normally-on operation with a pich-off voltage (V_p) of -1.5V, that aligns closely to the experimental value. The MOS-HEMT with 4 nm AlN thin barrier shows a normally-on operation with a pich off voltage of -6.7V with a current density of 1000 mA/mm at VGS=-3V (Fig. 2(b)).

Figure 3 shows the device hysteresis transfer characteristics (IDS–VGS) at VDS=10 V. The pinch-off voltage (Vp) was shifted from -1.1V to 1V after changing the AlGaN barrier with a AlN barrier. This demonstrates the critical role of thin AlN/GaN barrier in enabling high carrier density, which is essential for achieving even even higher frequencies and enhanced power performance .

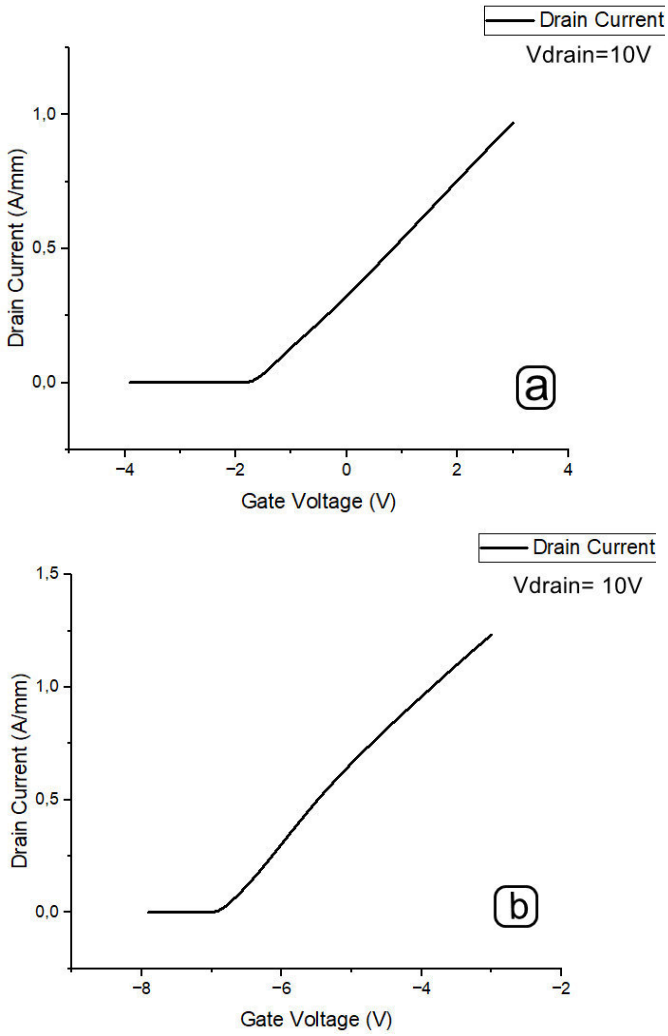


Fig.3. Linear transfer characteristics IDS –VGS of the simulated MOS HEMTs device (a) with 4 nm AlGaN barrier, (b) with 4 nm AlN barrier.

IV. Conclusion

In summary, a normally-on MOS-HEMT device was simulated by Silvaco TCAD and demonstrated by using a thin AlGaN/GaN barrier with a pinch-off voltage (V_p) of -1.5V which is shifted to -6.7V using a thin AlN/GaN barrier layer which is essential for achieving even higher frequencies and enhanced power performance.

References

[1] M.K. Reddy, J. Lakshmi, A. Hemanth, B.H. Kumar, L. Bandi and G. Sheu , "Physics Based TCAD Simulation and Calibration of GaN/AlGaN/GaN HEMT Device," IEEE. 27 February 2020.

[2] K. Mukherjee, F. Darracq, A. Curutchet, N.Malbert, N. Labat, "TCAD simulation capabilities towards gate leakage current analysis of advanced AlGaN/GaN HEMT devices," Microelectronics Reliability. 2017.

[3] A. Chakroun, A. Jaouad, M. Bouchilaoun, O. Arenas, A. Soltani. ''Normally-off AlGaN/GaN MOS-HEMT using ultra-thin Al0.45 Ga0.55N barrier layer'' Physica Status Solidi A (applications and materials science), 214 (8), 10.1002/pssa.201600836. hal-01914374, 2017.

[4] A. Chakrabarty, N. Sahoo, A.K. Panigrahy, V Bharath Sreenivasulu, R. Swain, '' DC and RF performance analysis of enhancement mode fin-shaped tri-gate AlGaN/GaN HEMT and MOSHEMT with ultra-thin barrier layer Phys. Scr, 99, 2024.

[5] X. Luo , S. Halder , W. R. Curtice , J. C. M. Hwang , K. D. Chabak , D. E. Walker, Jr , and A. M. Dabiran, "Scaling and High-Frequency Performance of AlN/GaN HEMTs", 2017.

[6] Kathia Harrouche, Riad Kabouche, Etienne Okada, F Medjdoub. High Power AlN/GaN HEMTs with record power-added-efficiency >70% at 40 GHz. IEEE/MTT-S International Microwave Symposium (IMS 2020), Aug 2020, Los Angeles, CA, United States. Session Tu3H - Advances in Microwave Semiconductor Devices, paper Tu3H-2, 285-288, 10.1109/IMS30576.2020.9223971 . hal-03043653f

Improving vehicle detection from complex urban aerial images with yolov8

Yasmine Silhadi
Faculty of Electrical Engineering, Djillali Liabes University of Sidi Bel Abbes, Laboratory Intelligent Control and Electrical Power System (ICEPS), Algeria,
djawedmeg08@gmail.com.

Hocine CHEBI
Faculty of Electrical Engineering, Djillali Liabes University of Sidi Bel Abbes, Laboratory Intelligent Control and Electrical Power System (ICEPS), Algeria.
chebi.hocine@yahoo.fr, hocine.chebi@univ-sba.dz,
chebi.hocine@univ-boumerdes.dz.

***Abstract*—Advances in computer vision and deep learning have enabled significant progress in real-time object detection, particularly with the introduction of powerful models such as YOLO. However, despite these advances, vehicle detection and tracking in complex urban environments from drone-captured data still pose significant challenges. These environments are characterized by the presence of numerous static and moving objects, variable lighting conditions, and frequent occlusions. These conditions affect the accuracy of current systems and highlight the need for specific improvements. It is in this context that we propose this model for vehicle detection based on You Look Only Once version 8 and pre-processing the images with the clahe filter to help improve the contrast of the latter. Experimental results and comparison show that the proposed method for vehicle detection gives an accuracy of 96.5% on the Vehicle Aerial Imaging From Drone (VAID) dataset.**

Keywords—Vehicle detection, Vehicle Aerial Imaging, yolov8, real-time object detection, tracking in complex urban environments.

I. INTRODUCTION

Object detection is an important branch of computer vision widely applied in daily life, such as civil or military security surveillance, autonomous driving, robotics [1], [2].

Previous object detection methods relied on machine learning methods [3], but their performance was often limited in complex or dense environments, partially occluded objects. These methods show limitations when dealing with fast moving objects and are often inefficient when dealing with small objects [4], [5].

But new methods based on Deep-learning have revolutionized the field of computer vision, with algorithms that can automatically extract complex features from data, largely surpassing conventional techniques. Thanks to Convolutional Neural Networks (CNN) [6], [7], YOLO [8], Transformer [9] that perform real-time detection, it is now possible to detect objects with high accuracy, even under varied scale and lighting conditions, with increased robustness, especially in the context of large-scale surveillance, especially by drone, where the management of altitude variations and viewing angles is crucial for object detection and tracking [10].

Our work consists in detecting vehicles in image sequences captured by drone, able to overcome the challenges of considerable scale variations, partial occlusion and light changes. The system will be tested on the VAID dataset [11].

The paper is organized as follows. Section II reviews the state-of-the-art work. The detection method used is explained in Section III. Performance measurements, experimental results and discussion are described in Section IV. The conclusion is presented in Section V.

II. REVIEW OF THE STATE OF THE ART OF THE WORKS

Object detection is a primary task in computer vision, consisting of identifying, locating objects in real time in image sequences. The advances developed in this field of vehicle detection using aerial images captured by drones are briefly described in this section.

The use of deep learning methods for vehicle detection [12], [13] is a trend in the computer vision community, and the one that excels in this field is the use of You Only Look Once (YOLO) [8]. Known for its ability to process multiple images very quickly compared to other two-stage methods such as region-based convolutional neural networks (RCNNs) [14]. YOLO is widely used for real-time detection applications, for its efficiency and speed in processing data. Yolov5 is widely used for precision in detecting vehicles despite the change of illumination and the variation of scale [15]. In [16] they used yolov4 combined with CNNs for aerial image classification. Yolov3 and RetinaNet [17] were used for multiple object detection, and in [13], [18] different versions of YOLO were used for road traffic monitoring. In this work we will use Yolov8[19] for its simplicity and efficiency in detecting vehicles in a complex environment.

III. DETECTION METHOD

This paper proposes a method for vehicle detection based on images captured by drone, a preprocessing step will be

performed for all the data and then the detection by the yolov8 method.

In the preprocessing step, all the images were converted to GRAYSCALE, then an application of the CLAHE filter to improve the contrast of the pixels was carried out on all the images [16].

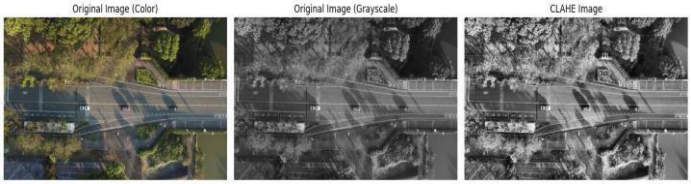


Fig. 1. the different phases of pre-treatment.

a) Vehicle detection using YOLOv8 :

Yolov8 offers various models, each one specialized in a different task in computer vision. These models are intended for various tasks such as segmentation, detection and classification, while minimizing the number of parameters and increasing the robustness and accuracy of the model.



Fig. 2. vehicle detection using yolov8.

IV. EXPERIMENTS AND RESULTS

The vision-based method for vehicle detection was implemented on the hosted GPU T4 runtime environment of Google Colab, using Python tools. The method was tested on the VAID dataset, and the details and results obtained are described below with a comparison with the state of the art.

1. Dataset description :

Aerial dataset for vehicle detection VAID[11], contains 5985 RGB annotated images in .jpg format, of dimension (1137*640) pixels, taken under different illuminations and angles at different locations in Taiwan.

2. Results obtained :

We used three metrics to evaluate our model, and the results obtained are described in Table 1.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (1)$$

Or TP corresponding to True Positives and FP False Positives.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (2)$$

FN stands for False Negatives.

$$F1\ Score = 2\ \frac{\mathrm{Precision}\,.\,\mathrm{Recall}}{\mathrm{Precision} + \ \mathrm{Recall}} \quad (3)$$

After applying the CLAHE filter, the YOLOv8 model achieved a high accuracy of 96.5%, indicating a remarkable ability to minimize false positives. A recall of 93.2% and an overall F1-score of 94.9%, the model demonstrates excellent overall performance for object detection in aerial images.

METRIC	VALUE
Precision	0.965
Recall	0.932
F1 Score	0.949

Table 1: Metrics and results.

a) Comparison with the state of art:

In this experiment, we compare our approach with other state of art methods, and the results obtained are described in table 2.

Article	Method	Dataset	Results
Vehicle Detection and Tracking in UAV Images Via Yolov3 and Kalman filter [20]	Yolov3	VAID	96,7%
Smart Traffic monitoring Through Pyramid Pooling Vehicle Detection and Filter-based Tracking [21]	CNN + pooling pyramidal	VAID VEDAI DLR3K	95,78% 95,18% 93,13%
Aerial Dataset Integration For Vehicle Dtection Based Yolov4 [20]	Yolov4	VAID UAVD DOTA	76,7% 72,9% 75,3%
OURS	Yolov8	VAID	96,5%

Table 2: Comparison of the accuracy of the proposed method with other state of art techniques.

The results show that the proposed method achieves an accuracy of 96.5% on the VAID dataset, approaching the best performance obtained with YOLOv3 (96.7%). And outperforms other methods.

V. CONCLUSION

In conclusion, This paper proposes a preprocessing followed by a detection method for vehicles in an urban environment and aims to provide an improvement for the accuracy of vision systems embedded in drones.

Références

[1] T. Zhang, S. Liu, C. Xu, B. Liu, et M.-H. Yang, « Correlation Particle Filter for Visual Tracking », IEEE Trans. Image Process., vol. 27, no 6, p. 2676‑2687, juin 2018, doi: 10.1109/TIP.2017.2781304.

[2] M. Yacine, « Deep Learning pour la détection des plaques d'immatriculation ».

[3] M. Abernot, S. Gauthier, T. Gonos, et A. Todri-Sanial, « SIFT-ONN: SIFT Feature Detection Algorithm Employing ONNs for Edge Detection », in Neuro-Inspired Computational Elements Conference, San Antonio TX USA: ACM, avr. 2023, p. 100 ‑ 107. doi: 10.1145/3584954.3584999.

[4] S. Ren, K. He, R. Girshick, et J. Sun, « Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks », IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no 6, p. 1137‑1149, juin 2017, doi: 10.1109/TPAMI.2016.2577031.

[5] W. Liu et al., « SSD: Single Shot MultiBox Detector », in Computer Vision – ECCV 2016, vol. 9905, B. Leibe, J. Matas, N. Sebe, et M. Welling, Éd., in Lecture Notes in Computer Science, vol. 9905. , Cham: Springer International Publishing, 2016, p. 21‑37. doi: 10.1007/978-3-319-46448-0_2.

[6] A. Khan, A. Sohail, U. Zahoora, et A. S. Qureshi, « A Survey of the Recent Architectures of Deep Convolutional Neural Networks », Artif. Intell. Rev., vol. 53, no 8, p. 5455 ‑ 5516, déc. 2020, doi: 10.1007/s10462-020-09825-6.

[7] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, et S. Belongie, « Feature Pyramid Networks for Object Detection », in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI: IEEE, juill. 2017, p. 936 ‑ 944. doi: 10.1109/CVPR.2017.106.

[8] J. Redmon, S. Divvala, R. Girshick, et A. Farhadi, « You Only Look Once: Unified, Real-Time Object Detection », in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, juin 2016, p. 779‑788. doi: 10.1109/CVPR.2016.91.

[9] A. Vaswani et al., « Attention Is All You Need », 1 août 2023, arXiv: arXiv:1706.03762. Consulté le: 8 octobre 2023. [En ligne]. Disponible sur: http://arxiv.org/abs/1706.03762

[10] Z. Zhang, X. Lu, G. Cao, Y. Yang, L. Jiao, et F. Liu, « ViT-YOLO:Transformer-Based YOLO for Object Detection », in 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada: IEEE, oct. 2021, p. 2799‑2808. doi: 10.1109/ICCVW54120.2021.00314.

[11] H.-Y. Lin, K.-C. Tu, et C.-Y. Li, « VAID: An Aerial Image Dataset for Vehicle Detection and Classification », IEEE Access, vol. 8, p. 212209‑212219, 2020, doi: 10.1109/ACCESS.2020.3040290.

[12] [12] S. Pouyanfar et al., « A Survey on Deep Learning: Algorithms, Techniques, and Applications », ACM Comput. Surv., vol. 51, no 5, p. 1‑36, sept. 2019, doi: 10.1145/3234150.

[13] S. Ali et A. Jalal, « Vehicle Detection and Tracking from Aerial Imagery via YOLO and Centroid Tracking ».

[14] Z. Cai et N. Vasconcelos, « Cascade R-CNN: Delving Into High Quality Object Detection », in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT: IEEE, juin 2018, p. 6154‑6162. doi: 10.1109/CVPR.2018.00644.

[15] A. M. Qureshi, « Semantic Segmentation and YOLO Detector over Aerial Vehicle Images », Comput. Mater. Contin., vol. 80, no 2, p. 3315‑3332, 2024, doi: 10.32604/cmc.2024.052582.

[16] M. O. Yusuf, M. Hanzla, et A. Jalal, « Vehicle Detection and Classification via YOLOv4 and CNN over Aerial Images », in 2024 International Conference on Engineering & Computing Technologies (ICECT), Islamabad, Pakistan: IEEE, mai 2024, p. 1‑6. doi: 10.1109/ICECT61618.2024.10581252.

[17] S. Kapania, D. Saini, S. Goyal, N. Thakur, R. Jain, et P. Nagrath, « Multi Object Tracking with UAVs using Deep SORT and YOLOv3 RetinaNet Detection Framework », in Proceedings of the 1st ACM Workshop on Autonomous and Intelligent Mobile Systems, Bangalore India: ACM, janv. 2020, p. 1‑6. doi: 10.1145/3377283.3377284.

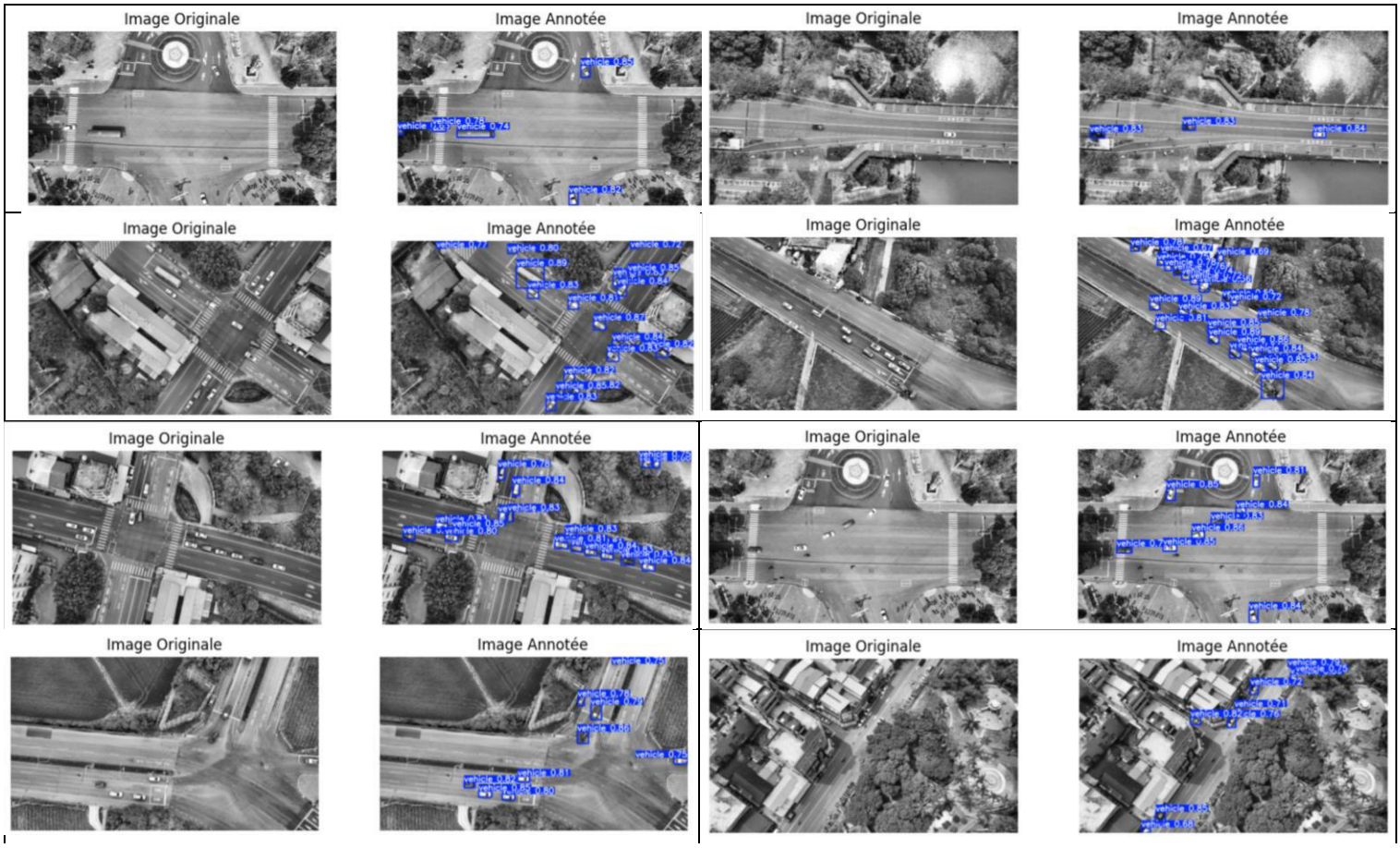
[18] W. Omar, Y. Oh, J. Chung, et I. Lee, « Aerial Dataset Integration For Vehicle Detection Based on YOLOv4 », Korean J. Remote Sens., 2021.

[19] M. Hanzla, M. O. Yusuf, et A. Jalal, « Vehicle Surveillance Using U-NET Segmentation and DeepSORT Over Aerial Images », in 2024 International Conference on Engineering & Computing Technologies (ICECT), Islamabad, Pakistan: IEEE, mai 2024, p. 1‑6. doi: 10.1109/ICECT61618.2024.10581338.

[20] S. Ali, A. Jalal, M. Hamad Alatiyyah, K. Alnowaiser, et J. Park, « Vehicle Detection and Tracking in UAV Imagery via YOLOv3 and Kalman Filter », Comput. Mater. Contin., vol. 76, no 1, p. 1249‑1265, 2023, doi: 10.32604/cmc.2023.038114.

[21] A. A. Rafique et al., « Smart Traffic Monitoring Through Pyramid Pooling Vehicle Detection and Filter-Based Tracking on Aerial Images », IEEE Access, vol. 11, p. 2993 ‑ 3007, 2023, doi: 10.1109/ACCESS.2023.3234281.

Fig. 3. Results obtained for vehicle detection



Drones In Digital Transformation For Industry :Applications, Challenges and Future Trends.

Tayeb AOUFI
Department of Industrial Engineering,
Mostapha BenBoulaid University
Automatic and Production Laboratory
(LAP, Laboratoire d'Automatique et Productique)
05000 Batna 2, Algeria.
tayeb.aoufi@univ-batna2.dz

Karima AKSA
Department of Industrial Engineering,
Mostapha BenBoulaid University
Automatic and Production Laboratory
(LAP, Laboratoire d'Automatique et Productique)
05000 Batna 2, Algeria.
k.aksa@univ-batna2.dz

***Abstract*— The swift proliferation of Industry 4.0 has emphasized the significance of drones in facilitating digital transformation across numerous industries. Drones equipped with data processing capabilities, advanced sensors, and cameras are essential for data collection and monitoring, safety and risk mitigation, as well as communication and collaboration. Their combination with new technology promotes the development of smart factories, allowing for predictive maintenance, improved safety, and efficient resource management. Drones help to improve operational efficiency, reduced costs, and make better decisions in industries such as manufacturing, agriculture, logistics, and construction. However, the widespread use of drones confronts substantial challenges. These include regulatory and legal barriers, system integration challenges, data security issues, and the requirement for workforce training. This paper explores the multifaceted applications and challenges of drones in digital transformation, emphasizing their revolutionary impact on industrial processes and contribution to the future of intelligent.**

Keywords—drones, digital transformation, industry 4.0, manufacturing, smart factory.

I. INTRODUCTION

Drones are widely acknowledged as essential elements in the digital transformation sphere, especially within the context of Industry 4.0. Their incorporation into many sectors improves operating efficiency and creates new opportunities for innovation. The utilization of drones across diverse sectors is swiftly revolutionising operating methodologies, improving efficiency, and facilitating creative solutions. Unmanned aerial vehicles (UAVs), or drones, are being used more and more for inspections, data collection, and tracking in fields like manufacturing, construction, and agriculture. Their addition to Industry 4.0 marks a shift towards processes that are more automatic and data-driven. Drones are an important part of Industry 4.0 because they make inspections, mapping, and surveying possible, which makes manufacturing and building more efficient [1]. Their adaptability facilitates regular inspections and maintenance, hence improving production and safety [2]. Drones are essential to the advancement of intelligent transport systems (ITS), enhancing logistics and minimizing environmental effects [3]. This evolution presents considerable challenges and new trends that influence their implementation. These challenges can be classified into technological, organizational, and environmental elements, affecting sectors including manufacturing, oil & gas, and construction. Challenges including regulatory compliance, cost, worker competencies, and system integration must be resolved to fully exploit their promise in industrial applications. The incorporation of drones into industrial operations is a notable trend in the current digital transformation.

The study is organized as follows: Section 2 present general information about drones; in Section 3, we provide materials and method of search; in Section 4, we discuss the application of drones in digital transformation for industry, in section 5, we present the challenges of drones in digital transformation then we discuss the most important future research pathways; and lastly, we summarize the major conclusion.

II. GENERAL INFORMATION ABOUT DRONES

A. *Main research areas of drone application:*

Drones, have emerged as versatile tools across various research areas, significantly enhancing efficiency and data collection capabilities. Their applications span multiple sectors, including environmental monitoring, infrastructure assessment, and emergency response.



Fig. 1. Figure showing some research area of drones [4]

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Below are the main research areas of drone application:

1) Emergency Response and Disaster Management

Drones are vital in search and rescue operations, providing real-time data and imagery during disasters [5].

They enhance law enforcement capabilities by facilitating crowd monitoring and controlling [5].

2) Military Applications

Drones are utilized for surveillance and reconnaissance, optimizing resource allocation and reducing operational costs in military contexts [6].

3) Precision Agriculture

Drones are utilized for pest control and crop monitoring, employing deep learning models for disease detection and efficient pesticide application [7].

They enhance agricultural productivity by enabling real-time data collection and analysis, which aids in decision-making processes.

4) Healthcare Delivery

Drones are increasingly used for delivering medical supplies, categorized into time-critical (e.g., emergency response) and non-time-critical deliveries, enhancing healthcare supply chains [8].

5) Industry 4.0 and digital transformation

Drones support smart factory operations by providing reliable wireless communication, crucial for the seamless integration of Industry 4.0 technologies [9].

6) Urban Planning and Smart Cities

Drones can provide high-resolution imagery for urban planning, enabling detailed mapping and monitoring of city layouts, greenery coverage, and infrastructure development [10].

The use of UAVs allows for real-time data acquisition, which is crucial for dynamic urban management and planning in rapidly expanding cities [10].

7) Logistics and Supply Chain Management

Drones can achieve time savings of nearly 80% in parcel delivery compared to traditional methods, particularly when combined with ground vehicles [11]

The combination of drones with automated vehicles is creating intelligent delivery systems, addressing various optimization challenges such as the flying sidekick traveling salesman problem [12].

8) Environmental Monitoring

Drones are utilized for precision agriculture, enabling efficient crop monitoring and pest detection through remote sensing technologies [13].

They play a crucial role in environmental conservation, allowing for the assessment of ecological boundaries and types in power grid projects [14].

9) Infrastructure and Engineering

UAVs are increasingly employed in infrastructure monitoring, such as aerial damage assessments and boundary mapping using ground-penetrating radar (GPR) [15].

Their application in power grid engineering helps identify control area boundaries, optimizing project planning [14].

10) Climate Research and Weather Forecasting

Drones can reach altitudes up to 10 km, filling critical observational gaps in the troposphere, particularly in remote and polar regions [16]. Drones are being integrated into operational weather forecasting systems, such as the pilot project in North Dakota, which utilizes drone data to refine weather models [17]

They also support UAV path planning by incorporating probabilistic weather forecasts, which enhance mission safety and efficiency [18]

11) Surveying and Mapping

Drones have transformed land surveying by providing high-resolution imagery and efficient data collection, significantly reducing time and costs compared to traditional methods [19].

They are integrated with Geographic Information Systems (GIS), improving accessibility and accuracy in geospatial data collection for applications like urban planning and disaster management [20]

B. The most common types of drones used in industry

Drones have become integral to various industries, primarily due to their efficiency and versatility in performing tasks that are often hazardous or difficult for humans. The most common types of drones utilized in industrial applications include quadcopters, fixed-wing UAVs, and hybrid models, each serving distinct purposes.

Quadcopters: These are widely used for visual inspections and non-destructive testing due to their ability to hover and maneuver in tight spaces. They are equipped with advanced navigation systems that allow for real-time mapping and data collection [21].

Fixed-Wing UAVs: Ideal for covering large areas, these drones are often employed in agricultural monitoring and infrastructure inspections, where extensive coverage is required [22].

Fixed-wing UAVs outperform rotary-wing counterparts in speed and endurance, making them suitable for long-distance missions [23].

Hybrid Drones: Combining the features of both quadcopters and fixed-wing UAVs, these drones can take off and land vertically while also flying long distances, making them suitable for complex industrial tasks [24].



Fig. 2. The most common types of drones used in industry

III. MATERIALS AND METHODS

Within the scope of the research, Pre-defined selection criteria were developed to extract papers from Scopus, Web of Science, and Google Scholar about drone technologies. content analysis helps the studies on the applicability and challenges associated with drones in digital transformation for industry evaluated.

From the Clarivate Web of Science and Scopus databases, we obtained the titles, keywords, and abstracts of every publication. We also incorporated grey literature gathered on Google Scholar since the topic of study is new. We requested many papers from the authors due to their unavailability in open access.
The project is to gather all the application and challenges of drones in digital transformation for next researches.

IV. Application of Drones in Digital Transformation for Industry

After our in-depth examination and study of papers related to drones in digital transformation, we concluded the following applications:

A. Data Collection and Monitoring

Within the framework of digital transformation in numerous industries, drones are becoming more and more important for data collecting and monitoring. Their adaptability, speed, and connection with cutting-edge technology help to greatly improve operational performance and data management.
The age of information (AoI) is substantially reduced by drones, which optimize flight trajectories and transmission schedules through deep reinforcement learning, thereby facilitating fresh data collection in IoT networks [25]. This method improves the availability of real-time data, which is essential for applications that require rapid response. In the same vein, networked unmanned aerial vehicles (UAVs) function as aerial sensor networks, thereby enhancing their surveillance and monitoring capabilities. In comparison to conventional methods, they provide improved mobility and adaptability, rendering them indispensable for industrial and public safety applications [26]. Furthermore, proven successful in difficult terrain, creative ideas like the Data Collection Fly (DCFly) use small aerial vehicles to reduce energy usage in wireless sensor networks. This approach lessens general communication requirements and improves operational effectiveness. Moreover, including deep learning for drone monitoring enables enhanced identification and tracking even in challenging surroundings, a capacity essential for assuring correct data collecting and monitoring [27].

B. logistics and delivery

Drone applications in distribution and logistics are fast changing and provide major benefits in cost control, environmental impact, and efficiency. Especially for last-mile logistics, where they can improve service speed and lower carbon emissions, drones are progressively included into delivery systems.
Achieving time savings of around 80% over conventional truck-only approaches, drones can significantly shorten delivery times [11]. As the Flying Sidekick Traveling Salesman Problem shows, the mix of drones with ground vehicles—such as trucks—optimizes delivery routes and reduces completion times [28].

C. Safety and Risk Mitigation:

During digital transformation, including drones into industrial operations greatly improves risk reduction and safety. Particularly in industries like oil and gas and construction, drones help to maximize operational efficiency while reducing human contact to dangerous surroundings.
By performing inspections and monitoring high-risk environments like deepwater oil platforms, drones help to minimize the need for workers in hazardous conditions [29]. Drone swarms are also used in building for site mapping and safety inspections, so guaranteeing adherence to safety regulations and so enhancing worker safety [30]. Compliance and risk management depend on environmental safety monitoring, which also depends critically on drones in controlling leaks and methane emissions detection [31]. Moreover, sensors aboard drones allow them to quickly spot and examine industrial process irregularities. By automating the monitoring of major activities, they increase response times and accuracy in problem diagnosis [32].
Although drones have great advantages in terms of risk reduction and safety, implementation of them still depends critically on issues such cybersecurity concerns and the necessity of strong privacy policies.

D. Communication and Collaboration

The integration of drones into industrial processes significantly enhances communication and collaboration, driving digital transformation across various sectors. Drones, particularly in the context of Industry 5.0, optimize operations and improve human-machine collaboration, leading to increased efficiency and reduced costs [33]. Their ability to gather real-time data and deliver payloads facilitates rapid decision-making and operational agility, especially in industries like healthcare, where drones can transport critical medical supplies [34]. Furthermore, the application of advanced technologies such as AI and blockchain in Drone-as-a-Service (DraaS) enhances data security and operational autonomy, fostering a more collaborative environment [35]. In the oil and gas sector, structured teams utilizing drones promote innovation and streamline processes, further supporting digital transformation efforts [36].

E. Integration with internet of thing, Big Data and artificial intelligent :

Driving digital transformation across several industries depends mostly on the integration of drones with the Internet of Things, artificial intelligence, and big data. This synergy improves operational effectiveness, data analysis, and decision-making procedures, therefore changing the scene of the business.
Improving efficiency in areas including agriculture and construction drones fitted with artificial intelligence may independently navigate and complete challenging jobs [37]. Furthermore, the fast gathering and analysis of large datasets made possible by the combination of drones and Big Data helps businesses to extract actionable insights for better decision-making [37]. Moreover, by automating tasks and enhancing asset management—qualities necessary for smart factories striving for re-industrialization—drones are becoming important in industrial internet of things (IIoT) [38]. Furthermore, including drones into IIoT systems improves the quality and coverage of services by means of sophisticated security assessments, therefore mitigating vulnerabilities [39]. Adoption of drones inside IIoT systems promotes digital transforming skills, like business model and

operational changes, which are necessary to keep competitiveness in a fast-changing market [40].
.

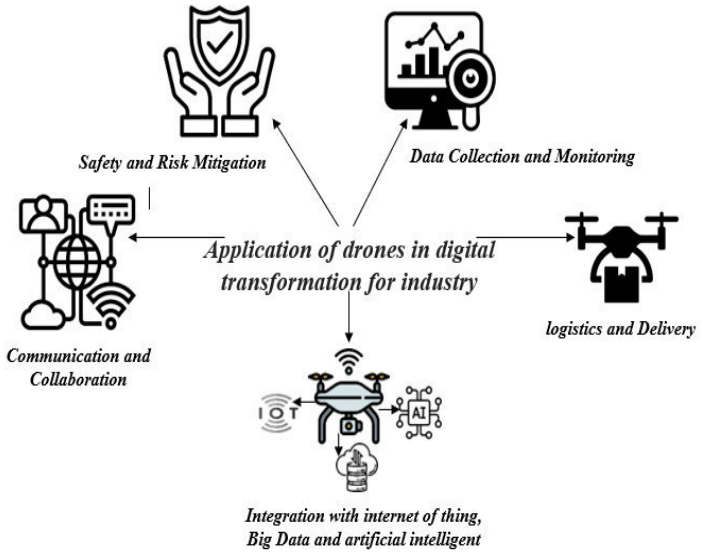


Fig. 3. The application of drones in digital transformation for industry

V. CHALLENGES OF DRONES IN INDUSTRIES UNDERGOING DIGITAL TRANSFORMATION

In this section we present the challenges of drones in digital transformation based on our study.

A. Regulatory and Legal Barriers :

Different legal and regulatory constraints greatly impede the acceptance of drones in sectors experiencing digital change. Lack of thorough systems, ethical considerations, and safety issues define these difficulties.
Current regulations are frequently outdated or insufficient, with a primary focus on military rather than commercial applications [41]. Moreover, drones could endanger human populations, which calls for strict safety rules that might discourage creativity [42]. Legal uncertainty about liability in case of mishaps makes companies reluctant to use drone technology [43]. Furthermore, the spread of drones complicates the legal environment by posing ethical issues about surveillance and privacy, therefore changing the scene [44].

B. Data Security and Privacy :

There are significant privacy and data security issues when drones are included into sectors going digital. Robust security measures are absolutely necessary to safeguard private information and guarantee operational integrity as drones find increasing presence in many different uses. Denial of Service (DoS) and Man-in----middle attacks are among the several threats that drones are vulnerable to, therefore compromising data integrity and availability [45]. Furthermore, aggravating these weaknesses in many drone systems is their absence of built-in security measures; so, customized security solutions are needed [46]. Moreover, particularly in surveillance and monitoring, the use of drones begs serious privacy concerns, particularly with relation to data collecting and distribution in civilian environments [47]. Also, by allowing real-time monitoring and anomaly detection, emerging technologies include Software Defined Networks (SDN) and Machine Learning can improve drone communication security [45].

C. Operational Challenges:

Integration of drones into industries experiencing digital transformation presents many operational difficulties mostly resulting from legislative, technological, and alignment problems. These difficulties have to be resolved if we are to fully use unmanned aerial systems (UAS).
The limited regulatory environment of today for UAS limits their operations mostly to military ones. This absence of a thorough structure limits business acceptance [41]. Furthermore, improving drone capabilities depends on the convergence of edge computing and artificial intelligence; hence, using these technologies presents difficulties including guaranteeing low latency and energy economy [48]. Moreover, effective drone integration depends on operational alignment between corporate strategy and IT capacity [49].

D. Cost and Investment

Including drones into industrial digital transformation offers serious financial and technological difficulties. Among the several reasons for these difficulties are operational expenses, technology needs, and regulatory compliance.
The profitability of drone delivery services is mostly dependent on electricity and battery buying expenses, which could rise with aging of batteries [50]. As the use of drones grows, the building industry has to pay more for safety and liability, which means they have to spend more on training and risk control [51]. Also, drones need to use edge computing technologies, which can be pricey, in order to do AI processing quickly [52].
Drones have the ability to change many industries, but they are expensive and hard to invest in, so they need to be carefully planned and strategically invested in order to be successfully integrated into digital transformation efforts.

E. System Integration Challenges:

In the framework of digital transformation across numerous industries, the inclusion of drones into current systems poses major difficulties. These difficulties result from the complexity of system integration, legal obstacles, and necessity of strong safety precautions.
As part of intricate systems-of- systems (SoS), drones suffer integration problems comparable to those in avionics and automotive sectors, where late-stage integration results in high costs and inefficiencies [53]. Service-oriented architectures (SOAs) can also help to solve some integration issues; but, incorrect application could lead to fragmentation and problems with requirements traceability [53]. Furthermore, compounded by regulatory concerns and public opinion is the integration of Unmanned Aircraft Systems (UAS) into non-segregated airspace, therefore impeding operating capabilities [54].

F. Workforce Skills and Training Needs :

Including drones into sectors of industry as part of digital transformation calls for major workforce adaptation and



Fig. 4. The challenges of drones in the digital transformation of industry.

training. The demand for new skills and job profiles consistent with Industry 4.0 technology fuels this change.

The move toward digital technologies including drones and automation is changing job profiles and calling for workforce reskilling and up-to-date skills [55]. Stakeholders have to understand that the effective application of drone technology depends on matching the interests of workers and companies, therefore guaranteeing that training programs are relevant and efficient [56].

All things considered, even if the digital revolution brought by drones brings difficulties, it also gives chances for workforce development via focused training and adaption policies. Still, the complexity of technology integration calls for careful management to guarantee a harmonic response to labor changes.

G. Ethical and Social Considerations :

Especially as the digital revolution accelerates, the integration of drones into many different fields raises major ethical and social questions. These challenges touch accountability, privacy, and legal systems—all of which are quite important for prudent drone deployment.

Effective digital transformation with drones must first give community relations top attention so that local people are informed and consulted concerning drone activities [57]. Furthermore, the larger digital economy deals with basic social and ethical problems that need to be resolved if technology is to help to bring about good changes in society [58].

Although drones offer great possible advantages, proper navigation of the ethical and social issues they raise guarantees responsible use and public confidence.

H. scalability and reliability

Particularly in industries like oil and gas, the scalability and dependability of drones in the framework of digital transformation generate main challenges. Dealing with these challenges requires a varied approach combining operational concepts, security, and technology.

Establishing a Center of Excellence (CoE) will help to enable the scalable use of drones by motivating invention and information sharing among interdisciplinary teams [59]. Furthermore, depending on tailoring drone applications to specific use cases enhancing operational efficiency and guaranteeing scalability in many contexts [60].

Combining intelligent reflective surfaces with UAVs will help to improve wireless connectivity, so enabling reliable drone operations in complex industrial environments [61].

VI. FUTURE TRENDS OF DRONES IN DIGITAL TRANSFORMATION FOR INDUSTRY

- LiDAR technology and high-resolution imagery will both contribute to an improvement in mapping accuracy. Within the realms of urban planning and environmental monitoring, drones will become indispensable instruments due to their ability to assist three-dimensional modeling and terrain analysis.
- AI developments will result in more autonomous drones' capability of completing difficult jobs with little human involvement. This will improve inspections, and logistics' efficiency as well as others.
- Environmental studies, conservation projects, and animal tracking will all benefit much from drones. For assessments of biodiversity and climate research, they can offer priceless data. A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- The deployment of 5G networks will enhance drone communication capacity, therefore facilitating quicker data transfer and more consistent remote operations.

- Industries will gain from more simplified operations as regulations change to meet drone technology. This will promote more general acceptance in fields including delivery systems and emergency response.
- Training simulations for numerous industries will include drones, providing a safe and regulated space for practice and skill development.

VII. CONCLUSION

Drones are at the vanguard of digital change in a variety of industries, providing unprecedented levels of efficiency, precision, and automation. Drones improve data collecting, monitoring, and operational processes by integrating with technologies such as IoT, AI, and big data, altering industries such as logistics, manufacturing, and agriculture. Still, there are challenges on the way to fully exploit drone technology. Wider application is hampered in extreme places by operating restrictions, security issues, and regulatory limitations. Future initiatives should concentrate on enhancing these areas, as well as developing better legal frameworks and increasing drone autonomy, so fully realizing the promise of drones in industry. These advances will keep drones a major driver of digital age industrial innovation and advancement. expansion within the digital era.

REFERENCES

[1] Adepoju, O., Aigbavboa, C., Nwulu, N., Onyia, M., & Adepoju, O. (2022). Drone/unmanned aerial vehicles (UAVs) technology. Re-skilling Human Resources for Construction 4.0: Implications for Industry, Academia and Government, 65-89.

[2] Javaid, M., Khan, I. H., Singh, R. P., Rab, S., & Suman, R. (2022). Exploring contributions of drones towards Industry 4.0. Industrial Robot: the international journal of robotics research and application, 49(3), 476-490.

[3] Beke, É., Bódi, A., Katalin, T. G., Kovács, T., Maros, D., & Gáspár, L. (2018, November). The role of drones in linking industry 4.0 and ITS Ecosystems. In 2018 IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI) (pp. 000191-000198). IEEE.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[4] https://www.freepik.com/premium-ai-image/3d-flat-icon-business-as-drone-fleet-warehouse-inventory-audits-warehouse-management-inno_169414630.htm

[5] Ayman, Yafoz. (2024). A Survey on the Use of Unmanned Aerial Vehicles (UAVs) in Monitoring Applications. Journal of Electrical Systems, doi: 10.52783/jes.3547

[6] Paucar, C. *et al.* (2018). Use of Drones for Surveillance and Reconnaissance of Military Areas. In: Rocha, Á., Guarda, T. (eds) Developments and Advances in Defense and Security. MICRADS 2018. Smart Innovation, Systems and Technologies, vol 94. Springer, Cham. https://doi.org/10.1007/978-3-319-78605-6_10

[7] Refaai, M. R. A., Dattu, V. S., Gireesh, N., Dixit, E., Sandeep, C. H., & Christopher, D. (2022). [Retracted] Application of IoT-Based Drones in Precision Agriculture for Pest Control. *Advances in Materials Science and Engineering*, *2022*(1), 1160258.

[8] Campbell, H. A., Bosiocic, V., Hvala, A., Brady, M., Campbell, M. A., Skelton, K., & Luiz, O. J. (2024). Emerging Research Topics in Drone Healthcare Delivery. Drones, 8(6), 258.

[9] Haq, A. U., Muhammad, B., & Mihovska, A. (2022, October). UAV Communication Aided by Intelligent Reflective Surfaces-A Performance Analysis in the Context of Industry 4.0. In 2022 25th International Symposium on Wireless Personal Multimedia Communications (WPMC) (pp. 412-417). IEEE.

[10] Lei, M., Manchun, L., Yafei, W., Lihua, T., & Liang, C. (2013, August). Using high-resolution imagery acquired with an autonomous unmanned aerial vehicle for urban construction and planning. In *2013 the International Conference on Remote Sensing, Environment and Transportation Engineering (RSETE 2013)* (pp. 199-202). Atlantis Press.

[11] Patchou, M., Sliwa, B., & Wietfeld, C. (2019, December). Unmanned aerial vehicles in logistics: Efficiency gains and communication performance of hybrid combinations of ground and aerial vehicles. In 2019 IEEE Vehicular Networking Conference (VNC) (pp. 1-8). IEEE.

[12] Roca-Riu, M., & Menendez, M. (2019). Logistic deliveries with drones: State of the art of practice and research. In 19th Swiss Transport Research Conference (STRC 2019). STRC.S. Liu, "Wi-Fi Energy Detection Testbed (12MTC)," 2023, gitHub repository. [Online]. Available: https://github.com/liustone99/Wi-Fi-Energy-Detection-Testbed-12MTC

[13] Gao, X., Bai, X., & Zhou, K. (2023, December). Research on the Application of Drone Remote Sensing and Target Detection in Civilian Fields. In 2023 International Conference on High Performance Big Data and Intelligent Systems (HDIS) (pp. 108-112). IEEE.

[14] Zhai, X., Shao, M., Cang, M., Cheng, X., Wu, S., & Wang, J. (2024, July). Application of drone technology in power grid engineering. In Journal of Physics: Conference Series (Vol. 2798, No. 1, p. 012007). IOP Publishing.

[15] Frid, M., & Frid, V. (2024). Advancements in Drone-Based (UAV) Ground Penetrating Radar for Accurate Boundary Mapping between Disturbed Clayey Soil and Natural Rock.

[16] Bärfuss, K. B., Schmithüsen, H., & Lampert, A. (2023). Drone-based meteorological observations up to the tropopause–a concept study. Atmospheric Measurement Techniques, 16(15), 3739-3765.

[17] Guay, B., Fengler, M., & Hammerschmidt, L. (2023, May). Improving nowcasts and forecasts via operational use of meteomatics meteodrones. In EGU General Assembly Conference Abstracts (pp. EGU-8632).

[18] Zhang, B., Tang, L., & Roemer, M. (2014). Probabilistic weather forecasting analysis for unmanned aerial vehicle path planning. Journal of Guidance, Control, and Dynamics, 37(1), 309-312.

[19] M., P.., K., S.., Ghuge, S.A., Nyaharkar, S.N., Tupe, A.A. (2023). Enhancing Research Outcomes through Drone-Generated Imagery and Photogrammetry Software Analysis. Indian Scientific Journal Of Research In Engineering And Management, doi: 10.55041/ijsrem24172

[20] Quamar, M. M., Al-Ramadan, B., Khan, K., Shafiullah, M., & El Ferik, S. (2023). Advancements and applications of drone-integrated geographic information system technology—A review. *Remote Sensing*, *15*(20), 5039.

[21] Nooralishahi, P., Ibarra-Castanedo, C., Deane, S., López, F., Pant, S., Genest, M., ... & Maldague, X. P. (2021). Drone-based non-destructive inspection of industrial sites: A review and case studies. *Drones*, *5*(4), 106.

[22] Galar, D., Kumar, U., & Seneviratne, D. (2020). *Robots, drones, UAVs and UGVs for operation and maintenance*. CRC Press.

[23] Basescu, M., & Moore, J. (2020, May). Direct nmpc for post-stall motion planning with fixed-wing uavs. In *2020 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 9592-9598). IEEE.

[24] Du, H., & Heldeweg, M. A. (2017). Responsible design of drones and drone services. *Synthetic Report (October 2, 2017)*.

[25] Yi, M., Wang, X., Liu, J., Zhang, Y., & Bai, B. (2020, July). Deep reinforcement learning for fresh data collection in UAV-assisted IoT networks. In IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (pp. 716-721). IEEE.

[26] Li, X., & Savkin, A. V. (2021). Networked unmanned aerial vehicles for surveillance and monitoring: A survey. Future Internet, 13(7), 174.

[27] Chen, Y., Aggarwal, P., Choi, J., & Kuo, C. C. J. (2017, December). A deep learning approach to drone monitoring. In 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) (pp. 686-691). IEEE

[28] Dell'Amico, M., Montemanni, R., & Novellani, S. (2021). Drone-assisted deliveries: New formulations for the flying sidekick traveling salesman problem. *Optimization Letters*, *15*, 1617-1648.

[29] Faustino, R., da Silva, G. L. L., Coelho, F. M., Ovalle, A., Santos, J. L. B., Silva, C. L., ... & dos Santos, D. R. M. (2023, October). Digital Transformation in the O&G Industry with the Use of Drones. In Offshore Technology Conference Brasil (p. D031S031R005). OTC.

[30] Ranjan, C., Chakraborty, B., Srinivas, J., & Kumar, K. (2024). Drone Swarms in Industry 5.0. In Drone Applications for Industry 5.0 (pp. 31-45). IGI Global.

[31] Aydin, H., & Temizel, C. (2022, April). Latest Applications of UAVs, Drones and Robotics in the Oil and Gas Industry. In Offshore Technology Conference (p. D021S019R007). OTC.

[32] Pensec, W., Espes, D., & Dezan, C. (2022, June). Smart anomaly detection and monitoring of industry 4.0 by drones. In 2022 International Conference on Unmanned Aircraft Systems (ICUAS) (pp. 705-713). IEEE.

[33] P., Srividya. (2024). Industry 5.0 and Industrial Drones. Advances in computational intelligence and robotics book series, 1-16. doi: 10.4018/979-8-3693-2093-8.ch001

[34] Yigit, K. A., Dalkiran, A., & Karakoc, T. H. (2023). Applications of Drones in the Health Industry. *Unmanned Aerial Vehicle Design and Technology*, 69-93.

[35] Nar, D., & Kotecha, R. (2022). Enhancement of Drone-as-a-Service Using Blockchain and AI. *International Journal of Next-Generation Computing*, *13*(4).

[36] Faustino, R., da Silva, G. L. L., Coelho, F. M., Ovalle, A., Santos, J. L. B., Silva, C. L., ... & dos Santos, D. R. M. (2023, October). Digital Transformation in the O&G Industry with the Use of Drones. In *Offshore Technology Conference Brasil* (p. D031S031R005). OTC.

[37] (2024). Skybound Intelligence: AI's Impact on Drone Technology. Indian Scientific Journal Of Research In Engineering And Management, 08(04):1-5. doi: 10.55041/ijsrem29819

[38] Lee, C. K. M., Zhang, S. Z., & Ng, K. K. H. (2017). Development of an industrial Internet of things suite for smart factory towards re-industrialization. Advances in manufacturing, 5, 335-343.

[39] Sharma, V., Choudhary, G., Ko, Y., & You, I. (2018). Behavior and vulnerability assessment of drones-enabled industrial internet of things (iiot). IEEE Access, 6, 43368-43383.

[40] Ghosh, S. (2019). Developing digital transformative capabilities of industrial businesses by leveraging the industrial internet of things (Doctoral dissertation, Loughborough University).

[41] Dalamagkidis, K., Valavanis, K. P., & Piegl, L. A. (Eds.). (2009). On integrating unmanned aircraft systems into the national airspace system: issues, challenges, operational restrictions, certification, and recommendations. Dordrecht: Springer Netherlands.

[42] Anthony, Harrington. (2015). Who controls the drones? [Regulation Unmanned Aircraft]. Engineering & Technology, 10(2):80-83. doi: 10.1049/ET.2015.0211

[43] Çoban, S., & Oktay, T. (2018). Legal and ethical issues of unmanned aerial vehicles. *Journal of Aviation*, *2*(1), 31-35.

[44] Esteve, J. S., & Domènech, C. B. (2017). Rights and science in the drone era: actual challenges in the civil use of drone technology. *Rights and Science: R&S*, 117-133.

[45] Hassija, V., Chamola, V., Agrawal, A., Goyal, A., Luong, N. C., Niyato, D., ... & Guizani, M. (2021). Fast, reliable, and secure drone communication: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *23*(4), 2802-2832.

[46] Ilgi, G. S., & Ever, Y. K. (2020). Critical analysis of security and privacy challenges for the Internet of drones: A survey. In *Drones in smart-cities* (pp. 207-214). Elsevier.

[47] Marin, L. (2017). The deployment of drone technology in border surveillance: Between techno-securitization and challenges to privacy and data protection 1. In *Surveillance, privacy and security* (pp. 107-122). Routledge.

[48] McEnroe, P., Wang, S., & Liyanage, M. (2022). A survey on the convergence of edge computing and AI for UAVs: Opportunities and challenges. *IEEE Internet of Things Journal*, *9*(17), 15435-15459.

[49] Wagner, H. T., & Weitzel, T. (2012). How to Achieve Operational Business-IT Alignment: Insights from a Global Aerospace Firm. *MIS Quarterly Executive*, *11*(1).

[50] Park, S., Zhang, L., & Chakraborty, S. (2016, November). Design space exploration of drone infrastructure for large-scale delivery services. In 2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD) (pp. 1-7). IEEE.

[51] Wang, G., Hollar, D., Sayger, S., Zhu, Z., Buckeridge, J., Li, J., ... & Hu, W. (2016). Risk considerations in the use of unmanned aerial vehicles in the construction industry. Journal of Risk Analysis and Crisis Response, 6(4).

[52] McEnroe, P., Wang, S., & Liyanage, M. (2022). A survey on the convergence of edge computing and AI for UAVs: Opportunities and challenges. IEEE Internet of Things Journal, 9(17), 15435-15459.

[53] Farcas, C., Farcas, E., Krueger, I. H., & Menarini, M. (2010). Addressing the integration challenge for avionics and automotive systems—from components to rich services. *Proceedings of the IEEE*, *98*(4), 562-583.

[54] Ramalingam, K., Kalawsky, R., & Noonan, C. (2011, April). Integration of unmanned aircraft system (UAS) in non-segregated airspace: A complex system of systems problem. In *2011 IEEE International Systems Conference* (pp. 448-455). IEEE.

[55] Sakurada, L., Geraldes, C. A., Fernandes, F. P., Pontes, J., & Leitao, P. (2021). Analysis of new job profiles for the factory of the future. In *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future: Proceedings of SOHOMA 2020* (pp. 262-273). Springer International Publishing.

[56] Stroud, D., & Weinel, M. (2020). A safer, faster, leaner workplace? Technical-maintenance worker perspectives on digital drone technology 'effects' in the European steel industry. *New Technology, Work and Employment*, *35*(3), 297-313.

[57] Dimitrios, J., Dimitriou., Maria, F., Sartzetaki., Aristi, G., Karagkouni. (2024). 1. Navigating Digital Transformation Challenges and Ethical Considerations. doi: 10.1016/b978-0-443-29109-8.00010-4

[58] Guryanova, A. V., Smotrova, I. V., Makhovikov, A. E., & Koychubaev, A. S. (2020). Socio-ethical problems of the digital economy: challenges and risks. In *Digital transformation of the economy: Challenges, trends and new opportunities* (pp. 96-102). Springer International Publishing.

[59] Faustino, R., da Silva, G. L. L., Coelho, F. M., Ovalle, A., Santos, J. L. B., Silva, C. L., ... & dos Santos, D. R. M. (2023, October). Digital Transformation in the O&G Industry with the Use of Drones. In *Offshore Technology Conference Brasil* (p. D031S031R005). OTC.

[60] Aydin, H., & Temizel, C. (2022, April). Latest Applications of UAVs, Drones and Robotics in the Oil and Gas Industry. In *Offshore Technology Conference* (p. D021S019R007). OTC.

[61] Haq, A. U., Muhammad, B., & Mihovska, A. (2022, October). UAV Communication Aided by Intelligent Reflective Surfaces-A Performance Analysis in the Context of Industry 4.0. In *2022 25th International Symposium on Wireless Personal Multimedia Communications (WPMC)* (pp. 412-417). IEEE.

Control of Anode Pressure and Electro-Valve Dynamics for Enhanced Stability in Hydrogen Fuel Cells

KABACHE Sabah[1], BOUREGBA Hicham[2]

[1]Laboratoire revêtement, matériaux et environnement département de physique, université de Boumerdès, Algérie.
s.kabache@univ-boumerdes.dz

[2]Laboratoire Energétique Mécanique & Ingénierie, Faculté de Technologie, Université de Boumerdes, Algérie.
mr.bouregbahicham@gmail.com

***Abstract*— In a Proton Exchange Membrane Fuel Cell (PEMFC) setup, maintaining precise pressure levels of the anode (H2) sides is crucial, particularly during load variations. Failure to control pressure appropriately can result in imbalances that may harm the fuel cell stack's delicate membrane. This study proposes a control approach using a conventional PID regulator to keep membrane pressure stable within acceptable ranges. The effectiveness of this method was tested on a 5 kW PEMFC stack model subjected to 50 A current jumps, with simulations conducted in Matlab/Simulink. Results indicate that employing conventional regulation effectively mitigated disturbances caused by load changes and when the setpoint changed, maintaining satisfactory performance metrics (response time, overshoot, stability), as evidenced by indices like integral absolute error (IAE), integral time absolute error (ITAE), and integral square error (ISE).**

***Keywords*— PEMFC, H2 pressure, conventional control.**

I. INTRODUCTION

Fuel cells (FC) are devices that generate electricity by directly converting the electrochemical energy of hydrogen fuel. Widely appreciated in industrial and research sectors, this technology has gained significant attention for its efficiency and environmental benefits. Among the various types, Proton Exchange Membrane Fuel Cells (PEMFCs) have emerged as particularly promising for both stationary and mobile applications. PEMFCs are renowned for their compact and lightweight design, high power density, and capability to function efficiently at relatively low operating temperatures, making them a versatile and attractive option compared to other fuel cell systems. [1][2][3].

Numerous studies in the scientific literature have investigated various linear and non-linear control strategies for the modeling and regulation of fuel cells. Prominent research works in this domain include: Andrés Moran Doran et al. [4] introduces a hybrid approach utilizing neural networks for modeling and controlling fuel cell systems. It emphasizes the significance of accounting for factors beyond reactant supply to enhance fuel cell performance. The findings reveal that a neuro-PID controller, on its own, is insufficient to stabilize voltage without incorporating inverse model control. Kai Ou et al. [5] present a feedforward fuzzy-PID controller designed to regulate the oxygen excess ratio in PEM fuel cell systems. This approach efficiently manages the oxygen excess ratio while minimizing parasitic power losses. The study also highlights the integration of PID and fuzzy logic, offering a simple yet effective solution for implementing air flow control. Nicu Bizon et al. [6] introduced a fuel-saving strategy for Proton Exchange Membrane Fuel Cell (PEMFC) systems. This approach involves transitioning the load-following mode to the fueling regulators, optimizing fuel consumption efficiency. The proposed strategy achieves a reduction in fuel consumption by 14–29% compared to conventional commercial methods.

In the realm of process control, Proportional-Integral-Derivative (PID) controllers continue to be one of the most widely adopted solutions owing to their straightforward design, ease of implementation, and proven effectiveness across a broad range of applications. Their versatility makes them suitable for managing various industrial processes, from temperature and pressure control to flow regulation. Despite their widespread use, the performance of PID controllers heavily depends on accurate tuning of their parameters, which determine how well the system responds to changes and disturbances [7][8].

In this study, the conventional method of utilizing a PID controller is presented as a foundational approach. The PID controller's performance is meticulously adjusted to optimize gas consumption efficiency and ensure the stability of power output at rated values, even in the presence of fluctuating load conditions. Maintaining stable pressure at the oxygen and hydrogen terminals is a critical aspect of this control strategy, as any variations in these pressures can significantly impact the overall lifespan and operational efficiency of the fuel cell system. These pressure variations affect both the durability of the membranes and the efficiency of the electrochemical reactions, making precise control essential for long-term system performance.

To thoroughly evaluate the effectiveness of the conventional PID controller, its transient performance is analyzed using a range of performance indices. These indices provide valuable insights into the controller's responsiveness, stability, and ability to adapt to dynamic changes in load. By assessing these factors, the study aims to determine the PID controller's capability to maintain optimal operation under varying conditions and highlight areas for potential improvement or adaptation in more complex systems.

II. SYSTEM DESIGN AND MODEL BUILDING

The objective of this article is to develop a fuel cell anode pressure control strategy that is more effectively aligned with practical technical applications. This strategy seeks to advance fuel cell control systems, particularly those with high power ratings. To achieve this, a 5-kW fuel cell system has been designed as part of the study.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

A. *Fuel cell system design*

A PEM fuel cell system is shown in Fig. 1 operates by utilizing hydrogen, which require humidification prior to entering the stack. Heat generated during operation must be dissipated through either natural or forced cooling methods to prevent membrane flooding. Consequently, a standard fuel cell system includes various auxiliary components such as hydrogen storage tanks, pressure regulators, humidifiers, air filters, heat exchangers, compressors, water separators, relief valves, pumps, and water tanks, etc. [9][10].

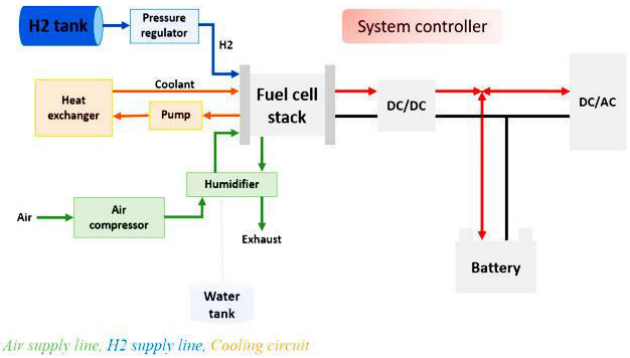


Fig. 1. Fuel cell system [10]

B. *Modelling of the fuel cell system*

The mathematical model of the fuel cell system incorporates several components, including the stack, anode gas supply system, and hydrothermal management system [10]. To clearly present and explain key system characteristics and facilitate controller design, the following assumptions are made in the development of the fuel cell system's mathematical model.

- It is assumed that all gases behave as ideal gases and follow the ideal gas law.
- The focus is solely on changes in gas inflow and outflow, with the assumption of a uniform distribution of gas composition throughout the system.
- The relative humidity of the air entering the cathode is set at 0.7, and the stack temperature is maintained at 65°C.
- Nitrogen permeation from the cathode to the anode is considered negligible.

This paper provides an in-depth analysis of a 5-kW fuel cell stack consisting of 35 individual cells, each with an active surface area of 232 cm2. The total output voltage of the stack exceeds 19 V, which corresponds to approximately 0.53 V per cell, and it is capable of delivering a maximum current of 300 A. The study is particularly focused on evaluating the system's performance during hydrogen (H2) operation, with pure hydrogen supplied to the anode. The investigation aims to assess the stack's efficiency, stability, and response to varying load conditions, with a specific emphasis on optimizing the fuel cell's performance and longevity under real-world operational scenarios [11][12]. Relevant parameter values are given in Table I.

TABLE I. PEMFC MODEL'S PARAMETERS

Parameters	Value	Unit
T	353	K
N_{fc}	35	
t_m	175x10^{-4}	cm
F	96485	C kmol^{-1}
R	8.31415	J kmol^{-1} K^{-1}
K_{O2}	2.11e-5	Kmol S^{-1} atm
K_{H2}	4.22e-5	Kmol S^{-1} atm
τ_{O2}	3.37	(s)
τ_{H2}	6.47	(s)
ξ_1	-0.944	
ξ_2	0.00354	
ξ_3	8x10^{-8}	
ξ_4	-1.96x10^{-4}	

C. *Output voltage model*

Figure 1 illustrates the fuel cell as an electrochemical device that efficiently converts the chemical energy stored in hydrogen into electrical power. The fuel cell operates based on an electrochemical redox reaction [13], which facilitates the interaction between hydrogen and oxygen to generate electricity, water, and heat. This process is summarized by the following overall chemical equation.

$$2H_2 + O_2 \rightarrow 2H_2O + Heat + energy \quad (1)$$

The output voltage of the fuel cell can be expressed using the following equation [14].

$$V_{FC} = E_{Nerst} - V_{act} - V_{ohm} - V_{conc} \quad (2)$$

E_{Nerst} is the reversible open-circuit voltage, it is described by the Nerst equation as:

$$E_{Nerst} = 1.229 - 8.5 * 10^{-4}(T_{fc} - 298.15) + 4.3085 * 10^{-5} * T_{fc}[\ln(P_{H2}) + 0.5(P_{O2})] \quad (3)$$

Where, P_{H2} hydrogen pressure (atm), P_{O2} Oxygen pressure (atm) and T Absolute temperature (K).

V_{act} denotes the activation voltage drop, as expressed in the Tafel equation as:

$$V_{act} = \xi_1 + \xi_2 T + \xi_3 T \ln(C_{O_2}) + \xi_4 T \ln I_{FC} \quad (4)$$

Here, I_{FC} denotes the fuel cell current (A), and ξ_i (where i=1- 4) represent parametric coefficients corresponding to each cell model.

CO_2 denotes the concentration of dissolved oxygen at the interface of the cathode catalyst, which can be determined through the following calculation:

$$C_{O_2} = \frac{P_{O_2}}{5.08 \times 10^6 \exp\left(\frac{-498}{T}\right)} \quad (5)$$

The total ohmic voltage loss, often denoted as V_{ohm}, can be formulated as:

$$V_{ohm} = IR_{mem} \quad (6)$$

Where Rm the membrane resistance is determined by dividing the thickness, t_{mem}, by the conductivity σ_{mem} as described by the following equation:

$$R_{mem} = \frac{t_{mem}}{\sigma_{mem}} \quad (7)$$

V_{conc} represents the concentration voltage drop, which is defined as:

$$V_{conc} = -B \ln\left(1 - \frac{J}{J_{\max}}\right) \quad (8)$$

Where, J_{max} is the maximum current density of the cell, J (A/cm2) represents the actual current density of the cell and B is a parametric coefficient, which depends on the cell and its operation state.

When N_{FC} fuel cells are linked in series, the fuel cell stack's output voltage is determined by:

$$V_{stack} = N_{FC} V_{FC} \quad (9)$$

The characteristic curve for this model is presented as follows:

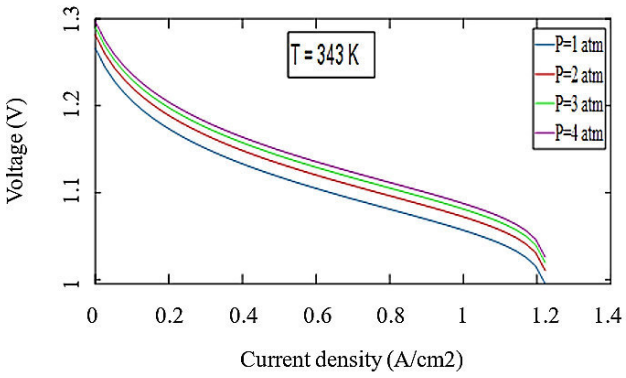


Fig. 2. characteristic curve of output voltage

D. Dynamic gas transport model

The hydrogen partial pressure can be determined by considering the input flow rate and the current of the fuel cell, establishing a functional relationship between these variables [15]:

$$P_{H2} = [(1/K_{H2})/(1 + \tau_{H2}S)] * \left(q_{H2}^{in} - 2K_r I\right) \quad (10)$$

In this context, q_{H2} represents the input hydrogen flow, while k_r signifies the hydrogen flow that is consumed. This consumption can be mathematically expressed as follows:

$$K_r = \frac{N_{fc}}{4F} \quad (11)$$

Using identical methods, the oxygen partial pressure is articulated as:

$$P_{H2} = [(1/K_{O2})/(1 + \tau_{O2}S)] * \left(q_{O2}^{in} - K_r I\right) \quad (12)$$

where in qO2 is the input oxygen flow, and k_r is the hydrogen flow which is consumed, and can be expressed as:

$$K_r = \frac{N_{fc}}{2F} \quad (13)$$

Equations (13) and (15) delineate the correlation between the current of the fuel cell and the partial pressures of hydrogen and oxygen, respectively. As the load initiates current flow, both hydrogen and oxygen reactants rise, leading to a corresponding decrease in their partial pressures.

III. CONTROL OF THE FUEL CELL SYSTEM

The conventional PID controller depicted in Fig. 3 is extensively employed in industrial process control owing to its uncomplicated design and effectiveness in managing both linear and non-linear systems.

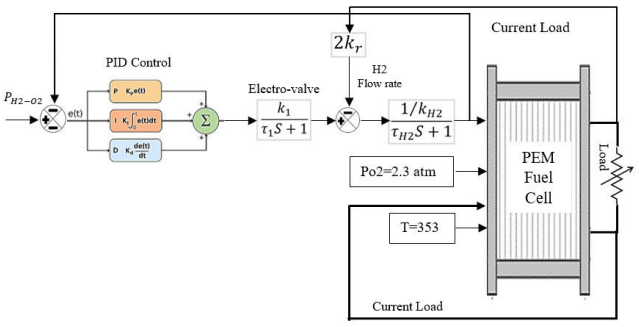


Fig. 3. Control strategy of the H2 pressure

The controller is given by the mathematical and canonical equation as follows [16]:

$$P(S) = K_p + \frac{K_i}{S} + K_d S \quad (14)$$

The pressure control loop consists of a direct chain and a unitary return chain. The open-loop transfer function (OLTF) is as follows:

$$OLTF = (K_p + \frac{K_i}{S} + K_d S) * (\frac{1/K}{\tau_r S + 1}) \quad (15)$$

The provided expression represents the closed-loop transfer function (CLTF):

$$CLTF = \frac{OLTF}{1 + OLTF} \quad (16)$$

$$CLTF = \frac{K}{(\tau_r + k_d)S^2 + (k_p + 1)S + k_i} \quad (17)$$

Identifying the transfer function parameters using the canonical form of a second-order system proceeds as follows:

$$\frac{K}{S^2 + 2\xi\omega_n S + \omega_n^2} \quad (18)$$

The calculated controller gains, which presented in the table below:

TABLE II. MANUEL REGULATION PARAMETERS

Parameters		Values
K_p		$2\xi\omega_n - 1$
K_i		ω_n^2
K_d		$1 - \tau_r$
Hydrogen regulator	Manuel regulation	Kp=1.9
		Ki=1
		Kd=0.15

IV. SIMULATION RESULT AND DISCUSSION

As part of the modeling study in Matlab/Simulink, the aim was to evaluate the performance of the "PEMFC" fuel cell. We examined the effects of varying the load while keeping the pressure constant. Other part, the robustness testing is evaluated when the pressure is change in the circumstance conditions. This configuration is crucial for understanding the complex interactions between load and pressure, providing valuable insight into system performance. Analyses of the Simulation Results like anode pressure (H2) and flow rate (H2), load current and power consumption are presented as follows:

The variation in load is outlined as follows:

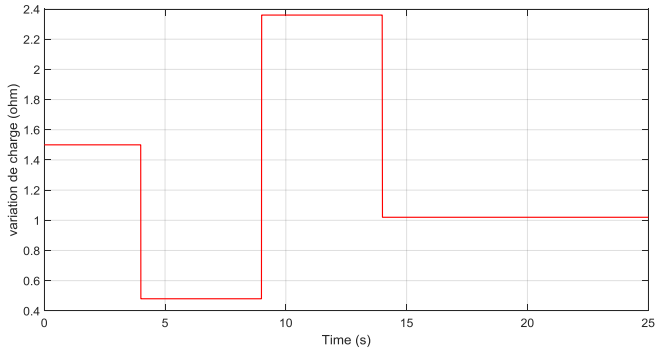


Fig 4. Load profile

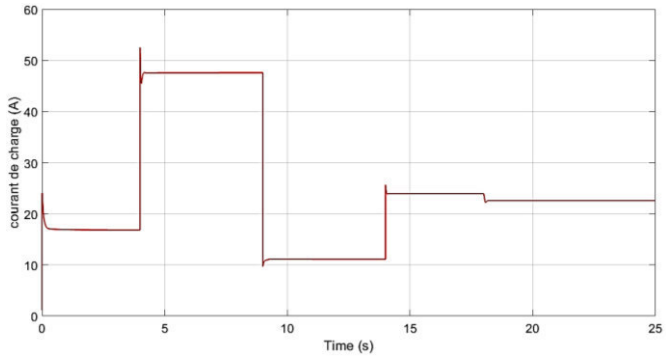


Fig. 5. Variation current load

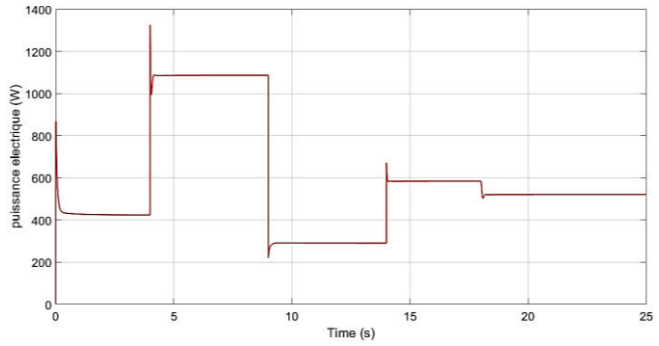


Fig. 6. Power rate profile

The load demand variation shown in the figure is made between [0-25s]. The pressure at the anode terminal is kept constant P=2.3 atm between [0-18s] where the load demand is variable, following a pressure variation at t= 18s where the load is constant (R=1.5Ω). The variation in load current and power shown in Fig. 2 and Fig. 3 are a direct consequence of the load demand between [0-18s], or the pressure variation [18-25s] giving no effect on the previous profiles. Conventional PID-based control shows how the control chain reacts in real time to meet energy requirements by adjusting the fuel cell's output.

Table III presents a performance analysis of power profile variations as a function of load demand, as follows:

TABLE III. ANALYSIS VARIATION OF POWER RATE

Time (s)	[0-4]	[4-8]	[8-14]	[14-18]	[18-25]
Response time (s)	0.2	0.001	0.001	0.001	0.001
Overshoot	18%	15%	5%	10%	3%
Stability	98%	98%	98%	98%	98%

Examination of the results presented in Table III reveals that the control chain is capable of maintaining stable and precise regulation through the use of a PID controller. The measurements taken show that the control chain is able to respond rapidly to load variations, while maintaining constant levels of current, hydrogen and oxygen flow (see Fig. 5, Fig. 6, Fig. 7 and Fig. 8). In addition, electrical energy production (Fig. 3) is also maintained at stable levels, even in the presence of load variations. These results confirm the efficiency and reliability of the PID controller's control chain. What's more, the pressure variation between [18-25s] shows no change in electrical power or flow rate (H2).

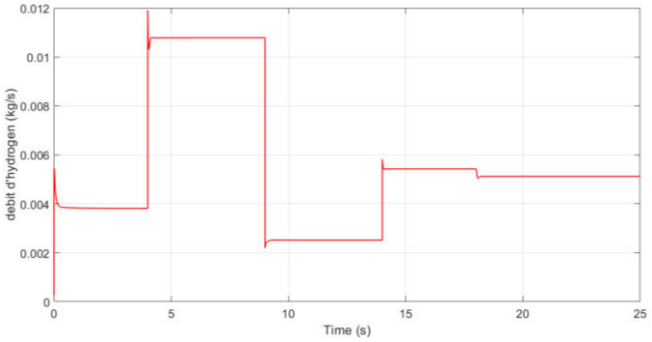


Fig. 7. Hydrogen flow rate profile

We carried out an in-depth analysis of the PID controller-based pressure control system, which involved assessing its ability to handle load variations over a time range of 14 seconds while maintaining a constant pressure P=2.3 atm. In addition, we incorporated a pressure setpoint change at t=18s with a new pressure setpoint (P=1 atm) to test its response to sudden changes. This gave a good result concerning the robustness of the PID-controlled system.

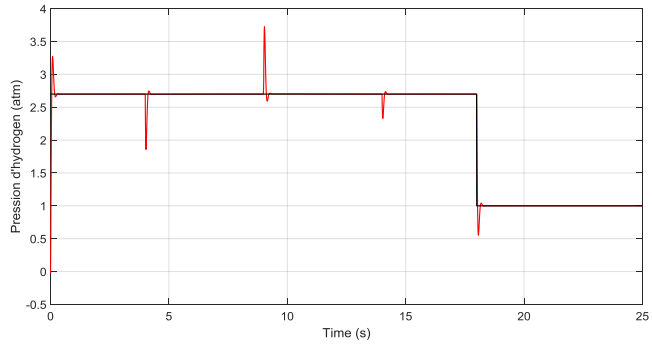


Fig. 8. Hydrogen pressure profile

Fig. 8 shows that the control system remained stable and was able to closely follow the setpoint, demonstrating its reliability in regulating pressure under different conditions. Overall, these results are significant, as they suggest that the PID controller is able to respond effectively to changing process conditions and maintain the desired setpoint.

Performance analysis of H2 pressure:

Performance analysis is based on anode pressure, as depicted in Fig. 8. Optimal control of this factor is crucial for the longevity of the fuel cell and its components, while ensuring good energy efficiency. The performance analyses are presented in the table below:

TABLE IV. ANALYSES PERFORMANCE OF H2 PRESSURE

Parameters		Performances	
Hydrogen pressure	Variations pics	A t=4	P=0.5 atm
		A t=8	P=1 atm
		A t=14	P=0.25 atm
		A t=18	P=0.3 atm
	Manuel regulation	Response time (s)	0.05
		Overshoot (%)	2%
		Stability	0.01

Analyses:

The characteristics provided for hydrogen conventional (PID) pressure regulation, including a response time of [0.05s] for H2, an overshoot of [2%], and stability at [0.01], collectively indicate a highly efficient and reliable system for fuel cell operation. The rapid response time ensures swift adjustments to changes in demand, minimizing stress on components and potential damage to the fuel cell membrane. The low overshoot signifies precise control, reducing fluctuations around the setpoint and safeguarding against excessive pressure variations. Moreover, the tight stability ensures consistent operating conditions, contributing to prolonged system lifetime and enhanced performance. In summary, these characteristics reflect a pressure regulation system that optimally maintains hydrogen pressure levels, mitigating risks and fostering the longevity and efficiency of the fuel cell system.

To develop the study and give more accuracy of regulated system, we analyzed the stability of the controller relative to its reference using many indicators: Integral Absolute Error (IAE), Integral Time Absolute Error (ITAE), Integral Square Error (ISE) as presented in the table below:

TABLE V. STABILITY ANALYSES OF STEADY STATE

Parameters		IAE	ITAE	ISE
PID manual regulation	Hydrogen pressure	0.01	0.037	0.0016

Integral Absolut Error (IAE), Integral Time Absolut Error (ITAE), Integral Square Error (ISE)

The comparison between PID tuned regulation and manual PID regulation across various control scenarios reveals significant differences in stability and performance. In all cases hydrogen pressure, oxygen pressure, and DC voltage the PID tuned regulator consistently exhibits lower values for stability indices (IAE, ITAE and ISE) compared to manual regulation. This indicates superior control performance and reduced oscillations, highlighting the effectiveness of PID tuning, especially when assisted by techniques like PSO. Overall, the results underscore the importance of optimizing control strategies to achieve better stability and performance in diverse control scenarios, ultimately enhancing system efficiency and reliability.

V. CONCLUSION

In this study, we investigate into the dynamic modeling, control, and simulation of PEMFC. Initially, we introduce the dynamic model of PEMFC based on existing literature, incorporating various evolutions of physical parameters. Subsequently, we outline the approach for calculating the flow rates of hydrogen. The efficacy of control strategy for the proton exchange membrane fuel cell system across a wide operating range is confirmed through simulations conducted using Matlab/Simulink software. The simulation results of the H2 pressure control loop, based on conventional PID controllers has been examined. The findings demonstrate that the conventional regulator significantly enhances system performance. This improvement has been evidenced through the analysis of performance indices such as IAE, ITAE, and ISE, particularly in managing load variation peaks. This optimization thereby aids in ensuring constant pressure at the PEMFC membrane terminals.

REFERENCES

[1] G. Bhansali and R. Kumar, "Design analysis and dynamic control of PEM fuel cell for standalone applications," 2014 IEEE Students' Conf. Electr. Electron. Comput. Sci. SCEECS 2014, 2014, doi: 10.1109/SCEECS.2014.6804455.

[2] Y. Li, X. Zhao, S. Tao, Q. Li, and W. Chen, "Experimental Study on Anode and Cathode Pressure Difference Control and Effects in a Proton Exchange Membrane Fuel Cell System," Energy Technol., vol. 3, no. 9, pp. 946–954, 2015, doi: 10.1002/ente.201500077.

[3] F. Gonzatti, M. Miotto, and F. A. Farret, "Proposal for Automation and Control of a PEM Fuel Cell Stack," J. Control. Autom. Electr. Syst., vol. 28, no. 4, pp. 493–501, 2017, doi: 10.1007/s40313-017-0322-2.

[4] R. Geethanjali and R. Sivakumar, "Design of intelligent controller for PEM fuel cell," ICONSTEM 2017 - Proc. 3rd IEEE Int. Conf. Sci. Technol. Eng. Manag., vol. 2018-Janua, no. V, pp. 1050–1053, 2017, doi: 10.1109/ICONSTEM.2017.8261316.

[5] H. Marzougui, A. Kadri, M. Amari, and F. Bacha, "Improvement of energy management algorithm for fuel cell electrical vehicle with fuzzy logic," 2017 18th Int. Conf. Sci. Tech. Autom. Control Comput. Eng. STA 2017 - Proc., vol. 2018-Janua, pp. 212–217, 2017, doi: 10.1109/STA.2017.8314963.

[6] Y. Zhu, J. Zou, C. Peng, Y. Xie, and L. Li, "Modelling and fuel flow control of PEMFC considering over-pressure case," Proc. - 2017 Chinese Autom. Congr. CAC 2017, vol. 2017-January, pp. 2222–2225, 2017, doi: 10.1109/CAC.2017.8243143.

[7] Y. Qi, M. Thern, M. Espinoza-Andaluz, and M. Andersson, "Modeling and Control Strategies of Proton Exchange Membrane Fuel Cells," Energy Procedia, vol. 159, pp. 54–59, 2019, doi: 10.1016/j.egypro.2018.12.017.

[8] H. P. Regulation, "PEM Fuel Cell Voltage Neural Control Based on," 2019.

[9] Y. Rao, Z. Shao, A. H. Ahangarnejad, E. Gholamalizadeh, and B. Sobhani, "Shark Smell Optimizer applied to identify the optimal parameters of the proton exchange membrane fuel cell model," Energy Convers. Manag., vol. 182, no. December 2018, pp. 1–8, 2019, doi: 10.1016/j.enconman.2018.12.057.

[10] H. C. Nejad, M. Farshad, E. Gholamalizadeh, B. Askarian, and A. Akbarimajd, "A novel intelligent-based method to control the output voltage of Proton Exchange Membrane Fuel Cell," Energy Convers. Manag., vol. 185, no. January, pp. 455–464, 2019, doi: 10.1016/j.enconman.2019.01.086.

[11] H. Deng, Q. Li, Y. Cui, Y. Zhu, and W. Chen, "Nonlinear controller design based on cascade adaptive sliding mode control for PEM fuel cell air supply systems," Int. J. Hydrogen Energy, vol. 44, no. 35, pp. 19357–19369, 2019, doi: 10.1016/j.ijhydene.2018.10.180.

[12] Y. Cao, Y. Li, G. Zhang, K. Jermsittiparsert, and N. Razmjooy, "Experimental modeling of PEM fuel cells using a new improved seagull optimization algorithm," Energy Reports, vol. 5, pp. 1616–1625, 2019, doi: 10.1016/j.egyr.2019.11.013.

[13] M. K. Singla, A. S. Oberoi, and P. Nijhawan, "Trends so far in hydrogen fuel cell technology: State of the art," Int. J. Adv. Trends Comput. Sci. Eng., vol. 8, no. 4, pp. 1146–1155, 2019, doi: 10.30534/ijatcse/2019/23842019.

[14] H. He, S. Quan, and Y. X. Wang, "Hydrogen circulation system model predictive control for polymer electrolyte membrane fuel cell-based electric vehicle application," Int. J. Hydrogen Energy, vol. 45, no. 39, pp. 20382–20390, 2020, doi: 10.1016/j.ijhydene.2019.12.147.

[15] T. Pangaribowo, "Modeling Of Control System for Hydrogen and Oxygen Gas Flow into PEMFC Based on Load Demand by Using Fuzzy Controller," Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 1.4, pp. 606–611, 2020, doi: 10.30534/ijatcse/2020/8491.42020.

[16] A. Polak, "Simulation of fuzzy control of oxygen flow in PEM fuel cells," Energies, vol. 13, no. 9, pp. 15–25, 2020, doi: 10.3390/en13092372.

A Novel Control to Enhance Solar Energy with Electrical Grid Integration

1st *Menzri Fatima*
University of L'arbi Ben M'hidi, Oum Bouaghi, Laboratory of LGEA, Algeria
menzri.fatima@univ-oeb.dz,

2nd *Naamane Debdouche*
University of Constantine 1, Laboratory of Electrical Constantine LEC, Algeria
Naamane.dabdouche@student.umc.dz,

Abstract— **Injecting solar energy from solar power generators into the electrical grid is a crucial strategy for advancing clean energy integration. This process leverages innovative control techniques, notably sliding mode control and direct power control. Sliding mode control enables precise tracking of the maximum power point for solar energy flow, aligning it with grid requirements to ensure stability and a reliable power supply. Simultaneously, direct power control enhances solar energy conversion, minimizing energy losses and maximizing efficiency during grid injection. Together, these control strategies enable seamless and efficient solar energy integration, promoting sustainable energy practices and reducing dependence on conventional energy sources. The obtained results confirmed the robustness and effectiveness of the proposed approach.**

Keywords-PV; Electrical grid; MPPT; SMC; DPC

I. INTRODUCTION

The integration of solar power into the electrical grid is a crucial and transformative step in the global transition toward cleaner and more sustainable energy sources [1]. Solar energy, harnessed from the sun's abundant and renewable resources, has emerged as a leading contender in the effort to reduce greenhouse gas emissions and combat climate change [2]. By connecting solar power systems to the electrical grid, we are unlocking the potential to generate clean electricity on a massive scale, making it accessible to communities, businesses, and industries worldwide [3].

This integration represents a pivotal shift in our energy landscape, as it offers numerous benefits beyond environmental sustainability [4]. Solar power reduces our dependence on fossil fuels, enhances energy security, and creates economic opportunities by fostering innovation and job growth in the renewable energy sector. Moreover, it allows for greater energy independence and resilience in the face of power outages or disruptions [5].

Connecting renewable energy generators to the electrical grid represents a significant exploration due to its profoundly positive impact on the environment [1-6]. However, it involves several complexities and challenges in harnessing these sources efficiently and dynamically, from electricity generation to consumer distribution [7]. Among the challenges faced is the inherent variability of renewable energy sources compared to the electrical grid, resulting in an imbalance in the overall system. This has led researchers to explore effective methods and technologies for injecting, managing, and optimizing energy resources [4-7-8]. These systems are often referred to as "smart grids.

In this research paper, we proposed an integrated method to improve the performance of solar energy production and its injection into the electrical network. The sliding mode controller SMC worked to track the maximum power point MPP of energy with high efficiency, no matter how the climatic conditions changed. On the other hand, the produced energy is injected into the electrical network through the direct power control DPC unit that is connected to the three-level inverter.

The SMC is a very effective control strategy, widely used in dynamic control systems, including applications in solar energy systems. Its main function is to adjust the system so that it precisely follows a specified reference value. In the context of solar energy, tracking the system's maximum power point is crucial, and the sliding mode controller is key to achieving this.

This controller technology improves the efficiency of solar energy use by optimizing the performance of the solar panels. It does this by regulating the electrical current generated by the solar cells to match the desired target value. The precise energy control facilitated by the sliding mode controller is essential to ensure that the solar energy system operates at its maximum power point, where energy production is most efficient. This level of control is essential to harness the full potential of solar energy and maximize the overall efficiency and performance of solar energy systems.

Direct Power Control (DPC) is a crucial control strategy in power electronics, particularly in the field of renewable energy systems. Its importance lies in its ability to efficiently convert power, providing fast and accurate regulation of active and reactive power in the transfer between the source (such as a renewable energy system) and the load. DPC plays a key role

in integrating intermittent energy sources such as solar and wind power into the grid, ensuring optimal energy utilization. Thanks to its fast and dynamic response, DPC is well suited to applications requiring rapid adjustments to maintain system stability. It also minimizes harmonic distortion in the output waveform, contributing to cleaner power transmission. DPC's simplified control structure allows for easier implementation and reduced hardware complexity, potentially reducing costs. It improves network performance by regulating the power factor and ensuring stability under various conditions, thereby supporting the reliability of the power distribution system. In addition, the adaptability of the DPC to different loads makes it versatile and applicable in a variety of contexts, from industrial processes to residential power systems.

Through the results obtained and after analyzing them, we confirmed the effectiveness of the method in improving the system's performance starting from production all the way to injecting the produced energy into the network without any major energy losses.

II. DESIGN OF INEGRATING SOLAR ENERGY INTO THE GRID VIA SMC AND DPC CONTROL METHODS

Figure1 illustrates the design of the proposed system in the research paper and the two methods employed to enhance the system's performance. This system consists of a solar power generator connected to a DC bus through a DC-DC boost converter, which is later linked to the electrical network via a three-level inverter. Filters are used to reduce harmonics when transferring energy and improve its quality when injected into the grid.

A sliding mode control SMC [9] is adopted to track the maximum power point MPP produced, despite changing weather conditions. The figure shows that the proposed control unit relies on the reference voltage value, which is estimated using an perturb and observe P&O [10] algorithm, then SMC directly influences the switch of the converter, thus affecting its dynamic behavior. After improving the generated power, it is injected into the electrical network with minimal power losses by utilizing a direct power control unit DPC. This unit estimates the active and reactive power values of the grid and compares them with the reference power values for both types. The method will be detailed further in the research paper.

TABLE 1. PV SYSTEM PARAMETERS

G(w/m^2)	1000
T(°C)	25
Oen circuit voltage Voc(V)	43.5
Short-circuit current I_{sc}(A)	4.75
Voltage at maximum power point V_{mp}(V)	34.5
Current at maximum power point I_{mp}(A)	4.35
P(w)	150
Ns	16
Np	10
DC-DC Converter	
C_{pv} (F)	200×10-6
L(mH)	10
f(kHz)	20

TABLE2. GRID POWER PARAMETERS.

Grid, Filter, Inverter	
eg (V)	220
f(Hz)	50
R_g((Ω)	0.34
L_g(mH)	0.15
C_{dc}(F)	2000×10-6
V_{dc}(V)	800

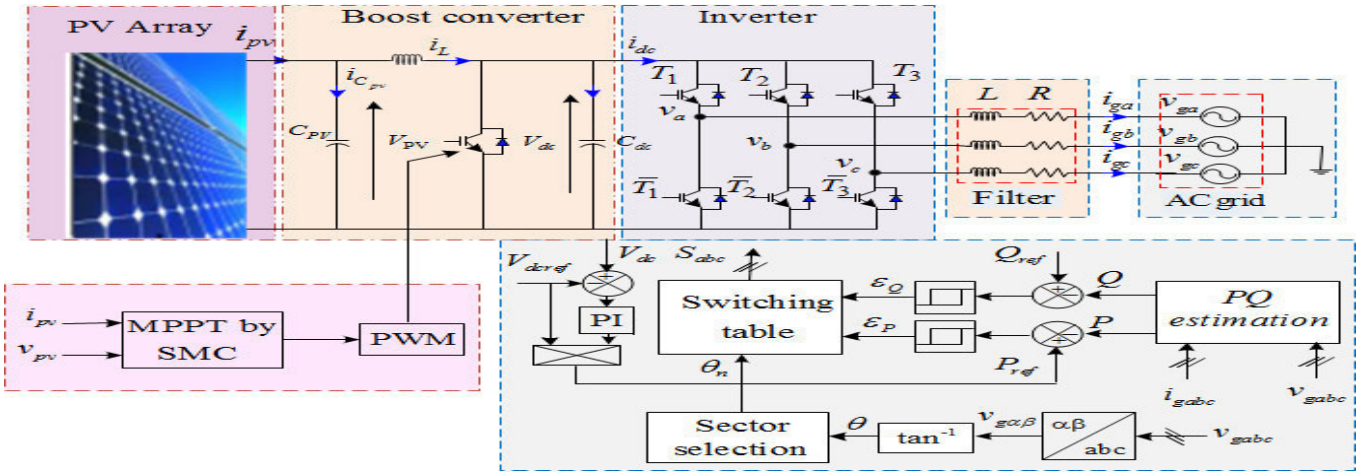


Fig.1. Scheme of the proposed system

A. *Principal of sliding mode control (SMC)*

The sliding mode control method represents a nonlinear controller, which is classified as one of the dynamic control approaches used to develop robust controllers for intricate high-order nonlinear dynamic systems operating in uncertain conditions [9],[12].

Figure 2 illustrates how to use the SMC to track the maximum power point MPP of the photovoltaic system so that the SMC directly affects the dynamic behavior of the DC/DC boost converter, which has been implemented as follows in figure 2:

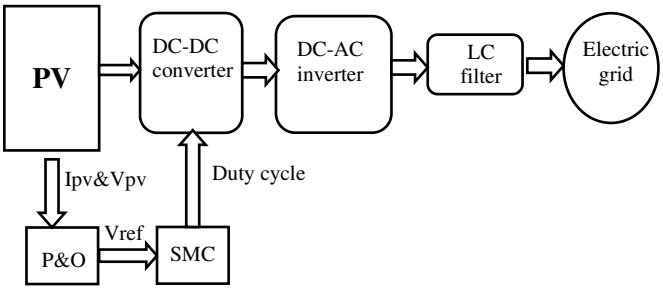


Fig.2. Block diagram of PV using SMC

PV system

Figure 3 represents the circuit equivalent of PV generator:

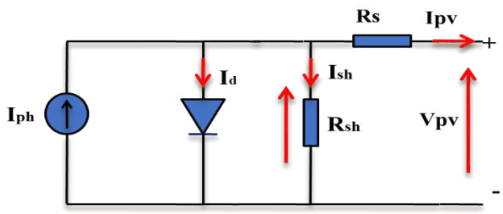


Fig.3. circuit of PV generator

The characteristic of PV cell (the current-voltage I-V) as given in [13]:

$$I_{pv} = N_p I_{ph} - N_p I_o \left[exp\left(\frac{q\left(V_{pv} + \left(\frac{N_s}{N_p}\right) R_s I_{pv}\right)}{N_s AKT} \right) - 1 \right] - \frac{V_{pv} + \left(\frac{N_s}{N_p}\right) R_s I_{pv}}{\left(\frac{N_s}{N_p}\right) R_{sh}} \quad (1)$$

Where: I_{pv}, V_{pv}: PV generator current and voltage respectively
I_{ph}, I_o: are the photo current and reverse saturation current
N_p, N_s: number of parallel modules and series cell
R_s, R_{sh}: the series and the shunt resistance
q: the electron charge ($1{,}602\times10^{-19}$C)
K: Boltzmann constant ($1{,}380649\times10^{-23}$J.K^{-1})
A: the diode ideality factor
T: the junction temperature in Kelvin (K)

The photo current and reverse saturation current can be expressed by the following equations:

$$I_{ph} = \left[I_{sc} + K_i (T - T_{ref}) \right] \frac{G}{G_{ref}} \quad (2)$$

$$I_o = I_{res} \left(\frac{T}{T_{ref}} \right)^3 \exp\left[\frac{qE_g}{AK} \left(\frac{1}{T_{ref}} - \frac{1}{T} \right) \right] \quad (3)$$

I_{sc}: *Short*-circuit current at reference conditions
K_i : Short circuit current temperature coefficient
E_g: Band-gap energy of the PV cell semiconductor
I_{res}: Saturation current at the reference temperature
G_{ref}: Reference conditions (1000W.m^{-1}) and T_{ref} (25°C)

Perturb and observe algorithm P&O was illustrated in [10], So that the value of the reference voltage Vref is estimated by tracking the maximum power point to be later compared with the actual voltage of the solar power generator under varying environmental conditions. Then, use the sliding mode controller SMC to generate a signal for turning the converter on (1) or off (0).
The analysis of SMC's dynamic behavior:
The sliding surface proposed is:

$$S = \frac{dP}{dV} = \frac{P_{pv} - P_{ref}}{V_{pv} - V_{ref}} = 0 \quad (4)$$

With: $$P_{ref} = I_{pv}.V_{ref} \quad (5)$$

This sliding surface enables us to define the duty cycle of the PV system's boost converter, as illustrate by [14]:

$$D = \begin{cases} 0 & , \ for \quad Deq + Dn \le 0 \\ D_{eq} + D_n \ , for & 0 < Deq + Dn < 1 \\ 1 & , \ for \quad Deq + Dn \ge 0 \end{cases} \quad (6)$$

With: $$D_{eq} = 1 - \frac{V_{ref}}{V_{out_pv}} \quad (7)$$

D_n represent a discrete control:

$$D_n = K.Sat(s) \quad (8)$$

While: $$Sat(s) = \begin{cases} \frac{S}{\varepsilon}, & if \ |s| < \varepsilon \\ sign(S), Otherwise \end{cases} \quad (9)$$

Where *K* is a positive constant and ε is the boundary layer thickness.

B. *Principal of direct power control (DPC)*

The figure 1 allows explaining the main idea of how a direct power controller works where the instantaneous values of the active Pg and reactive power Qg of the electrical grid are measured and compared to the reference values Pg-ref and Qg-ref, Then the difference between their values are linked to the inputs of two "hysteresis" comparators. These comparators, with the assistance of a "switching table" and considering the mains voltage, determine the switching states of the switches [15-16].

To estimate the instantaneous values of active and reactive power, we resort to applying the following two equations [17]:

$$P = v_a i_a + v_b i_b + v_c i_c \quad (10)$$

$$Q = \frac{1}{\sqrt{3}}[(v_b - v_c)i_a + (v_c - v_a)i_b + (v_a - v_b)i_c] \quad (11)$$

Depending on the output data "Sp" and "Sq" of the hysteresis comparator, the selected vectors must provide the increase or decrease of the active and reaction power [18], given the switching table 3:

TABLE 3. SWITCHING FOR DPC

Sp	Sq	θ 1	θ 2	θ 3	θ 4	θ 5	θ 6
1	0	V_5	V_6	V_1	V_2	V_3	V_4
	1	V_3	V_4	V_5	V_6	V_1	V_2
0	0	V_6	V_5	V_4	V_3	V_2	V_1
	1	V_1	V_2	V_3	V_4	V_5	V_6

III. SIMULINK RESULT

This research paper processes a method to inject a power solar into electric grid. Sliding mode control was used to improve the tracking performance of the maximum power point of the solar power generator. This was done due to its instantaneous responsiveness, despite the random nature of weather conditions. As shown in Figure 6, the accuracy of the results obtained in solar energy production was confirmed according to instantaneous changes in solar irradiance (figure 4).

Figure 5 illustrates the dynamic characteristics associated with the solar power generator. We can observe that the electric current also varies according to the recorded changes in solar irradiance. However, the voltage value stabilizes at 560 v.

Solar energy was injected into the electrical grid using the direct power control unit, which was connected to the inverter. The analysis of the results in figure 7 demonstrates the effectiveness of the method by showing the alignment of the energy produced by the solar power generator with the amount of energy injected into the grid, Despite the slight loss of injected energy compared to the generated energy.

Figure 9 represents the current (Ig) and voltage (Vg), we observe that THD over all modes is maintained at less than 5%, achieving healthy operation. However, the THD value is better in the case of power injection to the grid (G=1000 W/m2)

Figure 10 illustrates the tracking performance of the DC bus voltage. The continuous voltage remains steady throughout the simulation period at a value of 800V, despite some minor fluctuations at moments 0.27s, 0.5s, and 0.78s. This indicates the accuracy of the proposed controller and the effectiveness of the overall system.

After analyzing the obtained results, it can be said that the sliding controller played a fundamental role in tracking the maximum power point generated by the solar system, despite changing weather conditions, to be injected into the electrical grid using direct energy control technology. The integration of these two methods together significantly improved the system's performance and efficiency.

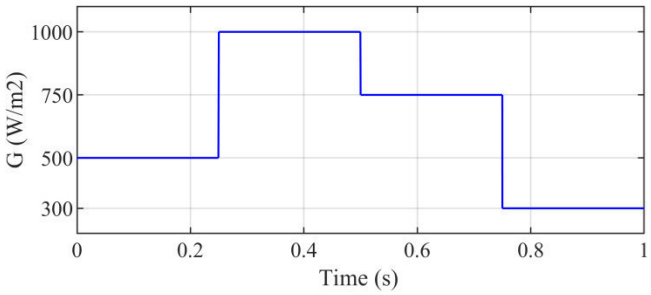
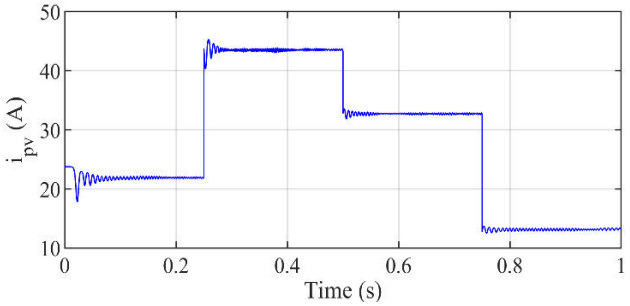
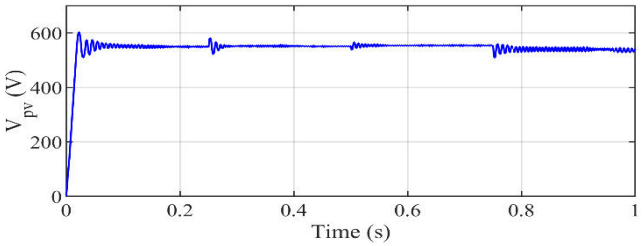


Fig.4. Solar irradiance



(a)



(b)

Fig.5. (a) Current of Pv generator. (b) Voltage of Pv generator

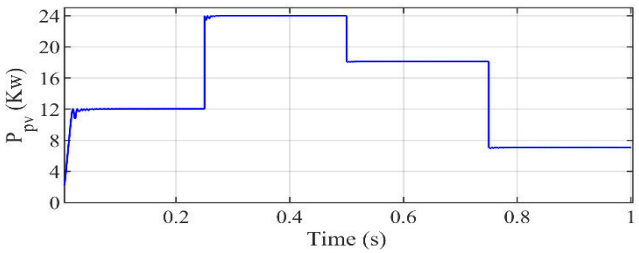


Fig.6. Solar power

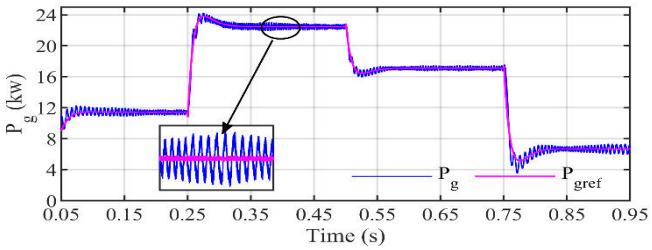


Figure 7. Active power of grid

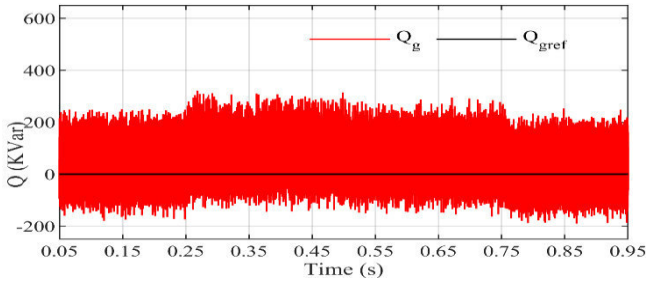


Fig.8. reactive of grid power

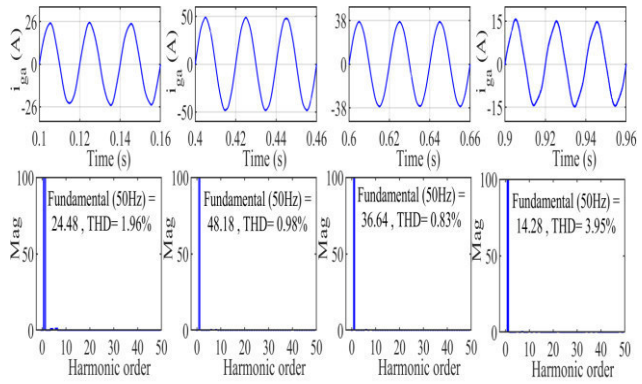


Fig.9.

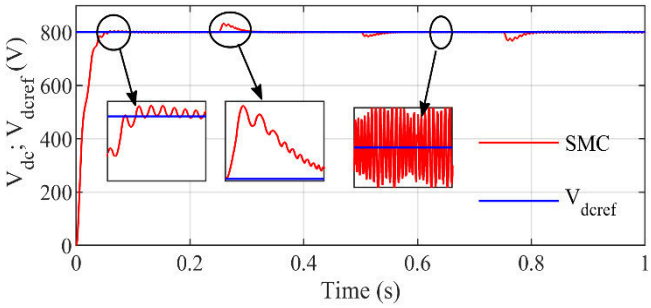


Fig.10. DC bus voltage

IV. CONCLUSION

In the proposed system for injecting solar energy into the electrical grid, two effective methods have been suggested for its success. The sliding control unit efficiently tracked the maximum electrical power point, regardless of changes in regulatory conditions, while the direct energy control unit estimated the value of the energy generated from the electrical grid. In turn, it provided operation commands to the transformer. The results obtained confirmed the effectiveness and accuracy of both techniques, as well as their rapid response.

REFERENCES

[1] M. S Alam and S. A. Arefifar "Energy management in power distribution systems: Review, classification, limitations and challenges," IEEE Access,2019, pp. 92979-93001

[2] S. Ali, Z. Zheng, M. Aillerie, J. P. Sawicki,M. C. Péra, D. Hissel, "A review of DC microgrid energy management systems dedicated to residential applications," Energies, **2021**, 14- 4308

[3] A. Moualdia, M. Mahmoudi, L. Nezli, O. Bouchhida "Modelling and control of a wind power conversion system based on the double-fed asynchronous generator," international journal of renewable energy, 2012, pp. 301-306.

[4] G. M. Shafiullah, A. M. T. Oo, D. Jarvis, S. Ali and P. Wolfs, "Potential Challenges: Integrating Renewable Energy with the Smart Grid," Proceedings of Australasian Universities Power Engineering Conference, Christchurch, 5-8 December 2010, pp. 1-6.

[5] Q. Yang, J. A. Barria and T. C. Green, "Communication Infrastructures for Distributed Control of Power Distribution Networks," IEEE Transactions on Industrial Informatics, Vol. 7, No. 2, 2011, pp. 316-327.

[6] S. Leonori, E. D. Santis, A. Rizzi, F. M. F. Mascioli, "Multi objective optimization of a fuzzy logic controller for energy management in microgrids," in: 2016 IEEE Congress on Evolutionary Computation (CEC), 2016, pp. 319–326.

[7] T.K.A.Brekken, A.Yokochi, A.von Jouanne, Z.Z.Yen,H.M. Hapkeand D.A.Halamay, "optimal energy storage sizing and control for wind power applications," IEEE Trans, Sustainable Energy, 2011, vol.2, no.1, pp.66-77.

[8] A. Mahesh and K. S. Sandhu, "Hybrid wind/photovoltaic energy systemdevelopments: Critical review and findings," Renewable and Sustainable Energy Reviews, vol. 52, pp. 1135–1147, 2015.

[9] Shankareppagol. L. M., Dodamani. S. N., Hampannavar. S, "Performance Analysis of P&O and INC MPPT for WECS," IEEE 3rd International Conference for Convergence in Technology. (I2CT)., 2018

[10] A.Bartoszewicz and J. Żuk "Sliding mode control – basic concepts and current trends," In: IEEE International Symposium on Industrial Electronics (ISIE), Bari, Italy, 4-7 July 2010, p. 3772–7.J.

[11] Clerk Maxwell "A Treatise on Electricity and Magnetism", 3rd ed. Oxford: Clarendon, 1892, pp.68–73.

[12] Z. Meng, W. Shao, J. Tang, H. Zhou "Sliding-mode control based on index control law for MPPT in photovoltaic systems," IEEE Trans Electr Mach Syst, 2018, pp. 303-311

[13] P. Garcia, C. A. Garcia., and M. Fernànde "ANFIS- Based Control of a Grid-Connected Hybrid System Integrating Renewable Energies", Hydrogen and Batteries. IEEE Trans. Industrial Informatics., 2014, pp. 1107-1117

[14] O. Zebraoui. And M. Bouzi. "Improved MPPT controls for a standalone PV/wind/battery hybrid energy system," International Journal of Power Electronics and Drive System. (IJPEDS)., 2020, pp.988-1001

[15] P. Xiong and D. Sun "Backstepping-based DPC strategy of a wind turbinedriven DFIG under normal and harmonic grid voltage," IEEE Trans. Power, 2016, pp. 4216-4252.

[16] X. Wang, D. Sun, Z. Q. Zhu "Resonant-based backstepping direct power control strategy for DFIG under both balanced and unbalanced grid conditions," IEEE Trans. Ind. Appl, 2017, pp. 4821–4830.

[17] J. Y. Lee, R. Verayiah, K H. Ramasamy, A. Marsadek, "Voltage oriented control and direct power control strategies in solving under and overvoltage conditions for heavy load application on Malaysian distribution representative network," Electr. Eng, 2021, pp. 1597–1612.

[18] Z. Ghadouche, Ch. Belfedal, T. Allaoui , M. Denai, M. Bey, "Hybrid Renewable Energy System Controlled with Intelligent Direct Power Control," Journal Européen des Systèmes Automatisés, 2022.

Automation and Supervision of Switching Multiple Power Generators within TRC-SONATRACH Headquarters

Maizi Abdelatif and Moussa Hamadache

Electrical Systems Engineering Department, Faculty of Technology, University of MHhamed Bougara, Boumerdes, Algeria

m.hamadache2@univ-boumerdes.dz; abdelatif.maizi@outlook.com

***Abstract* - The research focuses on automating the switching of SONATRACH's five backup power generators, which currently face issues with manual operation such as start-up delays and simultaneous activation. The project aims to improve the reliability and efficiency of the power supply to various buildings in SONATRACH's headquarters located in the Sidi Arcine industrial zone, Algiers. It explores power generators, the automation language and programming, including the Human-Machine Interface (HMI), to develop an automatic switching system. The ultimate goal is to enhance the company's operation by automating the switching between multiple backup power generators.**

***Index Terms* -Backup power generators, Automatic switching, Automation language, Programming, HMI.**

I. Introduction

This study aims to develop an automatic switching system to improve the operation of backup power generators at SONATRACH's Sidi Arcin site.

A. *Presentation of the Company*

SONATRACH was created on December 31, 1963. It is at the forefront due to the importance of its activities: Liquefaction and Separation, Refining, Pipeline Transport, Exploration, and Marketing. This project is held within the Pipeline Transport Activity (commonly known as the TRC). TRC covers several divisions, including: operation division, study and monitoring division, telecommunication division, and maintenance division, where this project is performed; specifically within the technical direction department [1].

B. *Backup Power Generators*

At the TRC headquarters, five backup electrical power generators are used, comprising one main generator and four secondary generators.

1) The Main Generator 1250 KVa: it is called the main generator because in case of power outage, it supplies the most critical half of the TRC headquarters, including the buildings A, C, J, Kitchen, CMS, TA OBM, and the six (06) Units Power Supply (UPS).

2) Secondary generators: like 425 KVa for B building, 250 KVa for K building, 230 KVa for C building and OBM, and 500 KVa for A, J buildings, kitchen, CMS

II. Existing Systems within TRC Headquarters

The TRC headquarters has various systems, including fire-fighting, service water, and electrical system. The main issue lies within the electrical system, specifically with the manual operation of the backup power generators. This paper focuses on addressing the problems associated with this manual operation [2].

Electrical system: The main electrical supply is a 30 kV line from a SONALGAZ delivery point, then to a 30/0.4 kV 2x1600 KVa Transformation Substation. In case of power cuts, the previous mentioned five backup power electrical generators are activated. Figure 1 shows these five-backup power electrical generators at the TRC Headquarters.

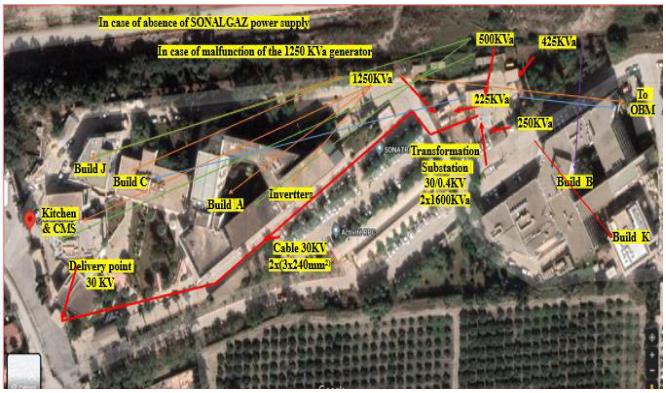


Fig. 1 Operating of the five backup power generators within TRC [2].

III. Operating Philosophy of the Current Manual Switching Between the Five Back-up Electrical Power Generators

SONALGAZ supplies all the buildings in the TRC headquarters. This process begins from departure cell number one transformer, then the current travels to the 30/0.4kV distribution transformer and then it goes to the Low Voltage Main Distribution Board (known as the TGBT). After that, it goes to the inverter, and then the buildings are powered. In case of a power cut from SONALGZAZ, the Main Generator 1250 KVa and 425 KVa secondary power generator are turned on manually to supply all the buildings without following the same process (i.e., the current goes directly from the power generators to the inverter and then to the buildings, excluding building K [2].

If one of these two power generators encounters an issue, TRC operators switch to the second process. They start the 500 KVa, the 250 KVa, and the 230 KVa power generators to supply all the buildings, excluding building B as we shown in Figure 2.

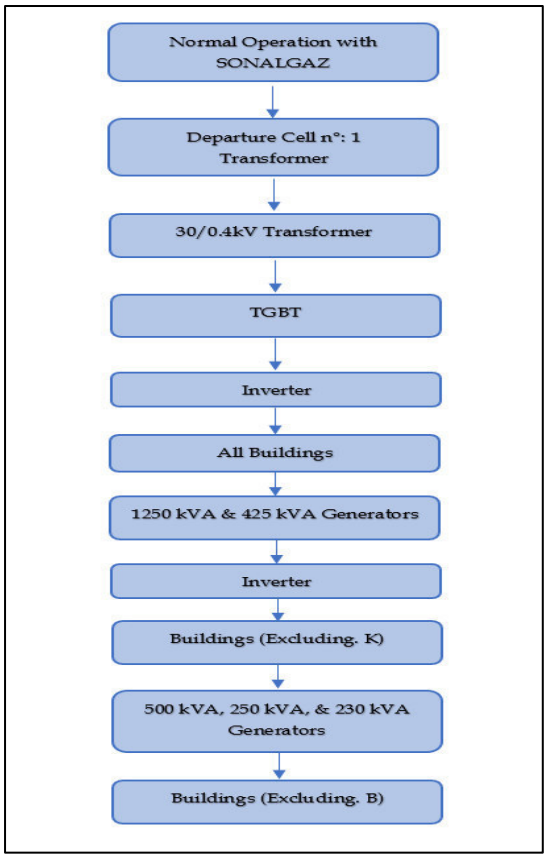


Fig. 2 The current manual switching between the five backup electrical power generators.

IV. Current Backup Electrical Supply System Problems

The current electrical supply to the TRC headquarters from SONALGAZ is ensured by two underground 30 KV arrivals originating from the DD-El Harrach, each with a power of 03 MW, where these two supplies have experienced disruptions caused by weather conditions, aging, and the instability of the two upstream and downstream delivery points of the TRC and the DP of domestic clients in the residential area. The recurring malfunctions recorded on these 30 KV arrivals lead to damages and unavailability of electrical, electronic, and computer equipment. The five (5) backup electrical power generators installed at the TRC headquarters (mentioned above in the subsection I) have problems with starting delay, simultaneous starting, switching, and too long a response time due to the currently manually operation and the logic used, which manifested as the separate installation of each group over the years. Furthermore, the fire prevention network at the headquarters lacks visibility (except for local cabinets) on the operation of the various pumps installed, as well as the status of the network whether (i.e., ON or OFF).

V. Proposed Solution

After defining the existing issues in the current electrical supply system (installation) in the TRC headquarters, especially in its manually back-up plan, the following steps are proposed in this project to deal with them, and finally to improve its efficiency, availability, and safety.

- Step 1: Understanding the existing system
- Step 2: Selecting and installing the programmable logic controller (PLC)
- Step 3: Development and integration with Human-Machine Interface (HMI):

Step 1. Understanding the Backup Power Generators

A. The Main Generator 1250 KVa

It has the following components: Perkins engine, MX321 voltage regulator, and a Stamford alternator. It covers two starting modes: a Diesel staring mode and a Pneumatic one. Its real view is shown in Figure 3.



Fig. 3 Real view of the 1250 KVa main power generator

B. The Power Generator 500 KVa

It has the following components: Volvo engine, Leroy Somer alternator, Control panel, Fuel system, Cooling system, and a Starting battery. Its real view is shown in Figure 4.



Fig .4 Real view of the power generator 500KVa

Its starting mode includes checking fuel levels, checking battery levels, activating the control panel, starting the engine, and then monitoring its operation.

C. The Power Generator 425 KVa

It has the following components: Perkins engine, Alternator, Control panel, Fuel system, Cooling system, and an Exhaust system. Its real view is shown in Figure 5.

Its starting mode includes preparing the control panel, starting the Black Start Diesel Generator (BSDG) to initiate power, starting the 425 KVa generator, and then

synchronizing the generator with the grid for efficient power supply.



Fig .5 Real view of the power generator 425KVa

D. The Power Generator 250 KVa

It has the following components: Alternator, Diesel engine (JD T4 Final or Volvo TAD series), Cooling system with ParCOOL coolant, Electrical distribution in a sound-attenuated enclosure, DSE 7320 control panel for engine monitoring and protection, and a Rugged steel enclosure for noise reduction. Its real view is shown in Figure 6.



Fig .6 Power generator 250KVa [3].

Its starting mode includes ensuring proper connections, checking oil and coolant levels, activating the battery switch, pressing the start button on the Qc2212 Advance digital controller, and then monitoring performance during synchronization and connection to the electrical load.

E. The Power Generator 230 KVa

It has the following components: MAN D2866 LXE40 diesel engine and the Stamford UCI274H alternator. Its real view is shown in Figure 7.



Fig .7 Real view of the power generator 230KVa

Its starting mode includes checking fuel levels, checking oil and coolant levels, activating the control panel, starting the engine, and then monitoring performance during operation.

F. Parameters of the Power Generators

It has few sensors (Fuel level, Oil pressure, and Emergency buttons)

The balance sheet of I/O is as follows: 36 digital inputs (DI), 24 analog inputs (AI), and 6 digital outputs (DO). Examples of some of these parameters (focusing on the Main Generator 1250 KVa) are shown in Table 1.

TABLE 1
PARAMETRES OF THE POWER GENERATORS

	parametres	Type
POWER GENERATOR 1250 KVa	Fuel Level	AI
	DC Battery Status	DI
	Circuit Breaker Status	DI
	Emergency Button	DI
	Heater	AI
	Commutator Man/Auto	DI
	Oil Pressure	AI
	Oil Temperature	AI
	Water Temperature	AI
	Water Lavel (Max / Min)	AI
	Compressor Operating Commande	DO
	Trip Alarm	DI
	Output Voltage Fault	DI
	Electric Motor (Start / Stop)	DO

Step 2. Selecting and Installing the PLC

A. Programmable Logic Controller (PLC)

1) Definition: PLC stands for programmable logic controller. A PLC is a programmable computing device that is used to manage electromechanical processes, usually in the industrial sector. A PLC is sometimes referred to as an industrial PC, a term that describes a PLC's main function as a specialized industrial computing machine.

2) Different types of PLC: There are various kinds of PLCs that are available, such as the Mini PLCs, the Modular PLCs, the Fixed PLCs, the Micro PLCs, the Nano PLCs, and the Safety PLC.

3) *Criteria for choosing a PLC:* For this project, the "S7-1500 CPU 1515-2 PN" PLC was selected because the Siemens S7-1500 PLC series offer:

- High performance
- Integrated diagnostics
- User-friendly engineering
- Scalability
- Built-in safety
- Robust communication
- Advanced technology
- Energy management
- Security features
- Modular design

B. Presentation of the TIA Portal V16

TIA Portal V16 (Totally Integrated Automation Portal) is Siemens' comprehensive engineering framework designed to streamline and enhance automation projects [4]. It integrates various tools for planning, engineering, and commissioning processes, offering new features to improve collaboration, version management, and software quality [5].

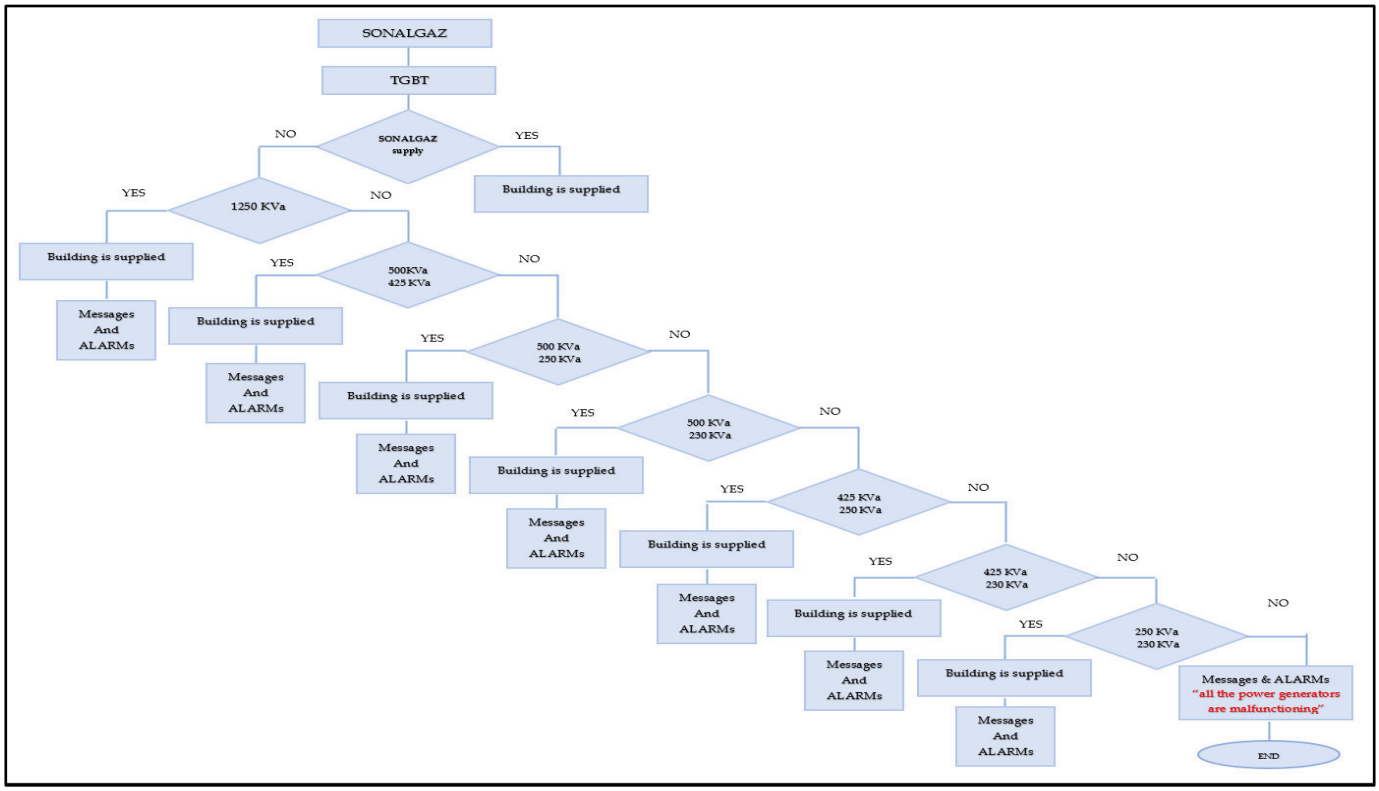


Fig. 8 Flow chart of the proposed automatic switching multiple power generators of the TRC-SONATRACH Headquarters

C. The Proposed Solution of Automating the Switching of the Multiple Backup Power Generators

1) *Description of the proposed solution*:

- When SONALGAZ supply to the buildings of the TRC headquarters is interrupted for any reason, the cooling system is shutdown as it is not critical and consumes a lot of power, and the main power generator (1250 KVa) starts supplying all the buildings. An alarm message is sent stating that SONALGAZ is not supplying the buildings and a message is also sent to state which buildings are supplied.
- If SONALGAZ does not restore the supply and the 1250 KVa generator has a problem with the diesel engine or compressor (i.e., cannot be turn ON), then switch to supply the power using the 500 KVa and 425 KVa generators and send alarm messages indicating that SONALGAZ is not supplying and the 1250 KVa generator is malfunctioning, plus a message indicating which building are supplied is generated.
- If the 425 KVa generator also has a problem, the proposed switching solution moves to the next step by supplying the power using the 500 KVa and 250 KVa generators and sending alarm messages stating that SONALGAZ is not supplying, and the 1250 KVa and 425 KVa generators are malfunctioning, plus a message indicating which building are supplied is generated.
- If the 250 KVa generator encounters a problem, the proposed switching logic flow proceed to the next step by powering on the 500 KVa and 230 KVa generators to guarantee the power supply with sending the adequate alarms and messages to the maintenance team and to the leadership. These steps are followed up until both the 250 KVa and 230 KVa generators encounter problems. The whole flow chart of the proposed solution of automatic switching multiple power generators to guarantee the power backup of the TRC-SONATRACH Headquarter is shown in Figure 8.

2. *Executing the Proposed Solution with TIA Portal V16*: As mentioned before, the S7-1500 with CPU 1515-2 PN is selected. In follow, the hardware and software configuration is described.

Hardware configuration: After creating the project in TIA Portal, the next step is to configure the workstation. Choose the device SEMATIC S7-1500 CPU 1515-2 PN, and add the following modules: Digital inputs: DI 16x24VDC BA, DI 32x24VDC BA; Analog inputs: AI 8xU/I/RTD/TC ST; Digital outputs: DO 8x24VDC/2A HF; and a Power supply: PS 25W 24VDC

Software configuration: It start with creating a Tags Tables, after writing the execution program using different blocks.

Program Development: To properly organize the project, the program is subdivided into several blocks; each assigned a specific task to execute.

A few lines of the switching between SONALGAZ and the 1250 KVa power generator within the FC 13 block, is shown in Figure 9.

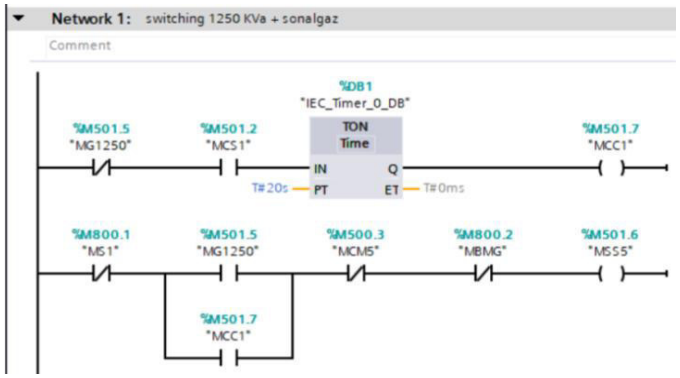


Fig. 9 Function of the switching

Step 3. Development and integration of the HMI

A. PC Station for Industrial Supervision

This station is used in industrial automation applications for: (a) representing and monitoring the system, (b) serving as an alarm manager, (c) acting as an archiving tool for maintenance, (d)recording fault history, and (e) tracking production.

B. Modules Uses in this PC Station

Are the WinCC advanced and the IE_general (Industrial Ethernet general).

C. Connecting PLC with PC Station

For retrieve data from the PLC to the PC station for supervision. Figure 10 shown the PLC and PC station connected with highlighting the IP address.

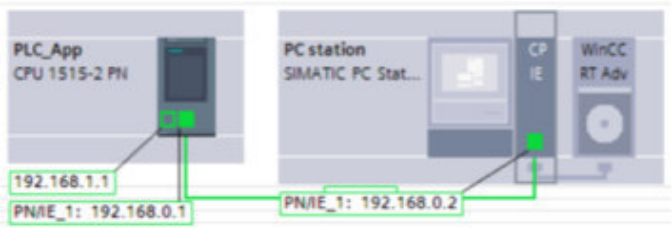


Fig 10 PLC / PC station connection

D. HMI TAGs Table:

Some of the HMI objects and variables from the PLC are given in Figure 11.

E. Configuration of HMI Interfaces (VIEWS)

HMI interfaces (i.e., views) are elements used for the control and command of machines and installations. To create views, there are predefined objects that allow representing the installation, displaying the procedures, and defining the process values.

1. Homepage view: The Homepage is the main or introductory page of this project. It serves as a starting point for the user/operator and provides an overview of the project's content, structure, and facilitates navigation within it. It can be seen in Figure 12.

The explanations of each button of the Homepage View shown above are: '**Main view**' to show the switching of power generator and whether the building is supplied or not. '**1250 KVa**' to display the parameters of the 1250 KVa power generator. '**500 KVa**' to display the parameters of the 500 KVa power generator. '**425 KVa**' to display the parameters of the 425 KVa power generator. '**250 KVa**' to display the parameters of the 250 KVa power generator. '**230 KVa**' to

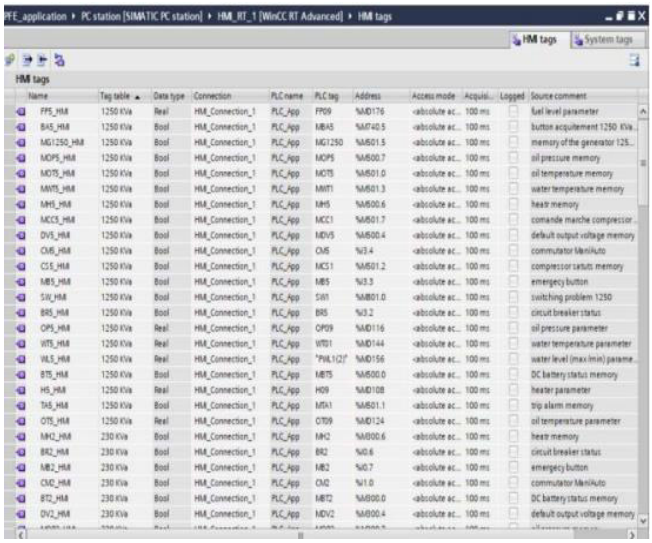


Fig .11 HMI TAGs table



Fig. 12 Homepage View

display the parameters of the 230 KVa power generator. '**ALARMS**' to display the alarms that occur when starting the simulation. And '**EXIT**' for stop the simulation.

2. Main View: This view shows how the buildings are supplied and by what power sources, either generators or SONALGAZ as can be seen in Figure 13. The buttons are added to facilitate access to the other pages.

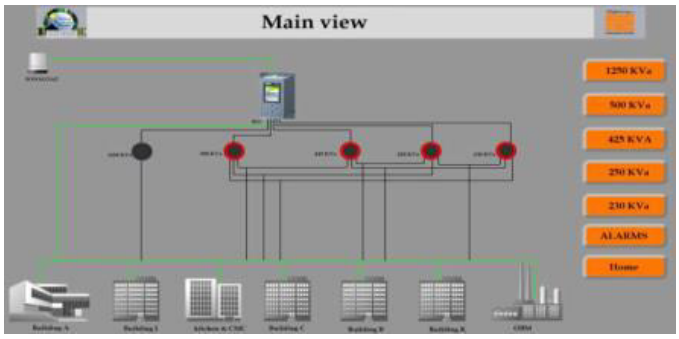


Fig. 13 Main View

The explanations of each button of the Homepage View

shown above are: '**Home**' to back to the homepage view. The remaining buttons (in orange color) have the same functions as in the previous homepage view.

3. Main Generator 1250 KVa View: This view is developed to supervise the Main Generator 1250 KVa power generator and its all parameters. It can be seen in Figure 14.

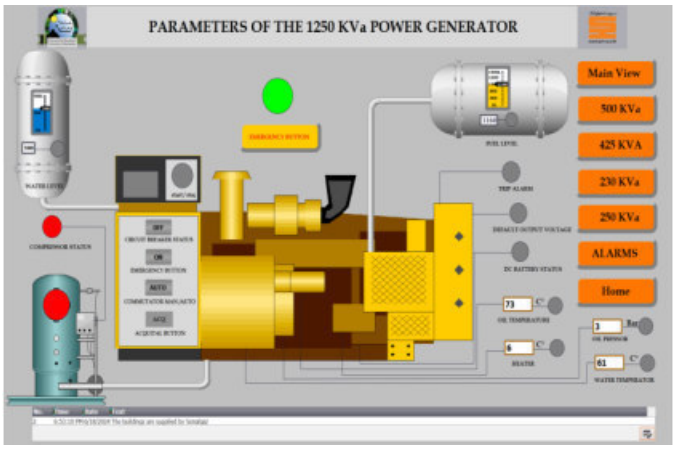


Fig. 14 Main Generator 1250 KVa View

The explanations of each button, circle, and rectangle of the Main Generator 1250 KVa View shown above are: the buttons (in orange color) have the same functions as in the previous views. In addition to the '**Global Emergency Button**' is shown by the green cirle (top middle), and is used for shutting down all power generators in case of danger. This button has two colors (Green color: Indicates the global emergency button is OFF; and Red color: Indicates the global emergency button is ON). '**ACQ (acquital) Button**' used to restart the 1250 KVa after setting all the parameters. '**Red circles**' indicates the condition is not verified. '**Gray circles**' indicates the condition is verified. '**Green circles**' indicates the power generator 1250 KVa is supply the buildings. '**Gray rectangle**' indicates the diesel engine is ready to start. And the '**Red rectangle**' indicates the diesel engine is not ready to start.

4. ALARMS View: This view displays what is happening in this project system, i.e., the power generators and switching, allowing the user/operator and the maintenance team to identify exactly where the problem is, and send messages to inform the system' status to the maintenance engineers and to the leadership. The ALARMS View can be seen in Figure 15.



Fig. 15 ALARMs View

The explanations of each button of the ALARMs View shown above are: the buttons (in orange color) have the same functions as in the previous views. In addition to the **Acknowledge button'** used to clear alarms on the screen after setting all the parameters.

VI. Conclusion

This paper presents an automated switching solution for SONATRACH's multiple backup power generators at the TRC Headquarters to address current challenges of start-up delays, synchronization, and manual operation inefficiencies. The proposed system leverages a PLC-based control scheme integrated with HMI for real-time monitoring and management, aiming to increase operational reliability and reduce response time. By adopting the S7-1500 PLC with TIA Portal V16 for programming and supervision, a scalable, robust framework that ensures a seamless transition between power sources during outages is achieved, ultimately enhancing SONATRACH's operational resilience. The anticipations that this proposed automation approach not only provides a model for similar industrial facilities, but also opens avenues for future advancements in automated power management systems.

Acknowledgment

I extend my gratitude to SONATRACH for its support and for providing the resources necessary for this project. Special thanks to the technical direction and maintenance divisions at TRC headquarters for their cooperation and invaluable expertise. I am also thankful to the Electrical Systems Engineering Department, Faculty of Technology at the University of M'Hamed Bougara for their academic guidance, which significantly contributed to the success of this work, and I appreciate Siemens for supplying the TIA Portal software, which played a crucial role in developing this system.

References

[1] A. Maizi and M. lakhdari, "Conception et Réalisation d'une application de gestion des accidents / incidents," Diplôme de Technicien Supérieur en Informatique, Option: Base de Données, Inst. Nat. Spéc. Form. Prof. Mohamed Tayeb Boucenna, Ministère de la Formation et de l'Enseignement Professionnels, Hussein Dey, Algérie, 2023.

[2] Sonatrach, "Contrôle et supervision des utilités du siège de l'Activité TRC," Algiers, Algeria, Feb. 2024.

[3] Atlas Copco. (2024). QAS 250 Large Diesel Generators. (Online). Available:https://www.atlascopco.com/en-us/construction-equipment/products/generators-us/large-diesel-generators/qas-250 (Accessed: 22-May-2024).

[4] Automation Sense. (2024). TIA Portal V16: Siemens présente la nouvelle version de TIA Portal. (Online). Available: https://www.automation-sense.com/blog/automatisme/tia-portal-v16_siemens-presente-la-nouvelle-version-de-tia-portal.html (Accessed: 14-Jun-2024).

[5] Siemens AG. (2024). 6ES7314-6EH04-0AB0. (Online). Available: https://mall.industry.siemens.com/mall/en/ww/Catalog/Product/6ES7314-6EH04-0AB0 (Accessed: 22-May-2024).

Stick-Slip Vibration Elimination in Rotary Drilling Systems Using a Tuned GA-PID-Based Controller

TELLAA Mawloud
Department of Automation and Electrification of Process Applied Automation Laboratory,Faculty of Hydrocarbons and Chemistry University M'Hamed Bougara-Boumerdes,
Boumerdes 35000, Algeria
m.tellaa@univ-boumerdes.dz

AIBECHE Abderrezak
Department of Automation and Electrification of Process Applied Automation Laboratory,Faculty of Hydrocarbons and Chemistry University M'Hamed Bougara-Boumerdes,
Boumerdes 35000, Algeria
aibeche_umbb@univ-boumerdes.dz

DOGHMANE Mohamed Zinelabidine
Department of Automation and Electrification of Process Applied Automation Laboratory,Faculty of Hydrocarbons and Chemistry University M'Hamed Bougara-Boumerdes,
Boumerdes 35000, Algeria
Bob L. Herd Department of Petroleum Engineering, Texas Tech University, Lubbock 79401, Texas, USA

KIDOUCHE Madjid
Department of Automation and Electrification of Process Applied Automation Laboratory,Faculty of Hydrocarbons and Chemistry University M'Hamed Bougara-Boumerdes,
Boumerdes 35000, Algeria
mkidouche@univ-boumerdes.dz

***Abstract*— Vibrations in rotary drilling systems can harm equipment, borehole quality, and drilling efficiency, while also increasing non-productive time. Current mitigation relies on manual parameter adjustments, an unreliable method. Researchers are developing automatic strategies, with recent success using a genetic algorithm (GA)-based Proportional-Integral-Derivative (PID) controller. However, its effectiveness in mitigating stick-slip vibrations remains untested. This paper investigates the use of a PID controller fine-tuned by the genetic algorithm using integral squared error (ITAE) fitness function to enhance vibration control in rotary drilling systems, with a focus on stick-slip vibrations.**

Keywords: Proportional–integral–derivative controller (PID); Genetic Algorithms (GA); rotary drilling systems; stick-slip vibrations

I. INTRODUCTION

The rotary drilling technique is vital in the petroleum sector for the extraction of oil and gas. Three kinds of vibrations can be generated during drilling: axial, lateral, and torsional vibrations. Stick-slip vibrations are the most detrimental, as they can initiate other vibrations. The objective of this study is to reduce stick-slip vibrations. In recent years, researchers have studied these vibrations. For instance, Tang et al. [3] discovered that reducing WOB could eliminate stick-slip, albeit with a delay. Increasing rotary table speed also reduces torsional vibrations, but it does not account for drill string stiffness [4]. Sliding mode control, which is applied to four degrees of freedom [5] and extended to n degrees [6], is designed to prevent oscillations, albeit with high stabilization times. Liu [7] experimentally compared rotational velocities and achieved some vibration reduction. Abdulgalil and Siguerdidjane [8] employed back-stepping control to eliminate stick-slip vibrations. Hybrid controllers, which combine various methods, have also been investigated [9, 10]. Although the results are satisfactory, it is necessary to enhance the robustness of the controller. This paper introduces a genetic algorithm (GA) in conjunction with a PID controller to improve vibration suppression. The objective of this innovative methodology is to enhance the performance of controllers that have been implemented in the past. The study contrasts the new method with a PID controller that lacks optimization, noting that PID with GA has not been investigated for this issue. This article presents a PID controller that has been optimized using a genetic algorithm (GA). The manuscript is structured as follows: Section 2 provides a description of the drilling system and the vibrations it produces. Section 3 introduces the mathematical model and the PID controller. Section 4 provides a detailed explanation of the genetic algorithm and how it is applied. Section 5 discusses the results obtained and compares the performance of the PID controller with and without GA optimization. The manuscript concludes by outlining potential areas for future research. In Section 2, the drilling system and vibrations are briefly explained. Section 3 introduces the mathematical model of vibrations, the PID controller, and its importance. Section 4 outlines the GA algorithm and its application to the drilling system. Section 5 discusses open and closed-loop results and compares them with traditional PID control.

II. DRILLING SYSTEM

The power system, hoisting system, circulating system, rotating system, well control system, and monitoring system are the six fundamental systems that make up the drilling rig

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

[13, 14]. The hoisting system includes elements including draw-works, an auxiliary brake, a traveling system, a hook, and a derrick and supports drilling and completing operations. The drilling strings inside the wellbore are spun by the rotary system to push the bit to fracture rock [15]. The circulating system's functions include moving cuttings from the bottom of the hole to the surface, preventing debris from settling, regulating the temperature of the drill bit, and preserving the structural integrity of the wellbore walls. The well control system is used to shut down and/or control the well in an emergency to restore pressure balance. The monitoring and control system ensures the effective operation of all rig components to satisfy the needs of the drilling process, while the power and transmission system supplies energy to the entire system., Fig. 1. offers a visual representation of the rotary drilling system.

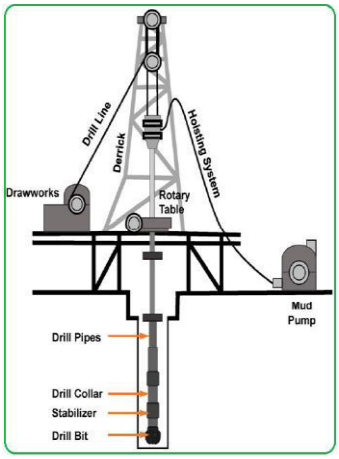


Fig. 1. Schematic of oil well drill string system [9]

III. MATHIMATICAL MODEL

In this study, we adopt a three-element model based on the mass-spring-damper framework.

A. Drilling string model

In this research, a mass-spring-damper model is applied to replicate the behavior of a rotary drilling system [3]. This model encompasses three degrees of freedom, with the upper disc symbolizing the rotary table, the middle disc signifying the tool string, and the lower disc representing the drill bit. You can find a concise summary of the model's parameters in Table 1. Additionally Fig. 2 portrays the three degrees of freedom model.

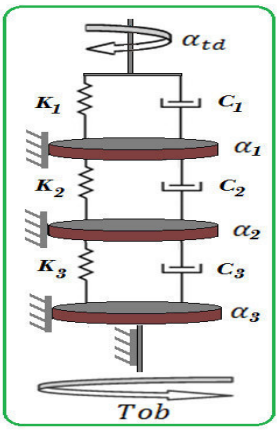


Fig. 2. Equivalent mass-spring-damper three-element model [3, 14].

The mathematical representation of the latter is articulated through equation (1) [14]:

$$\begin{cases} \ddot{\alpha}_1 = \frac{1}{J_1}\left(c_1(\dot{\alpha}_{td}-\dot{\alpha}_1)-k_1(\alpha_{td}-\alpha_1)-c_2(\dot{\alpha}_1-\dot{\alpha}_2)-k_2(\alpha_1-\alpha_2)-\mu\dot{\alpha}_1\right) \\ \ddot{\alpha}_2 = \frac{1}{J_2}\left(c_2(\dot{\alpha}_1-\dot{\alpha}_2)+k_2(\alpha_1-\alpha_2)-c_3(\dot{\alpha}_1-\dot{\alpha}_2)-k_3(\alpha_2-\alpha_3)-\mu\dot{\alpha}_2\right) \\ \ddot{\alpha}_3 = \frac{1}{J_3}\left(c_3(\dot{\alpha}_2-\dot{\alpha}_3)+k_3(\alpha_2-\alpha_3)-\mu\dot{\alpha}_3-Tob\right) \end{cases} \quad (1)$$

we define $u = \dot{\alpha}_{td}$, and the states variables as :

$x_1 = \dot{\alpha}_1$ $x_2 = \dot{\alpha}_2$, $x_3 = \dot{\alpha}_3$, $x_4 = \alpha_{td} - \alpha_1$,
$x_5 = \alpha_1 - \alpha_2$ and $x_6 = \alpha_2 - \alpha_3$,
the Equation (1) is written then as:

$$\begin{cases} \dot{x}_1 = -\frac{c_2+c_1+\mu}{J_1}x_1 + \frac{c_2}{J_1}x_2 + \frac{k_1}{J_1}x_4 - \frac{k_2}{J_1}x_5 + \frac{c_1}{J_1}u \\ \dot{x}_2 = \frac{c_2}{J_2}x_1 - \frac{c_2+c_3+\mu}{J_2}x_2 + \frac{c_3}{J_2}x_3 + \frac{k_2}{J_2}x_5 - \frac{k_3}{J_2}x_6 \\ \dot{x}_3 = \frac{c_3}{J_3}x_2 - \frac{c_2+\mu}{J_3}x_3 + \frac{k_3}{J_3}x_6 - \frac{1}{J_3}Tob \\ \dot{x}_4 = \dot{\alpha}_{td} - \dot{\alpha}_1 = u - x_1 \\ \dot{x}_5 = \dot{\alpha}_1 - \dot{\alpha}_2 = x_1 - x_2 \\ \dot{x}_6 = \dot{\alpha}_2 - \dot{\alpha}_3 = x_2 - x_3 \end{cases} \quad (2)$$

TABLE I. PARAMETERS DESCRIPTION FOR THE DESIGNED ROTARY DRILLING MODEL [13, 19].

Parameter	Description	Unit
α_{td}	The top drive angular displacement	[rad]
$\alpha_{i=1,2,3}$	The angular displacements	[rad]
$k_{i=1,2,3}$	Torsional stiffness coefficient	[N.m/rad]
$c_{i=1,2,3}$	Internal damping coefficient	[N.m.s/rad]
μ	Wall friction coefficient	[N.m]
$J_{i=1,2,3}$	The inertia	[kg.m2]
Tob	The torque on bit	[N.m]

B. Generalized model of Rock-bit interaction :

Within the literature, various functions have been employed to characterize the interaction between the drill bit and the rock. However, in the context of our specific investigation, we will utilize the generalized model as outlined by the expression [14] :

$$Tob = \mu_n Nr\left(\frac{x_3}{\sqrt{x_3^2+\Omega_0^2}} + \frac{p\Omega_0 x_3}{x_3^2+\Omega_0^2}\right) + Dx_3\left(\frac{x_3}{\Omega_1} - 1\right) \quad (3)$$

Where the model coefficients are given in the following table [14]:

TABLE II. PARAMETERS USED IN ROCK-BIT GENERAL MODEL

parameter	description	value
μ_n	friction coefficient	28 nm
n	the force vector	9.81 * wob
r	the contact radius vector	0.1 m
Ω_0	chain transition speed	1
Ω_1	transition speed for the well	31.4159
p	the initial friction parameter	1.5
d	the linear damping vector	0.28

C. PID controller

The Proportional Integral Derivative (PID) control stands as the most commonly employed control strategy for regulation. Its widespread adoption across industries can be attributed to its simplicity in tuning and its reliable operation. The general expression for the PID control equation is as follows [10]:

$$C(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau + K_d \frac{de(t)}{dt} \qquad (4)$$

Where, K_P, K_i , K_d represent the parameters of the PID controller denote respectively : the proportional coefficient, the integral time constant and the differential coefficient.
To ensure stability, the derivative component is filtered to prevent singularities. Equation (4) represents the standard PID configuration, which is employed prior to applying the GA algorithm for parameter optimization.

IV. GENITIC ALGOIRITHM

A. History and principle:

The Genetic Algorithm (GA) is a metaheuristic algorithm that utilizes a random search technique with parametric coding. It was first introduced by John Holland in 1975 and was initially employed to address control problems in gasoline pipelines by his student Goldberg. The GA is particularly advantageous for solving control problems in situations where an analytical approach is not feasible due to a vast search space. The algorithm operates based on the principle of natural selection, where the fittest candidates survive and reproduce, while the fewer fit ones are eliminated. The Genetic Algorithm (GA) is a metaheuristic algorithm that utilizes a random search technique with parametric coding.

B. Fitness function

To achieve the desired performance, the parameters of a controller are chosen to minimize the dynamic error

$$e(t) = y_{ref}(t) - y_{sys}(t) \qquad (5)$$

To do this, a cost function needs to be reduced. There are many basic performance criteria used in PID controller tuning [14]. In our case we use the ITAE (Integral of Time-weighted Absolute Error):

$$J_{ITAE} = \int_0^T t\,|e(t)|dt \qquad (6)$$

C. PID-GA Based controller structure :

The architecture of the suggested approach for implementing the PID-GA in rotary drilling systems is depicted in Fig 2. The process of designing an appropriate PID controller hinge on identifying the optimal tuning parameters for the PID controller.

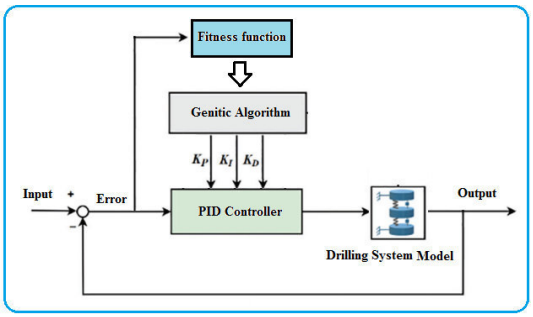


Fig. 3. GA-PID based control system structure [16]

V. SIMULATION RESULTS AND DISCUSSION

A. Open loop response

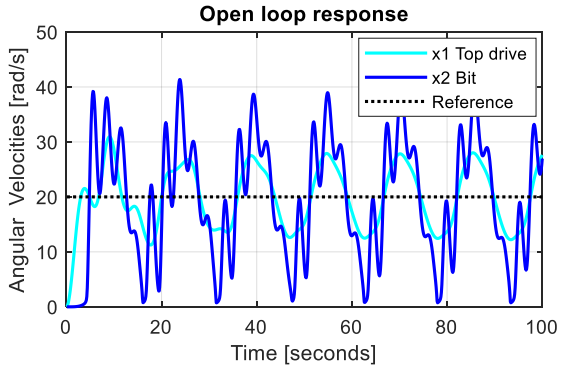


Fig. 4. The open-loop responses of rotary drilling systems without control

The Fig.4 illustrates the step response of a drilling rotary system without controller. We can see the bit velocity has big vibration, the regulation time is very long, the dynamic and static character of the system is not good, and the ability of tracking rotary table speed is bad.

B. Closed loop using PID Controller without optimisation

Below are the appropriately selected PID controller parameter values:

K_p=0.015 ;
K_i=0.0023 ;
K_d=0.011;

The simulation yields the following results

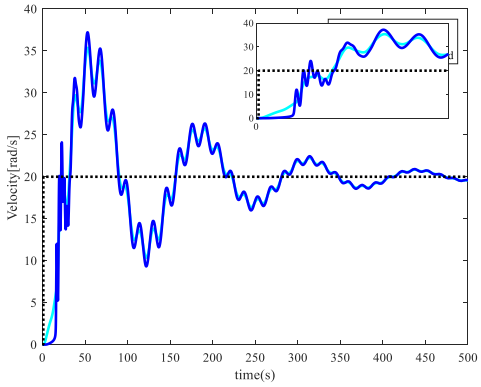


Fig. 5. The closed-loop responses of rotary drilling systems with PID controller without GA

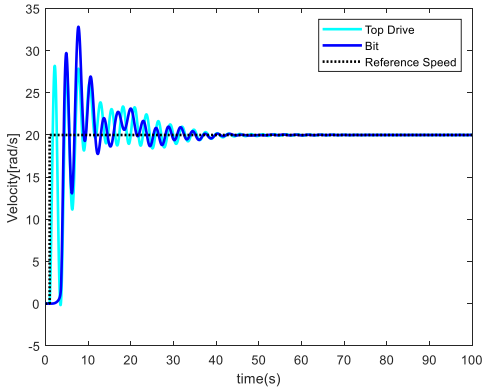


Fig. 6. The closed-loop responses of rotary drilling systems with PID controller optimized by GA

In Fig. 4, the step response of a drilling rotary system is shown, featuring a PID controller without any tuning. Notably, the figure reveals that the bit velocity exhibits minor oscillations, but the time response is considerably protracted. The system shows a delayed ability to track the rotary table speed, characterized by substantial oscillations and eventual stabilization after an extended period

C. Closed loop using PID Controller with GA Algorithm

The GA algorithm was initialized with the subsequent parameter settings :

TABLE III. GENITIC ALGORITHM PARAMETERS

Parameter	Value
Generations number	20
Population size	80
Lower bounds=[kp,Ki,Kd]	lb=[0 0 0]
Upper bounds=[kp,Ki,Kd]	ub=[10 10 10]

After the optimization process by GA algorithm was finished we get these parameters below

K_p=2.5124 ;
K_i=0.3029;
K_d=8.9386;

The simulation results of the drilling string model controlled with PID tuned by GA is given below

Reference of Angular velocity = 20 rad/s for 0<t< 100 s

In Fig. 5, the step response of a drilling rotary system is depicted, featuring a PID controller fine-tuned using the GA algorithm. Notably, the figure showcases the rapid elimination of stick-slip vibrations, achieving a considerably shorter time response compared to the original system. Furthermore, there is no overshooting observed beyond the 30-second mark

D. Comparative Study

The simulation results visualized in Fig. 4, revealing pronounced stick-slip vibrations within the drill string system. Various components of the system exhibit periodic vibrations, with the drill bit being the most severely affected, indicated by both viscous and sliding states. These outcomes validate the model's effectiveness in safeguarding drilling tools from damage.

Our study delves into the PID controller's ability to mitigate stick-slip vibrations. As demonstrated in Fig. 5 and Fig.6, the controller successfully reduces stick-slip vibrations, but it's the PID controller with GA Algorithm that completely eliminates the phenomenon in a short time, ensuring a stable drilling process. These findings underscore the practical applicability of our proposed control methods within the drilling industry

VI. CONCLUSION

The purpose of this work was to enhance the reduction of stick-slip vibration in rotary drilling systems. By utilizing a PID controller tuned by a GA algorithm with various fitness functions, the incidence of stick-slip vibration could be rapidly suppressed. The obtained results are pretty acceptable; the power of the GA might be enhanced by boosting the population size but that require more time to reach optimal parameters. Future work linked to expand integer order PID controller to non-integer order PID for vibration control in drillstring system the future study could include applying alternative tuning methods as well as more complicated artificial intelligence method.

.

REFERENCES

[1] R. I. Leine , D. H. van Campen , W. J. G. Keultjes (2002) ,Stick-slip Whirl Interaction in Drillstring Dynamics.

[2] Doghmane, M. Z. Kidouche,M. Ahriche ,A. (2021) Decentralized Overlapping Control Design with Application to Rotary Drilling System, IETE Journal of Research, DOI: 10.1080/03772063.2021.1886602

[3] Tang, L., Zhu, X., Shi, C., Tang, J. and Xu, D. (2015), "Study of the influences of rotary table speed on stick-slip vibration of the drilling system", Petroleum, Vol. 1 No. 4, pp. 382-387.

[4] Tang, L., Zhu, X., Qian, X. and Shi, C. (2017), "Effects of weight on bit on torsional stick-slip vibration of oilwell drill string", Journal of Mechanical Science and Technology, Vol. 31 No. 10,pp. 4589-4597.

[5] Navarro-López, E. and Licéaga Castro, E. "Non-desired transitions and sliding-mode control of a multi-DOF mechanical system with stick-slip oscillations," Chaos Solitons Fract, Vol. 41, no. 4, pp. 1561–2200, Aug. 2009.

[6] Vaziri, V. Kapitaniak, M. and Wiercigroch, M. "Suppression of drill-string stick–slip vibration by sliding mode control: numerical and experimental studies," Eur. J. App. Math, Vol. 29, no. 5, pp. 805–25, Oct. 2018.

[7] Liu,Y., "Suppressing stick-slip oscillations in underactuated multibody drill-strings with parametric uncertainties using sliding-mode control," IET Control Theory, Vol. 9,no. 1, pp. 91–102, Jan. 2015.

[8] F. Abdulgalil, and H. Siguerdidjane, "Backstepping design for controlling rotary drilling system," in Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005, Toronto, Ont., Canada, Aug. 28–31, 2005, pp. 120–4.

[9] Mendil,C. Kidouche,M. Doghmane,M Z. Hybrid sliding PID controller for torsional vibrations mitigation in rotary drilling systems, Indonesian Journal of Electrical Engineering and Computer Science, Vol. 22, No. 1, April 2021, pp. 146~158, ISSN: 2502-4752, DOI: 10.11591/ijeecs.v22.i1.pp146-158

[10] Mendil, C., Kidouche, M., Doghmane, M.Z. (2022). Sliding Mode Controller Design for Torsional Vibrations Minimization Under Rock-Bit Interaction Effects. In: Hatti, M. (eds) Artificial Intelligence and Heuristics for Smart Energy Efficiency in Smart Cities. IC-AIRES 2021. Lecture Notes in Networks and Systems, vol 361. Springer, Cham. https://doi.org/10.1007/978-3-030-92038-8_61

[11] China University of Petroleum Press and Springer Nature Singapore Pte Ltd. 2021 Z. Guan et al., Theory and Technology of Drilling Engineering, https://doi.org/10.1007/978-981-15-9327-7_2

[12] Thiago G Ritto1 and Maryam Ghandchi-Tehrani2 Active control of stick-slip torsional vibrations in drill-strings Journal of Vibration and Control 2019, Vol. 25(1) 194–202 ! The Author(s) 2018 Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/1077546318774240.

[13] Mendil, C., Kidouche, M., Doghmane, M.Z., Benammar, S. and Tee, K.F. (2021), "Rock–bit interaction effects on high-frequency stick-slip vibration severity in rotary drilling systems", Multidiscipline Modeling in Materials and Structures, Vol. 17 No. 5, pp. 1007-1023. https://doi.org/10.1108/MMMS-10-2020-0256

[14] Abushawish,A, Hamadeh,M. Bou Nassif. ,A PID Controller Gains Tuning Using Metaheuristic Optimization Methods: A survey INTERNATIONAL JOURNAL OF COMPUTERS (2020) DOI: 10.46300/9108.2020.14.14

[15] Israa Ahmed.A, Muntadher Khamees,M. A review of adaptive tuning of PID-controller: Optimization techniques and applications Int. J. Nonlinear Anal. Appl. 15 (2024) 2, 29–37 ISSN: 2008-6822 (electronic) http://dx.doi.org/10.22075/ijnaa.2023.21415.4024

[16] Kanhua Su, Jian Wei*, Meng Li, Hao Li, Wenghao Da and Lang Zhang. Research on Stick-Slip Vibration Suppression Method of Drill String Based on Machine Learning Optimization SV, 2023, vol.57 DOI: 10.32604/sv.2023.043734

[17] Problems and Solutions Drilling Engineering Problems and Solutions A Field Guide for Engineers and Students M. E. Hossain and M. R. Islam

[18] R. Riane, MZ. Doghmane, M. kidouche, KF Tee "Stick Slip vibration suppression in drill string using observer based LQG controller" Sensors 22 (16) 5979, 2022

[19] M. Kidouche, A. Charef " A constructive methodology of Lyapunov function of composite system with nonlinear interconnection term " International journal of robotic and automation, 21 (1), 19-24, 2006

Comparative Analysis of Various Adaptation Mechanisms in a Sliding Mode Observer-Based MRAS for Vector Control of induction Machine

Zoubir Gadouche
Laboratory of Electrical and Computer Engineering (L2GEGI) of Electrical Department University Ibn Khaldun of Tiaret Tiaret, Algeria
gadouche.zoubir@gmail.com

Djilali Toumi
Laboratory of Electrical and Computer Engineering (L2GEGI) of Electrical Department University Ibn Khaldun of Tiaret Tiaret, Algeria
toumi_dj@yahoo.fr

Mokhtar Mahmoud Mohammedi
Laboratory of Electrical Drives Development (LDEE) of Electrical Department University of Sciences & Technology of Oran Mohamed Boudiaf Oran, Algeria
mokhtar.mohammedi@yahoo.fr

***Abstract* — Carrying out control of the induction machine without a speed sensor has many advantages. It eliminates the sensor cable, provides a reduction in noise, provides an increase in reliability and makes hardware less complex and less expensive. The objective of this work is the study by simulation in the Matlab/Simulink environment, of the indirect vector control without speed sensor, using a speed estimator which is called "MRAS" (model reference adaptive system). We propose to analyze the behavior of a sliding mode observer (adjustable model), associated with the "MRAS" speed estimator, with the application of a robustness text (variation of the machine parameters) and the use of different adaptation mechanisms carried out by a classic PI regulator and by a fuzzy PI regulator. We conclude our work with a comparative study between the two adaptation mechanisms.**

***Keywords*— induction machine, sliding mode observer, model reference adaptive system (MRAS), adaptation mechanism, fuzzy regulator.**

I. INTRODUCTION

The induction machine associated with static converters is currently the most used in industrial applications with variable speed, where high torque performances are required [1].

The performance of the induction machine control system depends on the information at any time from sensors (speed, position, flow, etc.). However, the reconstruction of the machine speed using an estimator or observer from measurable quantities allows an increase in the robustness of the control structure. An improvement in this quality is obtained by the estimation of the flux, generally not measurable [2].

Among them, speed estimators are based on the theory of adaptive system with reference model and an adaptation mechanism, usually a PI controller, makes the behaviour of the adaptive model tend towards the behaviour of the reference model; in our work using two types of classical and fuzzy controllers, to make the comparison between them.

These estimators are called MRAS (model reference adaptive system) and are the most popular techniques which have been implemented for the control of induction motor without speed sensor using only stator voltage and current measurements [2].

II. IDUCTION MACHINE MODEL

The mathematical model of the induction machine, in the (d q) reference frame of Park, is described by a system of five differential equations, four of which are electrical and one mechanical [3]:

$$\begin{cases} \dfrac{d\phi_{sd}}{dt} = \dfrac{-1}{\sigma T_s}\phi_{sd} + \omega_s \phi_{sq} + \dfrac{M_{sr}}{\sigma T_s L_r}\phi_{rd} + V_{sd} \\ \dfrac{d\phi_{sq}}{dt} = \dfrac{-1}{\sigma T_s}\phi_{sq} - \omega_s \phi_{sd} + \dfrac{M_{sr}}{\sigma T_s L_r}\phi_{rq} + V_{sq} \\ \dfrac{d\phi_{rd}}{dt} = \dfrac{-1}{\sigma T_r}\phi_{rd} + \omega_r \phi_{rq} + \dfrac{M_{sr}}{\sigma T_r L_r}\phi_{sd} + V_{rd} \\ \dfrac{d\phi_{rq}}{dt} = \dfrac{-1}{\sigma T_r}\phi_{rq} + \omega_r \phi_{rd} + \dfrac{M_{sr}}{\sigma T_r L_r}\phi_{sq} + V_{rq} \end{cases} \quad (1)$$

And using the torque equation

$$C_{em} = p\left(\Phi_{sd} i_{sq} - \Phi_{sq} i_{sd}\right) \quad (2)$$

III. VECTOR CONTROL

For our study, we opted for indirect vector control with oriented rotor flux.The orientation of the reference frame linked to the rotor field is determined by the integration of the pulsation ω_s which is calculated by:

$$\omega_s^* = p\Omega + \omega_r^* \quad et \quad \omega_r^* = \frac{M_{sr} i_{sq}^*}{T_r \phi_r^*} \quad (3)$$

Fig.1 shows that we have three regulators, a PI-P type speed regulator and two PI type current regulators (i_{sd} and i_{sq}).

The decoupling terms are:

$$\begin{cases} \hat{e}_d = -\omega_s \sigma L_s i_{sq} - \dfrac{M_{sr}}{L_r^2} R_r \phi_r \\ \hat{e}_q = \omega_s \sigma L_s i_{sd} + \dfrac{M_{sr}}{L_r} \omega \phi_r \end{cases} \quad (4)$$

IV. MRAS SPEED ESTIMATOR ASSOCIATED WITH SLIDING MODE OBSERVER

A. *MRAS estimation method*

Fig. 2 shows the rotor speed estimation technique by an adaptive system without a reference model [4].

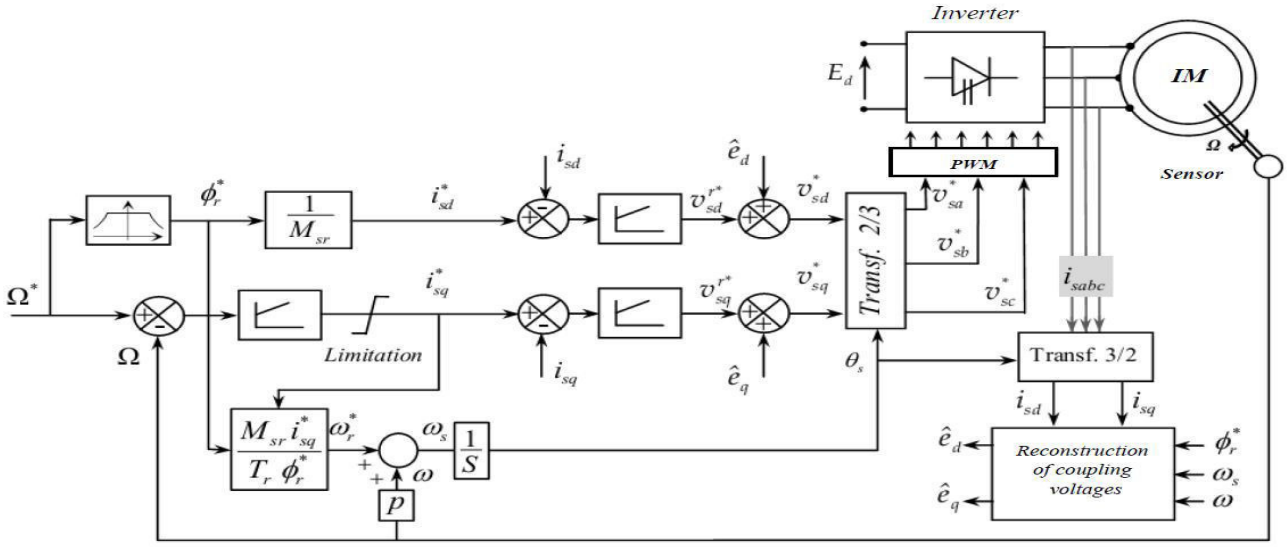


Fig.1. Block diagram of indirect vector control with oriented rotor flux

This structure of the MRAS estimator is introduced by kubota and Matsuse in [5], for the estimation of rotor speed and resistance.

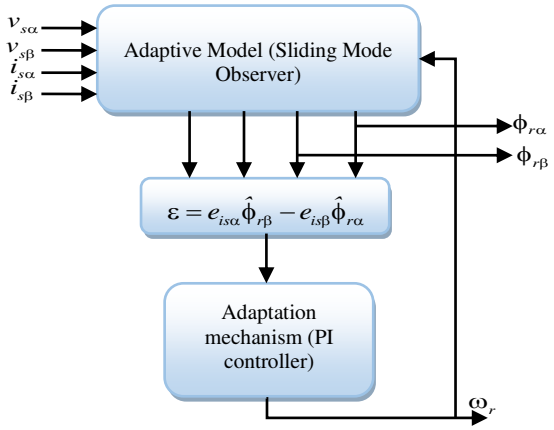


Fig. 2. Scheme of MRAS estimator associated with the sliding mode observer

1) Adaptative Model (without sliding mode observer):

For the adaptive model, we use the model of the induction machine in a fixed reference frame. We consider as state variables the stator currents ($i_{s\alpha}$,$i_{s\beta}$), the rotor fluxes($\phi_{r\alpha}$,$\phi_{r\beta}$).

$$\begin{cases} \dfrac{d\hat{i}_{s\alpha}}{dt} = -\lambda i_{s\alpha} + K\dfrac{1}{T_r}\hat{\phi}_{r\alpha} + pK\omega_r\hat{\phi}_{r\beta} + \dfrac{1}{\sigma L_s}V_{s\alpha} \\ \dfrac{d\hat{i}_{s\beta}}{dt} = -\lambda i_{s\alpha} + K\dfrac{1}{T_r}\hat{\phi}_{r\beta} - pK\omega_r\hat{\phi}_{r\alpha} + \dfrac{1}{\sigma L_s}V_{s\beta} \\ \dfrac{d\phi_{r\alpha}}{dt} = \dfrac{M_{sr}}{T_r}i_{s\alpha} - \dfrac{1}{T_r}\hat{\phi}_{r\alpha} - p\omega_r\hat{\phi}_{r\beta} \\ \dfrac{d\phi_{r\alpha}}{dt} = \dfrac{M_{sr}}{T_r}i_{s\beta} - \dfrac{1}{T_r}\hat{\phi}_{r\beta} + p\omega_r\hat{\phi}_{r\alpha} \end{cases} \quad (5)$$

2) Adaptation Mecanism

The input of the adaptation mechanism is the error calculated by equation (6):

$$\varepsilon = e_{is\alpha}\hat{\phi}_{r\beta} - e_{is\beta}\hat{\phi}_{r\alpha} \quad (6)$$

With: $\begin{cases} e_{is\alpha} = i_{s\alpha} - \hat{i}_{s\alpha} \\ e_{is\beta} = i_{s\beta} - \hat{i}_{s\beta} \end{cases}$

Adaptation mechanism is represented by a "PI" regulator [6]. So, the speed $\hat{\omega}_r$ is expressed by:

$$\hat{\omega} = k_p(\varepsilon_\alpha\hat{\phi}_{r\alpha} - \varepsilon_\beta\hat{\phi}_{r\beta}) + k_i\int_0^t(\varepsilon_\alpha\hat{\phi}_{r\alpha} - \varepsilon_\beta\hat{\phi}_{r\beta})d\tau \quad (7)$$

The gains k_p and k_i are positive gains and arbitrarily chosen.

B. Sliding mode observer associated with the MRAS estimator (adaptive model)

For our work we used the speed estimation technique (MRAS), without a reference model with the use of a sliding mode observer instead of the classic adaptive model.

The sliding mode observer has been proposed for the observation of rotor flux and stator current. It is a copy of the asynchronous machine model [7].

$$\begin{cases} \dfrac{d\hat{i}_{s\alpha}}{dt} = -\lambda i_{s\alpha} + K\dfrac{1}{T_r}\hat{\phi}_{r\alpha} + pK\omega_r\hat{\phi}_{r\beta} + \dfrac{1}{\sigma L_s}V_{s\alpha} + \Lambda_1 I_s \\ \dfrac{d\hat{i}_{s\beta}}{dt} = -\lambda i_{s\alpha} + K\dfrac{1}{T_r}\hat{\phi}_{r\beta} - pK\omega_r\hat{\phi}_{r\alpha} + \dfrac{1}{\sigma L_s}V_{s\beta} + \Lambda_2 I_s \\ \dfrac{d\phi_{r\alpha}}{dt} = \dfrac{M_{sr}}{T_r}i_{s\alpha} - \dfrac{1}{T_r}\hat{\phi}_{r\alpha} - p\omega_r\hat{\phi}_{r\beta} + \Lambda_3 I_s \\ \dfrac{d\phi_{r\alpha}}{dt} = \dfrac{M_{sr}}{T_r}i_{s\beta} - \dfrac{1}{T_r}\hat{\phi}_{r\beta} + p\omega_r\hat{\phi}_{r\alpha} + \Lambda_4 I_s \end{cases} \quad (8)$$

Where Λ_1, Λ_2, Λ_3 and Λ_4 are observer gains, with Λ_j=[Λ_{j1} $\Lambda j2$] for.jϵ{1,2,3,4}.

I_2 is a column vector.

$$I_s = \begin{bmatrix} \mathrm{sign}(s_1) \\ \mathrm{sign}(s_2) \end{bmatrix} \quad (9)$$

$$\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \frac{1}{\beta}\begin{bmatrix} \frac{M_{sr}}{T_r}(i_{s\alpha} - \hat{i}_{s\alpha}) - p\omega K(i_{s\alpha} - \hat{i}_{s\alpha}) \\ \frac{M_{sr}}{T_r}(i_{s\beta} - \hat{i}_{s\beta}) + p\omega K(i_{s\alpha} - \hat{i}_{s\alpha}) \end{bmatrix} \quad (10)$$

And:

$$\beta = \left[\frac{K}{T_r}\right]^2 + p^2K^2\omega^2 \quad (11)$$

$$\Gamma = \frac{1}{\beta}\begin{bmatrix} \frac{K}{T_r} & -p\omega_r K \\ p\omega_r K & \frac{K}{T_r} \end{bmatrix} \quad (12)$$

The stability study of this observer was done by Benchaib in [7]; and after their study we have the following result:

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} = \begin{pmatrix} \frac{K}{T_r}\delta_1 & p\omega_r K\delta_2 \\ -p\omega_r K\delta_1 & \frac{K}{T_r}\delta_2 \end{pmatrix} \quad (13)$$

$$\begin{bmatrix} \Lambda_{31} & \Lambda_{32} \\ \Lambda_{41} & \Lambda_{42} \end{bmatrix} = \begin{pmatrix} \left(q_1 - \frac{1}{T_r}\right)\delta_1 & -p\omega_r\delta_2 \\ p\omega_r\delta_1 & \left(q_2 - \frac{1}{T_r}\right)\delta_2 \end{pmatrix} \quad (14)$$

V. FUZZY CONTROLLER ASSOCIATED WITH ADAPTATION MECHANISM

The principle of the fuzzy PI controller [8] is to replace the control calculated using a classic PI with the Defuzzification value of the control. The basic algorithm of a classic PI controller is expressed by:

$$U_{PI} = K_p e + K_I \int e\,dt = K_p \int (de + e/T_I)\,dt \quad (15)$$

e : The setpoint-measurement error.

de **:** Derivative of error.

$T_I = K_P/K_I$: Integral time constant.

The digital structure of the fuzzy PI controller is as follows:

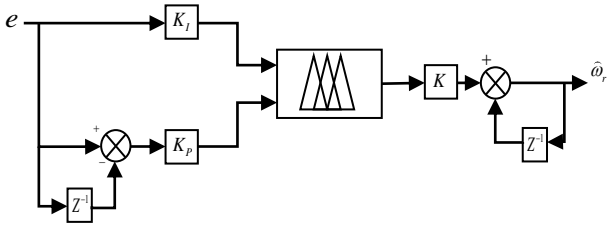


Fig.3 Fuzzy PI controller block diagram

A. *Fuzzification*

The fuzzy controller inputs are:

- e : The error is defined by

$$e(k) = e_{isd}(k).\hat{\phi}_{rq}(k) - e_{isq}(k).\hat{\phi}_{rd}(k) \quad (16)$$

- *de:* The derivative of the error, it is approximated by:

$$de(k) = \frac{e(k) - e(k-1)}{T_e} \quad (17)$$

With the membership functions for error and error derivative and output they are the same:

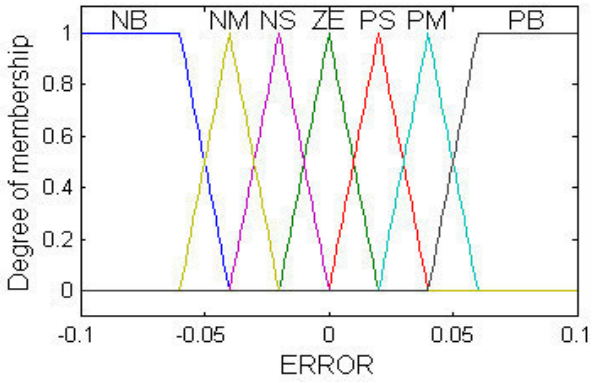


Fig.4 Membership function

B. Inference

Our inference matrix will take the following form:

TABLE I. INFERENCE MATRIX [9].

de \ *e*	*NG*	*NM*	*NP*	*ZE*	*PP*	*PM*	*PG*
PG	*ZE*	*PP*	*PM*	*PG*	*PG*	*PG*	*PG*
PM	*NP*	*ZE*	*PP*	*PG*	*PG*	*PG*	*PG*
PP	*NM*	*NP*	*ZE*	*PP*	*PM*	*PG*	*PG*
ZE	*NG*	*NM*	*NP*	*ZE*	*PP*	*PM*	*PG*
NP	*NG*	*NG*	*NM*	*NP*	*ZE*	*PP*	*PM*
NM	*NG*	*NG*	*NG*	*NM*	*NP*	*ZE*	*PP*
NG	*NG*	*NG*	*NG*	*NG*	*NM*	*NP*	*ZE*

C. Défuzzification

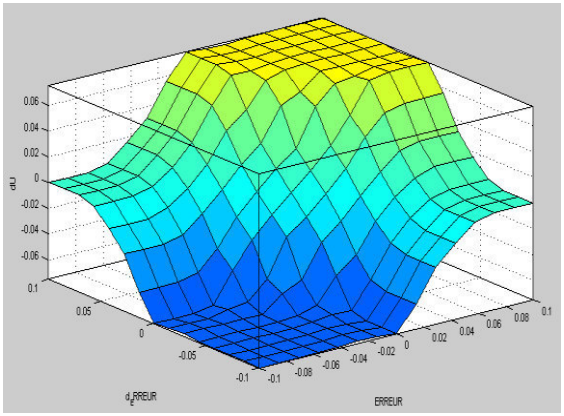


Fig.5 Fuzzy regulator surface diagram

VI. SIMULATION RESULTS

This test is carried out by varying the direction of rotation for a simulation time equal to 5 seconds with a duration of '2.25 s' for each direction of rotation and application of a load torque during time intervals from 1 s to 2.2s and 3.3s to 4.5s “fig.6”.

We will study the influence of the variation of the induction machine parameters on the performances of the MRAS estimator with PI-classic, PI-Fuzzy and the control. We consider the variation of the following parameters: stator resistance ‘R_s’, rotor resistance ‘R_r’, rotor inductance ‘L_r’ and moment of inertia ‘J’.

We present the maximum estimation errors in the transient regime in “TABLE II” for MRAS with PI-classical and in “TABLE III” a comparison between the two adaptation mechanisms. For the parametric variation of rotor and stator resistances and rotor inductance we can say that there is no difference between the results of the classic and fuzzy regulators.

For the variation of the moment of inertia ‘J’ and the operation at low speed, we notice that the speed responses are more efficient for the MRAS estimator with PI-Fuzzy than for the one with PI-Classic which is clearly represented by “fig.7” and “fig.8” for the comparison.

For all figures, the first is the classic adaptation mechanism and the second concerns the fuzzy adaptation mechanism.

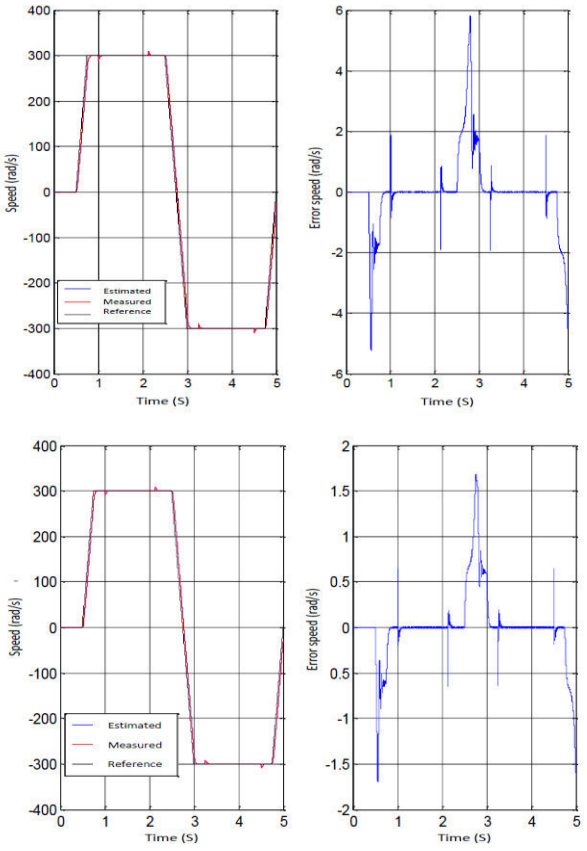


Fig.6 Comparison of rotation speed behavior for normal operation.

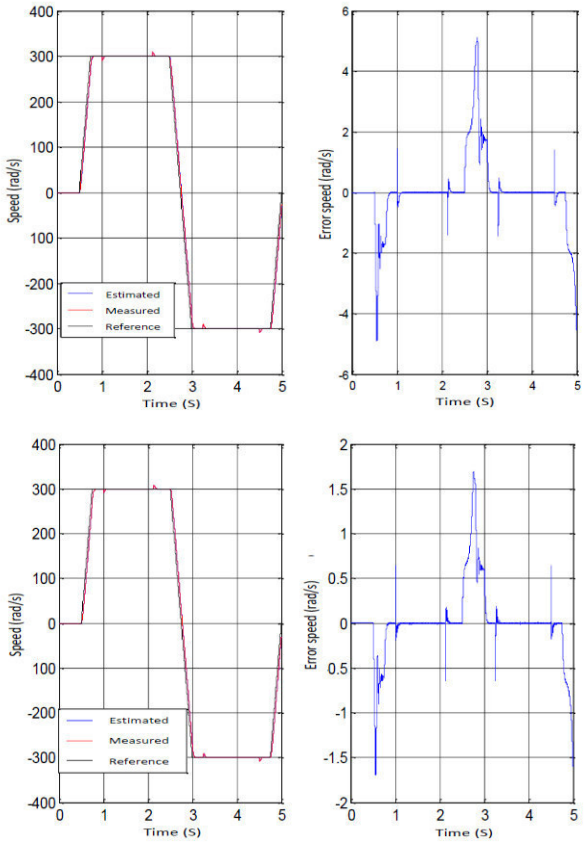


Fig.7 Comparison of the behavior of the rotation speed for a variation of +50% in moment of inertia.

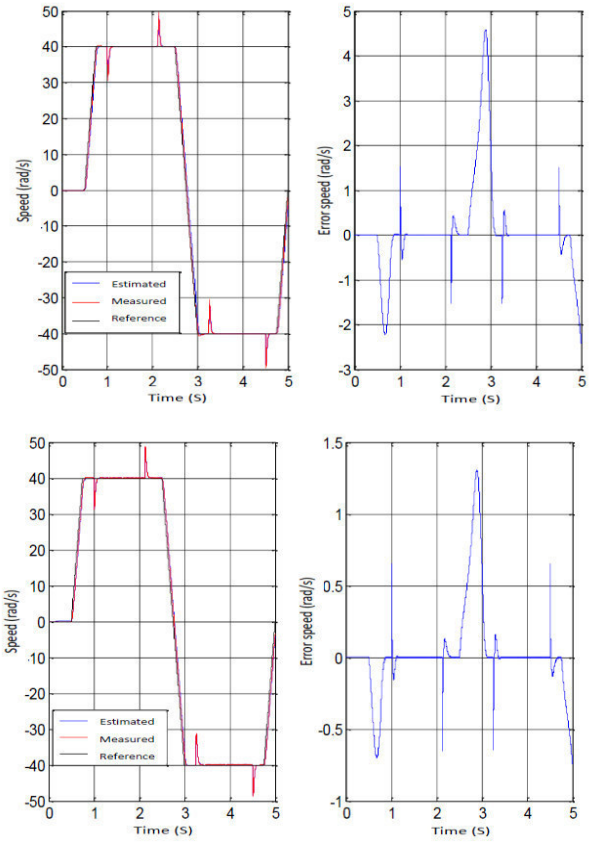


Fig.8 Comparison of rotation speed behavior for low speed operation.

TABLE II. MAXIMUM ESTIMATION ERRORS IN DYNAMIC REGIME WITH CLASSICAL ADAPTATION MECHANISM.

Variables / Errors	**Normal**	**+50% J**	**Low speed**	**+50%Rs**	**+50% Rr**	**+20 % Lr**
Speed (rad/s)	2 %	1.67%	11.25%	6.67%	1.33%	1.67 %
Current Isd (A)	0.4%	0.35%	0.16%	0.94%	0.58%	0.58 %
Current Isq (A)	0.15%	0.15%	0.1%	0.62%	0.2%	0.42 %
Flow module (wb)	8%	10%	10%	80%	28%	50%

TABLE III. THE ESTIMATION ERRORS IN DYNAMIC REGIME FOR THE TWO CONTROLLERS

	Classic controller			**Fuzzy contoller**		
	Nominal	+50% J	Low speed	Nominal	+50% J	Low speed
Speed (rad/s)	2 %	1.67%	11.25%	0.53%	0.53%	3%
Current Isd (A)	0.4%	0.35%	0.16%	0.1%	0.1%	0.4%
Current Isq (A)	0.15%	0.15%	0.1%	0.05%	0.05%	0.1%
Flow module (wb)	8%	10%	10%	3.4%	3.4%	2.4%

VII. CONCLUSION

The work presented in this paper is the study subject of the control without speed sensor.

The estimated speed , flow and observed current quickly converge to the corresponding real variables. The estimator (MRAS) validity with the association of the sliding mode observer was shown by the simulation results. The robustness tests with respect to parametric variations developed in this work confirm the robustness of the two regulators used, in particular that using fuzzy regulator.

Finally, we can say that this work has allowed us to know and test one of the most recent and promising techniques in the field of process control, which is the speed sensorless control with the use of MRAS estimator associated with fuzzy logic.

REFERENCES

[1] B.K.Bose, « Technology Trends in microcomputer of electrical machines », *IEEE Trans Ind Electron.*, vol. 35, nº 1, pp. 160-177, fév. 1988.

[2] A. Derdiyok, M. K. Guven, H. ur. Reman, N. Inanc, and L. Xu, « Desingn and implementation of a new sliding-mode observer for speed-sensorless control of induction machine », *IEEE Trans. Ind. Electronics.*, vol. 49, nº. 5, pp. 1177-1182, oct. 2002.

[3] F.-Z. Peng et T. Fukao, « Robust speed identification for speed-sensorless vector control of induction motors », *Industry Applications, IEEE Transactions on*, vol. 30, nº. 5, pp. 1234 -1240, oct. 1994.

[4] I. Landau, « Elimination of the real positivity condition in the design of parallel MRAS », *IEEE Transactions on Automatic Control*, vol. 23, nº. 6, pp. 1015- 1020, déc. 1978.

[5] H.Kubota et K.Matsuse, « speed sensorless field-oriented control of induction motor with rotor resistance adaptation », *IEEE Trans. Ind. Appl.*, vol. 30, nº 5, pp. 1219-1224, oct. 1994.

[6] C. Schauder, « Adaptive speed identification for vector control of induction motors without rotational transducers », *IEEE Trans. Ind. Appl.*, vol. 28, nº 5, pp. 1054-1061, oct. 1992.

[7] A. Benchaib, A. Rachid, E. Audrezet, et M. Tadjine, « Real-time sliding-mode observer and control of an induction motor », *IEEE Trans. Ind. Electron.*, vol. 46, nº 1, pp. 128−138, févr. 1999.

[8] P.Pivonka et J.Zizka, « The use of fuzzy pi/pd/pid controlers in fuzzy control », *Third Eur. Congr. Intell. Tech. Soft Comput.*, vol. 2, pp. 1041−1045, août 1995.

[9] Hang-xiong li et H.B. Gatland, « conventienal fuzzy control and its enhancement », *IEEE Trans. Smc*, vol. 26, nº 5, oct. 1996.

Analysis of the Impact of Objective Function Choice on Optimal FOPID Controller Design for Three Tanks Liquid Level Control.

Karim Benaouicha
Process Automation and Electrification Department, AAL Applied Automation Laboratory, Faculty of Hydrocarbons and Chemistry, University of Boumerdes (1).
Electrical Engineering Department, Faculty of Sciences and Applied sciences, University of Bouira (2).
Bejaia, Algeria
k.benaouicha@univ-boumerdes.dz

Yassine Bensafia
LISEA Laboratory, Electrical Engineering Department, Faculty of Sciences and Applied sciences, University of Bouira
y.bensafia@univ-bouira.dz

Hamza Akroum
AAL Applied Automation Laboratory, Faculty of Hydrocarbons and Chemistry, University of Boumerdes,
akroum@univ-boumerdes.dz

Abdelhakim Idir
Electrical Engineering Department, University of M'sila Algeria.
AAL Applied Automation Laboratory, Faculty of Hydrocarbons and Chemistry, University of Boumerdes, 35000 Boumerdes, Algeria.
M'sila, Algeria
abdelhakim.idir@univ-msila.dz

***Abstract*— Three-tank liquid level control systems are inherently complex, characterized by varied response times among tanks and sensitivity to disturbances, such as changes in inlet flow rates or desired level variations. To maintain the liquid level within desired ranges despite these perturbations, a stable and robust control strategy is paramount. The design of Fractional Order Proportional-Integral-Derivative (FOPID) controllers for such systems necessitates meticulous selection of the objective function to ensure optimal performance. This study investigates the impact of four commonly used objective functions - Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Square Error (ITSE), and Integral Time Absolute Error (ITAE) - on the optimal design of FOPID controllers for a three-tank system. Through detailed simulations and comparative analyses, we evaluate the performance of FOPID controllers designed based on each objective function, considering response time, overshoot, steady-state error, and robustness against disturbances. The results reveal significant differences in system performance based on the chosen objective function, highlighting the importance of appropriate objective function selection to meet specific performance requirements. This research provides invaluable guidelines for control system designers, underscoring the impact of the objective function on FOPID controller optimization in liquid level control applications.**

Keywords— FOPID Controller Design, Three-Tank Liquid Level Control, Objective Functions (ISE, IAE, ITAE, ITSE), Stability, Control System Optimization

I. INTRODUCTION

fractional calculus has seen a significant rise in popularity, particularly in its applications to science and engineering, its origins date back over 300 years [5]. Control theory is one of the primary domains where fractional-order systems find extensive use, with an increasing amount of theoretical and practical research [6]. Fractional-order (FO) differential calculus, commonly referred to as fractional calculus, extends the standard operations of differentiation and integration to non-integer orders. This mathematical field began to take shape around 1695, initiated through exchanges between mathematicians Leibniz and Hospital [7]. Incorporating fractional-order filters in closed-loop control systems offers notable benefits in the system's dynamic performance. These advantages are largely due to the hereditary characteristics of fractional-order operators [8], which provide enhanced resilience against external disturbances [9, 10]. In recent decades, fractional-order PID controllers have seen rapid development and widespread implementation in control engineering and related research fields [11, 12]. It is widely recognized that optimizing controllers (whether PID or FOPID) is one of the most efficient techniques for handling parametric uncertainties in both linear and nonlinear systems, regardless of whether they are fast or slow [13-15]. Particular attention has been given to applying fractional-order systems in slow and fast applications to achieve optimal control strategies.

The control of liquid levels in interconnected tank systems is a fundamental challenge in various industrial processes, such as chemical, pharmaceutical, and water treatment plants. Effective regulation of liquid levels ensures system stability and optimal performance, preventing overflow, underfill, and operational inefficiencies. The inherent nonlinearities, time delays, and dynamic behavior of such systems make the design of robust and efficient control

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

strategies critical to their successful operation. Among the most widely used control strategies, Proportional-Integral-Derivative (PID) controllers are favored for their simplicity and general effectiveness. However, when faced with more complex and nonlinear systems, fractional-order PID (FOPID) controllers have emerged as a promising alternative, offering enhanced flexibility and performance.

The Fractional Order Proportional-Integral-Derivative (FOPID) controller, an extension of the classical PID controller, introduces two additional tuning parameters: the fractional orders of the integral (λ) and derivative (μ) actions. This added degree of freedom allows for more precise control, particularly for systems exhibiting memory and hereditary properties. As a result, FOPID controllers are increasingly being explored for applications in systems with complex dynamics, such as multi-tank liquid level control systems. These systems are often characterized by slow dynamics, interaction between tanks, and the need for accurate level control to ensure process efficiency and safety [29].

The design of FOPID controllers typically involves selecting appropriate values for the proportional (k_p), integral (k_i), derivative (k_d), and the fractional orders (λ and μ) that minimize a performance criterion or objective function. The choice of objective function plays a crucial role in determining the overall performance of the controller. Common objective functions include time-domain performance indices such as Integral of the Absolute Error (IAE), Integral of the Squared Error (ISE), Integral of Time-weighted Absolute Error (ITAE), and Integral of Time-weighted Squared Error (ITSE), each of which emphasizes different aspects of the system's response, such as rise time (T_r), settling time (T_s), overshoot (OS), and steady state error (E_{ss}). The selection of the appropriate objective function can significantly affect the controller's ability to meet the desired performance specifications.

This paper aims to analyze the impact of different objective functions on the optimal design of a FOPID controller for a three-tank liquid level control system. The three-tank system, a benchmark for control studies, presents a multi-input multi-output (MIMO) structure with complex dynamics that pose a significant challenge for traditional control approaches. By exploring the effect of different objective functions on the tuning of the FOPID controller parameters, we seek to identify the most suitable performance criterion for achieving optimal control in this context.

The remainder of the paper is structured as follows: Section 2 provides an overview of the three-tank liquid level system and its mathematical modeling. Section 3 outlines the design methodology for the FOPID controller and discusses the various objective functions considered in the optimization process. Section 4 presents the simulation results and a comparative analysis of the controller performance based on the chosen objective functions. Finally, Section 5 concludes the paper with key insights and recommendations for future research.

In summary, this study highlights the importance of the objective function in the design of FOPID controllers, offering valuable guidance for practitioners seeking to optimize control performance in complex multi-tank systems.

I. Approximations of Fractional-Order Operators and Systems

Dynamical systems form the foundation of mathematical models used to describe many physical phenomena, which are often of fractional order.

If one wishes to design an integer-order dynamic system from a fractional one, an effective and commonly used method is to design a continuous linear integer-order filter. The filter approximation of fractional-order operators, proposed by the French professor Oustaloup and his colleagues, marked a new era in the simulation of complex fractional-order systems [5]. These filters allow the user to actively select the frequency band and the order of interest, in order to approximate the fractional-order operator using an integer-order transfer function model. The design of these filters will be presented below.

A. Filters Approximations

As demonstrated by the previous examples, the application of these filters is greatly compromised by the fact that the filter design method based on continued fraction expansion is not very effective at fitting frequency response data, particularly because this method does not allow the user to freely choose the appropriate frequency band to be fitted. This section introduces the more practical Oustaloup filter.

B. Oustaloup Filter

Assuming the frequency band of interest is (w_b, w_h), the linear properties of the magnitude-frequency response of the fractional-order operator can be approximated by the set of poly-lines. The French professor Alain Oustaloup and his colleagues proposed a filter design method based on this idea [10]. All these poly-lines are generated frominteger-order zeros and poles, so that the slopes of the asymptotic lines of the Bode magnitude-frequency characteristics alternate between 0 dB/dec and 20 dB/dec. The frequency response itself will be well approximated by a straight line. The standard form of Oustaloup filter is [10]

$$\hat{G}(s) = C \prod_{k=-N}^{N} \frac{1+s/z_k}{1+s/w'_k} \quad (1)$$

Using the following set of synthesis formulas:

$$w_0' = \alpha^{0.5} w_u;\ w_0 = \alpha^{0.5} w_u;\ \frac{w'_{k+1}}{w'_k} = \frac{w_{k+1}}{w_k} = \alpha.\eta > 1 \quad (2)$$

$$\frac{w_{k+1}}{w_k} = \eta > 1;\ \frac{w_k}{w'_k} = \alpha > 0; N = \frac{\log(w_N/w_0)}{\log(\alpha.\eta)};\ \mu = \frac{\log \alpha}{\log(\alpha.\eta)} \quad (3)$$

w_u is the unity frequency gain and the central frequency of a band of frequencies geometrically distributed around it. Thus, $w_u = \sqrt{w_h . w_b}$; where w_h et w_b are the high and low frequencies, respectively.

C. Reduction of Fractional-Order Models [26, 27]

Model reduction, as the name suggests, involves approximating a high-order model with a lower-order counterpart. The use of filters to approximate fractional-order operators often results in high-order transfer functions. To achieve the best reduced-order model, defining an optimization criterion is crucial. For instance, by exciting the original model and the reduced-order model with the same signal, an error signal can be derived from the two output signals. An objective function is then formulated based on this error signal. By leveraging this objective function, the problem is transformed into a numerical optimization problem. The original problem can only be solved using numerical optimization techniques. The idea proposed in Reference [12] leads to the optimal model reduction algorithm.

Optimal Model Reduction Algorithm:

1. Input the High-Order Model: Enter the initial high-order model.
2. Select Desired Orders: Choose the expected orders for the numerator and denominator of the reduced-order model.
3. Define the Objective Function: Write a function to describe the objective function based on the error signal.
4. Solve for Optimal Reduction: Utilize a solver to find the optimal reduced-order model.

II. SYSTEMS MODELING:

Figure 1 illustrates a fundamental liquid, level control system consisting of three tanks, B, C, and D, in addition to a primary tank, E, which provides liquid to the pump. The system initially computes the discrepancy between the target liquid level and the actual measured level to regulate the liquid level within the tanks. The PID controller processes this discrepancy to provide a control signal. The control signal regulates an electric valve with an adjustable aperture ranging from 0% to 100%. The system regulates the liquid level in the tanks by adjusting the flow rate through this valve [28, 30].

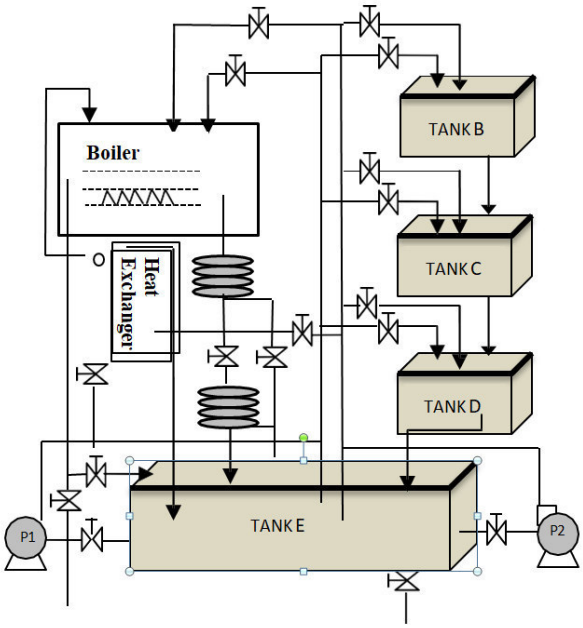


Fig. 1. Three Tanks Liquid Level System.

A simplified representation of Tank B, C, or D is shown in Figure 2.

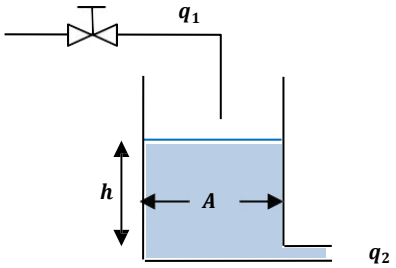


Fig. 2. Simple Structure of the Tanks.

Based on the principle of mass balance:

$$q_2 - q_1 = A(\frac{dh}{dt}) \quad (1)$$

$$q_1 = k_1 \times v_0 \quad (2)$$

$$q_2 = k_2 \times \sqrt{h} \quad (3)$$

In this context, inflow and outflow rates are denoted by q_1 and q_2, respectively. The tank's cross-sectional area is represented by A, while h signifies the liquid level height within the tank. Furthermore, v_0 indicates the valve's operational range, with k_1 and k_2 representing the proportional flow coefficients for the inlet and outlet valves, respectively.

Upon examining Equation (3), it becomes apparent that the liquid level dynamics exhibit nonlinear characteristics**.** Nevertheless, to facilitate a more straightforward analysis, linearization of Equation. (3) is permissible in the vicinity of

the equilibrium state (V_0, h_0), yielding an approximated linear representation.

$$\frac{q_2}{h} \approx \frac{k_2}{\sqrt{h_0}} = k_3 \tag{4}$$

Perform the Laplace transformation of Equations. (1), (2), and (4):

$$\frac{\mathrm{H(s)}}{\mathrm{V_o(s)}} = \frac{k}{\mathrm{Ts+1}} \tag{5}$$

Where

$k = k_1/k_3, T = A/k_3$. For $k_1 = 2,\ k_3 = 5, \tau = 0, A = 20$; according to (Kumar et al. 2011), the three-tank system's total transfer function is as follows:"

$$G_p(s) = \frac{H(s)}{Q(s)} = \frac{1}{64s^3+9.6s^2+0.48s+0.008} \tag{6}$$

The dynamic characteristics of our system, as encapsulated in Equation 6, indicate a slow-response system that closely resembles the behavior of a first-order system, particularly in its time-domain response. Figure 3 provides a graphical representation of the system's step response, offering further insight into its dynamic behavior.

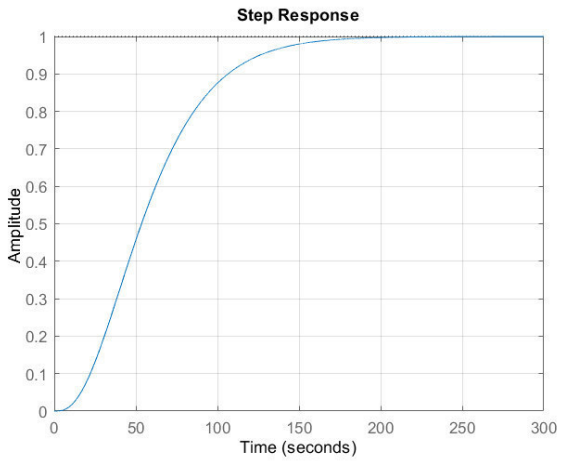


Fig. 3. Step response of the system.

III. Design of Fractional-Order - PID Controllers

PID control, a stalwart in industrial control systems, boasts a rich history as one of the earliest and most ubiquitous strategies. Its widespread adoption can be attributed to the relative simplicity of its design methodologies and intuitive control structures, making it an ideal fit for industrial applications. Notably, PID controllers do not necessitate precise system modeling, and their performance is generally satisfactory, solidifying their position as the most prevalent controller type in process industries. Over the past three decades, PID controllers have experienced a resurgence in academic interest, driving the development of innovative design approaches.

In contrast to traditional PID controllers, where integration and differentiation orders are uniformly set to one, Professor Igor Podlubny's seminal work in 1999 introduced the paradigm-shifting concept of fractional calculus into PID controller design. This innovation gave rise to the $PI^{\lambda}D^{\mu}$ structure, offering heightened flexibility through an increased number of tunable parameters. The optimal design methodologies that have since emerged, particularly for fractional-order PID controllers, have demonstrated marked effectiveness. Here, objective functions play a pivotal role in achieving optimal control outcomes, especially within the realm of servo control systems.

Fractional-Order PID Controllers

The fractional-order PID controller, denoted as $PI^{\lambda}D^{\mu}$, owes its development to Professor Igor Podlubny [22], who innovatively introduced two supplementary tuning parameters: λ, governing the integrator's order, and μ, determining the differentiator's order. This enhancement enables more nuanced control over the system's response.

The typical $PI^{\lambda}D^{\mu}$ controller is mathematically expressed as

$$G_c(s) = \left(k_p + \frac{k_i}{s^{\lambda}} + k_d s^{\mu}\right) \tag{9}$$

Where K_p (Proportional Gain), K_i (Integral Gain) and K_d(Derivative Gain) are three tunable parameters. The three parameters K_p, K_i and K_d can be tuned such that satisfactory control behaviours are finally reached.

IV. Objective Functions for Servo Control :

In the context of control systems, input signals are denoted by $r(t)$, while output responses are represented by $y(t)$. The primary objective of such systems is to achieve precise tracking of the input signal's dynamics by the output, a concept commonly encapsulated within servo control paradigms [2, 23].

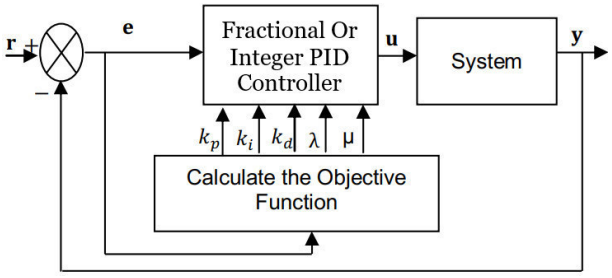


Fig. 4. Fractional/Integer PID Control of System.

Within the framework of servo control systems, two signals play crucial roles:

- $e(t)$, known as the error signal, which quantifies the discrepancy between desired and actual outputs.
- $u(t)$, referred to as the control signal, which is responsible for mitigating this error.

The principal aim in servo control systems is to minimize the error signal's magnitude over time [24]. Given the dynamic nature of e(t), integral-based performance metrics are frequently employed, as they effectively capture the cumulative error over a specified timeframe, providing a weighted assessment of system performance [2, 22].

The objective functions are commonly used criteria in the field of industrial control to evaluate the performance of a control system. These objective functions allow for the quantification of a system's response quality to disturbances or reference signals and are therefore essential for designing and optimizing control systems. The most commonly used objective functions are:

A. Integral of Squared Error (ISE)

The criterion ISE is defined as

$$J_{ISE} = \int_0^\infty \mathrm{e}^2(\mathrm{t})\mathrm{dt} \tag{10}$$

ISE minimizes large errors because the square of the error amplifies them.

- **Advantages**: ISE is often used in systems where large errors must be heavily penalized.
- **Disadvantages**: It can be sensitive to initial errors or sudden disturbances. This criterion prioritizes the rapid reduction of large errors but may neglect small persistent errors.

B. Integral of Absolute Error (IAE)

The criterion IAE is defined as

$$J_{\mathrm{IAE}} = \int_0^\infty |e(t)|\mathrm{dt} \tag{11}$$

IAE is less severe than ISE on large errors, as it does not square them.

- **Advantages**: IAE favors a more uniform reduction of errors, minimizing total error without giving too much importance to large errors.
- **Disadvantages**: It may take longer to reduce large errors, as it does not penalize them as much as ISE. This criterion may favor a solution where small errors persist over a long period.

C. Integral of Time-weighted Absolute Error (ITAE)

The criterion ITAE is defined as

$$J_{\mathrm{ITAE}} = \int_0^\infty t|e(t)|\mathrm{dt} \tag{12}$$

ITAE weighs errors that occur later in time more heavily, favoring rapid and stable responses.

- **Advantages**: ITAE is often used in systems where it is important to minimize the impact of persistent errors. It favors rapid transient responses with fewer oscillations.
- **Disadvantages**: ITAE can be more difficult to calculate and interpret. It may also penalize long-term errors more severely than initial errors.

D. Integral of Time-weighted Squared Error (ITSE)

The criterion ITSE is defined as

$$J_{\mathrm{ITSE}} = \int_0^\infty t\mathrm{e}^2(\mathrm{t})\mathrm{dt} \tag{13}$$

ITSE, by weighting the squared error by time, gives more importance to errors that occur later in the response.

- **Advantages**: This criterion is well-suited for systems where it is critical to minimize long-term errors. It heavily penalizes persistent errors, particularly those of large amplitude.
- **Disadvantages**: Like ITAE, ITSE is more complex to implement due to the temporal weighting and squaring of the error.

V. GLOBAL OPTIMIZATION ALGORITHM [16-20] :

The analysis and design of control systems heavily rely on optimization techniques to achieve optimal performance. However, local minimum points can be encountered if the initial search points are not selected appropriately. To overcome this challenge, our goal is to avoid local minimum points and discover the global minimum points that represent the optimal solution.

In control systems, the implementation is formulated as a minimization problem, where the objective function is employed as a time-domain metrics tool to evaluate the system's performance. The objective function is formulated to minimize the steady-state error (E_{ss}), percentage overshoot (OS), settling time (T_s), and rise time (T_r).

To address the issue of local minima, we employ a global search algorithm that conducts repeated local searches from random starting points. This approach allows for a thorough

exploration of the search space and increases the likelihood of identifying the global optimal solution. By employing this algorithm, we can ensure that the control system is optimized to achieve the best possible performance, avoiding local minima and converging to the global minimum point.

This approach is particularly useful in control systems where the objective function has multiple local minima, and the optimal solution is not easily attainable through traditional optimization methods. By using a global search algorithm, we can overcome this challenge and achieve optimal performance in control systems.

A global optimization algorithm, [19, 23]. Proceed in the following way.

1. Assign a large value f_x for the initial objective function.
2. Select randomly an initial search point x0.
3. Find the optimal solution from x0, and find x and the objective function f_1.
4. Check whether $f_1 < f_x$. If so, $f_1 < f_x$, record x; otherwise, go to (2). Run this in a loop for N times to find the best x and f_x.

VI. Comparative Simulation Resultats and Discussion

This study investigates the application of an optimized fractional PID controller, where the objective function is varied to match the transfer function of Equation 6 of the closed-loop regulation system for liquid level control.

A. The Parameters of the FOPID

The controller parameters for the three-tank system are summarized in Table 1.

TABLE I. PARAMETERS OF FOPID

Lower bounds for $[K_p, K_i, K_d, \lambda, \mu]$	Upper bounds for $[K_p, K_i, K_d, \lambda, \mu]$	initial values
[0, 0, 0, 0, 0]	[1, 1, 1, 1, 1]	0<Random values<1

B. Resultats and Discussion

The results of the simulation, which include rise time, settling time, overshoot, and steady-state error, as well as the minimum value of the objective function and the closed-loop transfer functions, are summarized in Table 2. This table offers a comprehensive overview of the performance of the proposed control system, allowing for a detailed analysis of its behavior and a comparison of its performance across different objective functions.

The closed-loop transfer functions corresponding to each of the selected objective functions are shown in Figure 5.

TABLE II. THE RESULTS

	J	$[K_p, K_i, K_d, \lambda, \mu]$	Closed-Loop Step Responses $G_J(s)$	T_r	T_s	OS	E_{ss}
J_{ITAE}	351.4	[0.01212 ,0.00086 ,0.263054 ,0.828983, 0.901314]	$\frac{0.005321}{s^2 + 0.1149s + 0.005363}$	32.18	51	0.771	0.0012
J_{IAE}	23.02	[0.011409 , 0.00088 , 0.240195 , 0.828939, 0.762986]	$\frac{-0.3959\,s^2 + 0.5463s + 0.01072}{s^6 + 20.91s^5 + 29.23s^4 + 84.95s^3 + 9.78s^2 + 0.804s + 0.011} e^{-0.538s}$	24.59	105	0.725	0.0054
J_{ISE}	9.62	[0.012239 ,0.00088 ,0.849994 ,0.828729, 0.802528]	$\frac{15.33\,s^2 + 38.33\,s + 0.4362}{s^4 + 2305s^3 + 220.3s^2 + 50.95s + 0.4381}$	10.55	114	20.72	0.0658
J_{ITSE}	128.47	[0.012153 ,0.00088 ,0.368538 ,0.828998, 0.86169]	$\frac{0.002595\ s^2 + 0.005635s + 0.0003075}{s^3 + 0.1084s^2 + 0.01277s + 0.0003099}$	26.15	104	0.00	0.0003

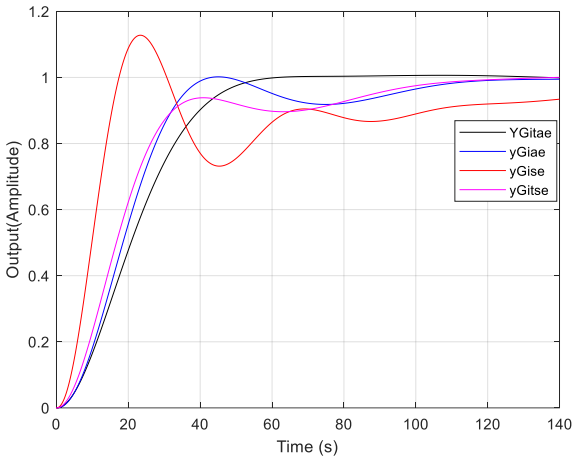


Fig. 5. Closed-Loop Step Responses for Different PID Controllers

A comparative study of the performance of FOPID controllers designed using different objective functions has revealed interesting results. According to the results of Figure 4 and Table 2, it is clear that the best FOPID controller is the one designed using the 'Integral of Time-weighted Absolute Error (ITAE)' objective function.

Indeed, the results of this study show that the four objective functions (J_(ITAE), J_(IAE), J_ISE, and J_ITSE) have different performances in terms of rise time, settling time, overshoot, and steady-state error. The results indicate that J_ISE has the lowest rise time, but with a high overshoot and steady-state error, while J_ITSE has the lowest overshoot and steady-state error.

However, it is important to note that the choice of objective function depends on the specific performance criteria of the system. For example, if the priority is given to the speed of response, J_ISE may be a good option, but if the priority is given to stability and accuracy, J_ITSE may be a better option.

On the other hand, J_ITAE seems to offer a good compromise between the two, with a reasonable rise time and stable performance in terms of overshoot and steady-state error. This is why it is considered the best FOPID controller in this study.

In conclusion, this comparative study has shown that the performance of FOPID controllers depends heavily on the chosen objective function. It is therefore important to choose the objective function that best suits the specific performance criteria of the system to obtain the best results.

VII. Conclusion

In conclusion, this study has demonstrated the importance of selecting the appropriate objective function in the design of Fractional Order Proportional-Integral-Derivative (FOPID) controllers for three-tank liquid level control systems. Through a comparative analysis of four commonly used objective functions - Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Square Error (ITSE), and Integral Time Absolute Error (ITAE) - we have shown that the performance of FOPID controllers depends heavily on the chosen objective function. The results of this study highlight the significance of considering response time, overshoot, steady-state error, and robustness against disturbances when selecting an objective function for FOPID controller design.

The findings of this research provide valuable insights for control system designers, emphasizing the need to carefully select the objective function to meet specific performance requirements. By choosing the right objective function, designers can optimize FOPID controller performance and achieve stable and robust control of three-tank liquid level systems.

Overall, this study contributes to the advancement of control system design for complex systems, highlighting the importance of objective function selection in FOPID controller optimization. The results and guidelines presented in this research can be applied to a wide range of control applications, including process control, robotics, and mechatronics.

References

[1] *I. Podlubny, "Fractional-order systems and controllers," IEEE Trans. Autom. Control, vol. 44, no. 1, pp. 208–214, Jan. 1999.*

[2] *A. A. Kesarkar and S. Narayanasamy, "Tuning of optimal fractional-order PID controller using an artificial bee colony algorithm," Syst. Sci. Control Eng. Open Access J., vol. 2, no. 1, pp. 385–391, Dec. 2014.*

[3] *J. Y. Cao and B. G. Cao, "Design of fractional-order controller based on particle swarm optimization," Int. J. Control Autom. Syst., vol. 4, no. 6, pp. 775–781, 2006.*

[4] *A. Tepljakov, E. Petlenkov, and J. Belikov, "FOMCON: A MATLAB toolbox for fractional-order system identification and control," in Proc. Int. Conf. Methods and Models in Automation and Robotics (MMAR), 2011, pp. 1–4.*

[5] *M. Axtell and M. E. Bise, "Fractional calculus applications in control systems," in Proc. IEEE Nat. Aerosp. Electron. Conf., New York, USA, 1990, pp. 563–566.*

[6] *A. Idir, L. Canale, Y. Bensafa, and K. Khettab, "Design and robust performance analysis of low-order approximation of fractional PID controller based on an IABC algorithm for an automatic voltage regulator system," Energies, vol. 15, no. 23, p. 8973, 2022.*

[7] *B. M. Vinagre, Y. Q. Chen, and I. Petráš, "Two direct Tustin discretization methods for fractional-order differentiator/integrator," J. Frankl. Inst., vol. 340, no. 5, pp. 349–362, 2003.*

[8] *A. Idir, M. Kidouche, Y. Bensafia, K. Khettab, and S. A. Tadjer, "Speed control of DC motor using PID and FOPID controllers based on differential evolution and PSO," Int. J. Intell. Eng. Syst., vol. 11, no. 4, pp. 241–249, 2018.*

[9] *A. Oustaloup, "Fractional order sinusoidal oscillators: Optimization and their use in highly linear FM modulation," IEEE Trans. Circuits Syst., vol. 28, no. 10, pp. 1007–1009, 1981.*

[10] *R. P. Tripathi, A. K. Singh, and P. Gangwar, "Innovation-based fractional order adaptive Kalman filter," J. Electr. Eng., vol. 71, no. 1, pp. 60–64, 2020.*

[11] *Y. Bensafia, A. Idir, A. Zemmit, and K. Khettab, "Performance improvement of aircraft pitch angle using the fractional order adaptive PID controller," Przegląd Elektrotechniczny, vol. 99, no. 5, pp. 98–101, 2023.*

[12] *A. Oustaloup, X. Moreau, and M. Nouillant, "The CRONE suspension," Control Eng. Pract., vol. 4, no. 8, pp. 1101–1108, 1996.*

[13] *V. Goyal, P. Mishra, A. Shukla, V. K. Deolia, and A. Varshney, "A fractional order parallel control structure tuned with metaheuristic optimization algorithms for enhanced robustness," J. Electr. Eng., vol. 70, no. 1, pp. 16–24, 2019.*

[14] *I. Podlubny, L. Dorcak, and I. Kostial, "On fractional derivatives, fractional-order dynamic systems, and PID controllers," in Proc. 36th IEEE Conf. Decision Control, San Diego, CA, USA, 1997, pp. 498–502.*

[15] *S. Ladaci, J. J. Loiseau, and A. Charef, "Adaptive internal model control with fractional order parameter," Int. J. Adapt. Control Signal Process., vol. 24, no. 10, pp. 944–960, 2010.*

[16] *D. Y. Xue and Y. Q. Chen, "Suboptimum H2 pseudo-rational approximations to fractional-order linear time-invariant systems," in Advances in Fractional Calculus – Theoretical Developments and Applications in Physics and Engineering, Springer, Dordrecht, 2007, pp. 61–75.*

[17] *D. Y. Xue and Y. Q. Chen, "OptimFOPID: A MATLAB interface for optimum fractional-order PID controller design for linear fractional-order plants," Proc. Fractional Derivatives Its Appl., Nanjing, 2012.*

[18] *D. Y. Xue and Y. Q. Chen, Modeling, Analysis, and Design of Control Systems in MATLAB and Simulink, Singapore: World Scientific, 2014.*

[19] *C. A. Monje, Y. Chen, B. M. Vinagre, D. Xue, and V. Feliu, Fractional-Order Systems and Controls: Fundamentals and Applications, Springer Science & Business Media, 2010.*

[20] *A. M. Dastjerdi, M. Zare, and H. Mohammadi, "Design and optimization of fractional-order PID controllers for higher-order time delay systems using meta-heuristic algorithms," ISA Trans., vol. 119, pp. 128–141, 2022.*

[21] *I. Petráš, Fractional-Order Nonlinear Systems: Modeling, Analysis and Simulation, Springer, 2019.*

[22] *V. Gupta and S. Banerjee, "A comparative study of optimization algorithms for fractional-order PID controller tuning," J. Control Autom. Electr. Syst., vol. 32, no. 3, pp. 542–552, 2021.*

[23] *H. Sun, Y. Zhang, Y. Chen, and Q. Yang, "Fractional-order PID controller optimization based on multi-objective genetic algorithm for industrial applications," Control Eng. Pract., vol. 132, p. 105365, 2023.*

[24] *R. Almeida and M. T. T. Monteiro, "On the design of fractional-order PID controllers for complex dynamical systems," J. Comput. Nonlinear Dyn., vol. 16, no. 10, p. 101014, 2021.*

[25] *F. Sabatino, B. B. Alagoz, and A. Tuna, "A comparative study on optimization algorithms for tuning FOPID controllers in real-time systems," Sensors, vol. 23, no. 1, p. 120, 2023.*

[26] D. Y. Xue and Y. Q. Chen, Suboptimum *H*2 pseudo-rational approximations to fractional-order linear time invariant systems, in: Advances in Fractional Calculus – Theoretical Developments and Applications in Physics and Engineering, Springer, Dordrecht (2007), 61–75

[27] D. Y. Xue and Y. Q. Chen, *Modeling, Analysis and Design of Control Systems in MATLAB and Simulink*, World Scientific, Singapore, 2014.

[28] X. Chen, G. Wu, W. Qiao, *et al.*, "3-Tank System Liquid Level Control Using Improved African Vulture Optimization Algorithm," in *2024 7th Int. Conf. Energy, Electr. Power Eng. (CEEPE)*, IEEE, pp. 635–641, 2024.

[29] L. Chen, "Principle and simulation PID controller of liquid level system," in *J. Phys.: Conf. Ser.*, IOP Publishing, p. 012187, 2021.

[30] X. Li, J. Li, and L. Shi, "Modeling and simulation of water level system," in *2008 IEEE Int. Conf. Autom. Logist.*, IEEE, pp. 2856–2859, Sep. 2008.

Comparative Analysis of Torsional Vibration Mitigation Strategies for Enhanced Rotary Drilling Efficiency

1st Zebboudj Ahmed Hichem
Department of Automation and Electrification of Processes
Laboratory of Applied Automation (LAA)
Faculty of Hydrocarbons and Chemistry (FHC), Institute of Electrical & Electronic Engineering (IGEE)
University M'hamed Bougara
Boumerdes, Algeria
a.zebboudj@univ-boumerdes.dz

2nd Akroum Hamza
Department of Automation and Electrification of Processes
Laboratory of Applied Automation (LAA), Systems Engineering and Telecommunications Laboratory
Faculty of Hydrocarbons and Chemistry (FHC)
University M'hamed Bougara
Boumerdes, Algeria
akroum@univ-boumerdes.dz,

3rd Doghmane Z.Mohamed
Department of Automation and Electrification of Processes
Laboratory of Applied Automation (LAA)
DOE, Exploration Division, SONATRACH, Bob L. Herd Department of Petroleum Engineering
Faculty of Hydrocarbons and Chemistry (FHC)
Boumerdes, Algeria
mohamedzinelabidine.Doghmane@Sonatrach.dz

4th Madjid Kidouche
Electrification of Processes
Laboratory of Applied Automation (LAA)
Faculty of Hydrocarbons and Chemistry (FHC), IGEE
University M'hamed Bougara
Boumerdes, Algeria
mkidouche@univ-boumerdes.dz

***Abstract*— Torsional vibrations, specifically the Stick-Slip phenomenon, present a persistent challenge in rotary drilling systems, affecting drilling efficiency, equipment durability, and hole quality. This study offers a comparative analysis of two key vibration mitigation strategies: sliding mode control (SMC) and a Kalman filter-based Proportional-Integral-Derivative (PID) control approach. The SMC method addresses parameter uncertainties and disturbances by acting as both observer and controller, maintaining stability through real-time adjustments to suppress Stick-Slip vibrations. Alternatively, the Kalman filter estimates drilling tool velocity using sensor data, which is then integrated into a PID control strategy to modulate torque and reduce torsional oscillations. Simulation results demonstrate each method's effectiveness in minimizing Stick-Slip behavior, enhancing drilling efficiency, and reducing equipment wear. This comparative analysis highlights the strengths and adaptability of each technique, offering insights for selecting optimal vibration mitigation strategies to improve rotary drilling performance across diverse operational scenarios.**

Keywords*— *torsional vibration, stick-slip, controller, observer, Kalman filter, sliding mode.

I. Introduction

Rotary drilling methds are the backbone of contemporary drilling operations, allowing the drill bit to successfully penetrate subterranean rock types by rotation. However, these systems are intrinsically complicated, with the drill string being a lengthy, flexible, rotating component subjected to a variety of dynamic stresses[1]. Torsional vibrations are distinct from other types of vibrations experienced during drilling because they have a direct influence on rotational stability[2]. These vibrations frequently show due to the Stick-Slip phenomena[3], in which the drill bit alternates between sticking from static friction and sliding from unexpected kinetic motion. Over the years, it has become evident that Stick-Slip not only reduces drilling efficiency but also increases mechanical strains, making it difficult to maintain steady drilling operations[4].

The negative impact of torsional vibrations and stick-slip on drilling operations cannot be emphasized. In the past, Stick-Slip causes rapid wear and tear on drill bits and other important components, resulting in a shorter equipment lifespan and higher maintenance expenses[5]. Beyond equipment damage, it reduces operating efficiency by creating irregular penetration rates and lowering overall hole quality. There have been examples where extreme Stick-Slip caused catastrophic drill string failures, resulting in costly downtimes and lengthy delays[6]. For these reasons, have concentrated heavily on reducing this occurrence in order to improve drilling performance, assure operational safety, and reduce costly disruptions.

For controlling torsional vibration, several control strategies have been devised[7], [8]. One method includes utilizing observers for estimating the drill bit's rotational velocity, which aids in detecting the commencement of Stick-Slip[9], [10]. Sliding Mode Observers (SMO)[11] and Kalman filters[12] are two popular types of observers used in this situation. The SMO effectively handles uncertainties and disturbances, resulting in a robust prediction of the system state, whereas the Kalman filter processes sensor data to provide exact real-time velocity estimations. These velocity estimations are then used in control systems like proportional-integral-derivative (PID) and sliding mode control (SMC)[1], [13] to regulate torque and other drilling parameters, successfully minimizing torsional vibrations. These systems operate by continually modifying control settings to stabilize drill bit rotation, decrease equipment wear, and increase drilling efficiency.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

II. Fundamentals of Drilling Systems

Rotary drilling systems are an essential part of contemporary drilling technology, used to drill boreholes in the Earth's subsurface to recover oil, gas, geothermal energy, and other resources. These systems work by transmitting rotating motion from the surface to a drill bit, which grinds and slices through the rock formations. The rotation is often powered by surface equipment, such as a rotary table or top drive, which allows for continuous penetration into deeper strata. Rotary drilling is popular because of its efficiency, flexibility to different geological conditions, and capacity to drill to vast depths, making it an important instrument in energy production and resource research[14], figure 1 illustrates a schematic representation of the rotary drilling system.

A rotary drilling system consists of several essential parts, including surface equipment, the drill string, and subsurface tools[15]. The drill string, made up of linked drill pipes, distributes torque and weight to the drill bit at the bottom of the borehole. During operation, dynamic forces are applied to subsurface components such as stabilizers, collars, and measuring equipment, as well as the bit itself. Torsional vibrations, among these forces, have the greatest impact on the drill string and bit, creating cyclic accelerations and decelerations that might result in Stick-Slip. These vibrations provide a significant challenge to drilling efficiency and equipment integrity, especially in the subterranean environment, where mechanical stresses are most severe.

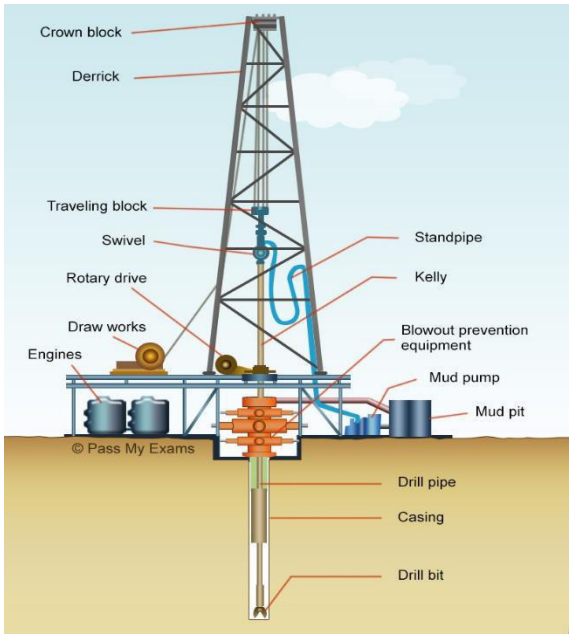


Fig. 1 Rotary drilling system descriptive schema.

III. Drill-String Model

The drill string is a fundamental component of rotary drilling systems, acting as a mechanical interface between surface equipment and the drill bit. It's a long, thin, and flexible structure made up of interconnecting drill pipes, hefty drill collars, and other subsurface instruments. The drill-string not only distributes rotational torque and axial force to the drill bit, but also serves as a conduit for drilling fluids, which lubricate the bit and transport cuttings to the surface.

drill-string model is required. This model generally depicts the drill-string as a mix of distributed mass, stiffness, and damping components to reflect its elastic and inertial characteristics. The model considers interactions with the borehole wall, changes in material qualities, and operational parameters like as rotating speed and applied weight-on-bit.

A. Drill-String Model

The drill-string model developed in this work, based on a three-degree-of-freedom (3-DOF) equation, is depicted in Figure 2, with its mathematical representation provided in eq (1).

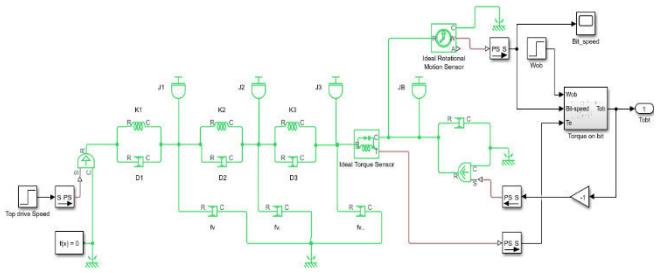


Fig. 2 Simscape model of the drill string

$$\begin{cases} \ddot{\Phi}_1 = \frac{K_1}{J_1}(\Phi_{td}-\Phi_1)+\frac{D_1}{J_1}(\dot{\Phi}_{td}-\dot{\Phi}_1)-\frac{K_2}{J_1}(\Phi_1-\Phi_2)-\frac{D_2}{J_1}(\dot{\Phi}_1-\dot{\Phi}_2)-\frac{f_v}{J_1}\dot{\Phi}_1 \\ \ddot{\Phi}_2 = \frac{K_2}{J_2}(\Phi_1-\Phi_2)+\frac{D_2}{J_2}(\dot{\Phi}_1-\dot{\Phi}_2)-\frac{K_3}{J_2}(\Phi_2-\Phi_3)-\frac{D_3}{J_2}(\dot{\Phi}_2-\dot{\Phi}_3)-\frac{f_v}{J_2}\dot{\Phi}_2 \\ \ddot{\Phi}_3 = \frac{K_3}{J_3}(\Phi_2-\Phi_3)+\frac{D_3}{J_3}(\dot{\Phi}_2-\dot{\Phi}_3)-\frac{f_v}{J_3}\dot{\Phi}_2-\frac{1}{J_3}T_{ob} \end{cases} \quad (1)$$

B. Rock-bit Interaction Model

The contact of the drilling tool and the rock, also known as rock-bit interaction, is crucial to drilling efficiency. As the drill bit cuts into the formation, it encounters resistance, which causes forces that alter the bit's movement. This interaction, controlled by rock hardness and weight-on-bit, can result in torsional vibrations such as Stick-Slip, in which friction forces the bit to alternate between sticking and sliding. Understanding this connection is critical for improving drilling efficiency and reducing damaging vibrations[16].

Tob is the torque on bit that represent the nonlinearity of the system, and it's given by eq (2).

$$T_{ob} = X_3\left(\frac{\alpha}{\sqrt{X_3{}^2+\Omega_0{}^2}}+\frac{\beta}{X_3{}^2+\Omega_0{}^2}\right)-DX_3\left(\frac{X_3}{\Omega_1}-1\right) \quad (2)$$

with $\Omega_0 = 1$, $\Omega_1 = 1$, $\alpha = \mu_n N_r$,and $\beta = Ap\Omega_0$

Table 1 presents a detailed description of the parameters used in the model

TABLE 1 Description of symbols used in the system model

Parameters	Description	Unit
$\dot{\Phi}_{td}$	the top drive angular velocity	*[Rad/s]*
$\Phi_{i;i=1,2,3}$	Respectively the angular displacement of the top drive, the drill pipes, and the BHA	*[rad]*

$J_{i=1,2,3}$	The equivalent moment of inertia of the rope section *i*	[$Kg.m^2$]
$K_{i=1,2,3}$	The stiffness coefficient of the rope section *i*	[$N.m/rad$]
$D_{i=1,2,3}$	The internal damping coefficient of the top drive, drill pipes and BHA respectively	[$N.m.S/rad$]
f_v	The viscous damping coefficient of the BHA	[N]
Tob	The nonlinear estimated value of the torque on bit	[$N.m$]
μ_n	Nominal friction coefficient	[40 Nm]
N_r	The force vector	[9.81Wob.N]
p	The initial friction parameter	1.5

C. Sliding Mode Observer-Based Controller Design

The sliding mode observer-based controller is an advanced technique for mitigating torsional vibrations in rotary drilling systems. The Sliding Mode Observer (SMO) is critical for accurately estimating the drill bit's rotational velocity in real time, even in the presence of system uncertainty and noise. The Sliding Mode Controller (SMC) uses this predicted velocity to alter drilling parameters such as torque, thereby suppressing torsional vibrations. One of the major benefits of this strategy is its resistance to disruptions and model imperfections. Furthermore, to improve performance, the controller is intended to reduce chattering, a typical issue in sliding mode control that can cause high-frequency oscillations and even damage the equipment[1]. By addressing both torsional vibrations and chattering, this approach improves drilling stability, efficiency, and equipment lifespan.

Eq (3) expresses the Sliding Mode Observer's open-loop model, which is derived from the system's state-space description in eq (1).

$$\begin{cases} \dot{\hat{X}}_1 = a_1\hat{X}_1 + a_2\hat{X}_2 + a_3\hat{X}_3 + a_4\hat{X}_4 + \dfrac{D_1}{J_1}U + L_1.v \\ \dot{\hat{X}}_2 = b_1\hat{X}_1 + b_2\hat{X}_2 + b_3\hat{X}_3 + b_4\hat{X}_4 - \dfrac{1}{J_2}Tob + L_2.v \\ \dot{\hat{X}}_3 = c_1\hat{X}_1 + c_2\hat{X}_2 + c_3\hat{X}_3 + c_4\hat{X}_4 + U + L_3.v \\ \dot{\hat{X}}_4 = d_1\hat{X}_1 + d_2\hat{X}_2 + d_3\hat{X}_3 + d_4\hat{X}_4 + L_4.v \end{cases} \quad (3)$$

where $v = K.\mathrm{sgn}(S)$ with $K > 0$ and the sliding variable is given as: $S = X_1 - \hat{X}_1$

To design the sliding mode conttroller a sliding surface is given by eq (4):

$$S_2 = (X_1 - Q_{ref}) + \lambda[\int_0^t (X_1 - Q_{ref})dz + \int_0^t (X_1 - \hat{X}_3)dz] \quad (4)$$

Where $\lambda > 0$

To ensure stability, the Lyapunov function described in eq (5) is used[17].

$$V = \frac{1}{2}{S_2}^2 \quad (5)$$

$$\dot{V} = S_2.\dot{S}_2$$

As a result, eq (6) represents the equivalent control command.

$$\dot{S}_2 = 0$$

$$U_{eq} = -\frac{J1}{D1}[a_1X_1 + a_2\hat{X}_2 + a_3\hat{X}_3 + a_4\hat{X}_4 + \lambda(X_1 - Q_{ref}) + \lambda(X_1 - \hat{X}_3)] \quad (6)$$

Where:

$$\begin{cases} a_1 = -(\dfrac{D_1 + D_2 + f_v}{J_1}); a_2 = 0; a_3 = \dfrac{K_1}{J_1}; a_4 = \dfrac{K_2}{J_1} \end{cases}$$

To make $\dot{V} < 0$, The control law is specified by eq(7).

$$U = Ueq + Ud \quad (7)$$

The discontinuous control law is specified in eq(8).

$$Ud = -K\,\mathrm{sgn}(S_2) \quad (8)$$

D. Kalman Filter -Based PID controller design

The Kalman filter-based PID controller is an useful solution for reducing torsional vibrations in drilling equipment. The Kalman filter calculates the drill bit's velocity in real time by analyzing noisy sensor data and taking into account system dynamics, resulting in correct state information. These velocity estimations are then input into a proportional-integral-derivative (PID) controller, which changes drilling parameters like torque to reduce vibrations and stabilize the machine. This technology improves drilling efficiency and reduces mechanical stress by combining the accuracy of the Kalman filter with the flexibility of the PID controller, even under changing operating circumstances[12].

- State Prediction Equation

This equation predicts the present state of a system at the next time step by taking the previous state and any control input into account. It is sometimes referred to as the "Time Update", eq (9).

$$\hat{x}(k) = A * \hat{x}(k-1) + B * u(k) + \Gamma * T_{ob}(k) \quad (9)$$

Predict the covariance matrix for the predicted state, eq (10):

$$P(k) = A * P(k-1) * A^T + Q \quad (10)$$

- Measurement Update Equation

This equation updates the state estimate by incorporating the difference between the actual measurement and the predicted state, along with the measurement noise. It is used to compute the Kalman gain, *K(k)*, as in eq (11)

$$K(k) = \frac{P^{-}(k) * C^{T}}{\left(C * P^{-}(k) * C^{T} + R\right)} \quad (11)$$

Update the state estimate, given by eq(12):

$$\hat{x}(k) = \hat{x}^{-}(k) + K(k) * \left[y(k) - C * \hat{x}^{-}(k)\right] \quad (12)$$

Update the error covariance matrix, given by eq (13):

$$P(k) = \left(I - K(k) * C\right) * P^{-}(k) \quad (13)$$

In this context, y(k) represents the actual measurement at time (k), while (C) is the measurement matrix that links the state to the measurements. $\hat{x}^{-}(k)$ denotes the predicted state estimate at time (k), and $P^{-}(k)$ is the predicted error covariance. $K(k)$ is the Kalman gain at time (k), which determines the weight given to the measurement update in refining the state estimate. (R) is the measurement noise covariance, reflecting the uncertainty in the measurements and their relationship with the state variables.

Eq (14) describes the mathematical basis of a PID controller. The tune function in MATLAB Simulink makes it simple to find the PID controller's settings and change the proportional, integral, and derivative gains to achieve optimal system performance.

$$u(k) = K_p e(k) + K_i \sum_{i=0}^{k} e(i) T_s + K_d \frac{e(k) - e(k-1)}{T_s} \quad (14)$$

And $e(k) = Q_{ref} - \hat{\dot{\Phi}}_3(k)$

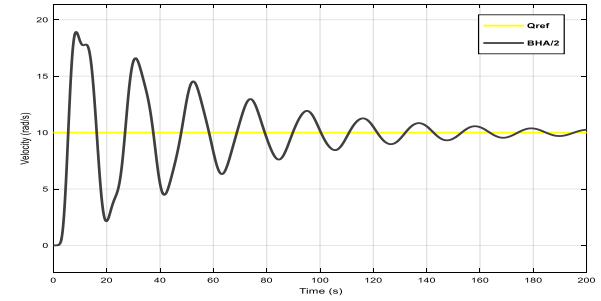
Here, $u(k)$ represents the control output, $e(k)$ is the error (the difference between the setpoint and the process variable). The constants K_p, K_i, and Kd correspond to the proportional, integral, and derivative gains, respectively. To achieve the desired control objectives and optimize system performance, the PID controller can be fine-tuned by adjusting these gains. However, this process of gain optimization may require some time to ensure that the system operates efficiently and effectively.

IV. Results And Discusion

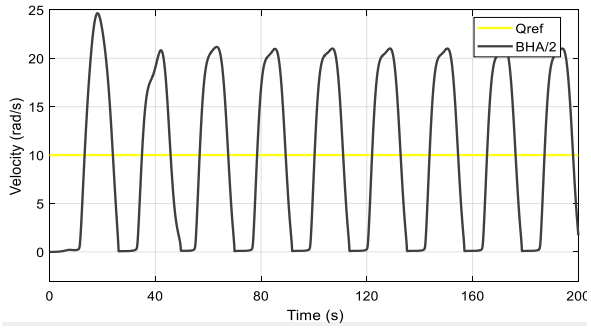
The MATLAB R2018a Simscape/Simulink environment was used to perform a thorough examination of the open-loop model's properties. This research revealed the existence of strong high-frequency torsional vibrations associated with the Stick-Slip phenomena in the system. Furthermore, the closed-loop model was tested with two sophisticated control strategies: a Sliding Mode Observer (SMO)-based controller and a Kalman filter-integrated PID controller. The study sought to evaluate the controllers' capacity to effectively attenuate torsional vibrations, increase system stability, and improve overall drilling performance by decreasing the negative impacts of the Stick-Slip phenomena.

A. *Open-Loop Responses*

Figure 3(a) displays the system's behavior at a Weight on Bit (WOB) of 150N, when torsional vibrations begin with considerable strength before progressively decreasing due to the reduced applied load. These vibrations indicate the system's response to the lowered WOB, showing modest oscillatory dynamics. Figure 3(b) shows the impact of using a greater WOB of 4005N, which results in substantially more aggressive drilling dynamics. The drill bit exhibits significant high-frequency oscillations, indicating an enhanced torsional reaction. This increased vibration causes the distinctive Stick-Slip cycle, which alternates between periods of stopped motion and fast sliding phases.



(a)

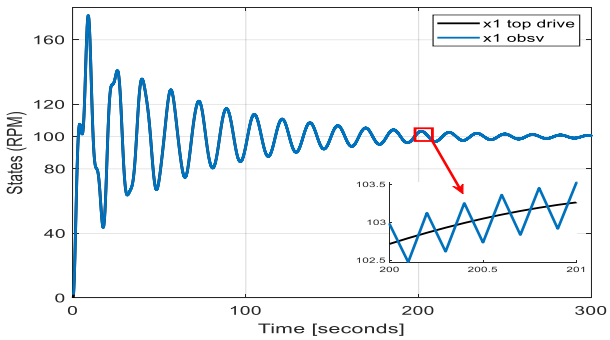


(b)

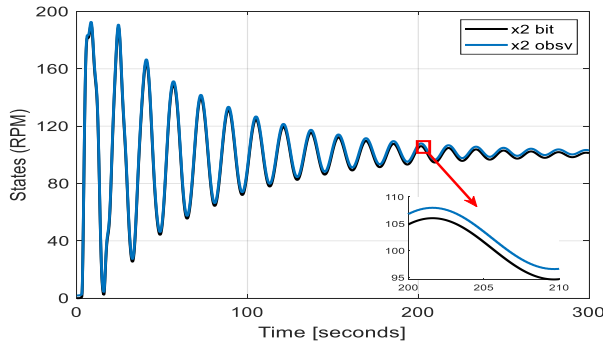
Fig. 3 Stick-Slip severity for Qref=10rpm with: (a) Wob=150N, (b) Wob = 4005N

B. *SMO-SMC Results*

Figures 4(*a*) and 4(b) show the Sliding Mode Observer's (SMO) step reaction for top drive and drill bit velocities, respectively, while Figure 5 shows the drill bit torque and its estimated value. The angular speed of the tool and its predicted counterpart have a constant offset, and both the drill bit and top drive velocities settle after around 200 seconds without the need of a controller. Despite this, the calculated values roughly match the required values in both situations. Furthermore, Figure 4(b) shows that in the absence of a controller, the drill bit velocity attempts to follow the top drive velocity but exhibits substantial oscillations and a long settling period.



(a)



(b)

Fig. 4 Step response of SMO in open loop for: (*a*) the Top Drive Velocity, (b) the Bit Velocity

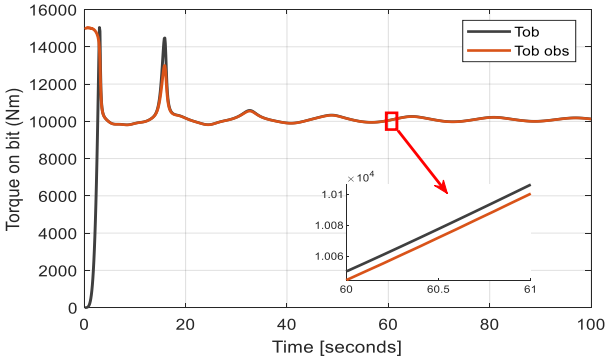


Fig. 5 Step response of SMO in open loop for the Torque on Bit.

Figure 6 depicts the rotary drilling system's (RDS) step reaction to angular velocities of the drill bit and top drive using the Sliding Mode Controller. In this investigation, a three-degree-of-freedom RDS model was used, which included a preset bit torque model. The findings show a notable overshoot, followed by the drill bit velocity aligning with the top drive angular speed with no static error, resulting in a quick settling period of 20 seconds.

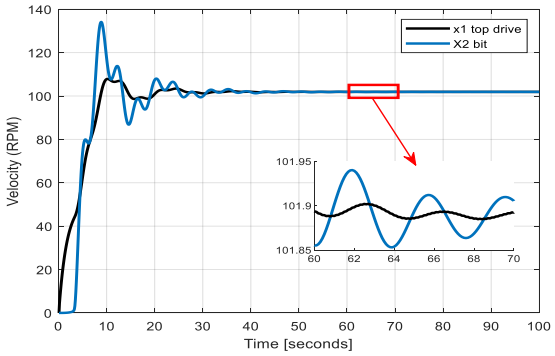


Fig. 6 SMC step response for the angular velocities

The response of the Sliding Mode Controller to the control input is depicted in Figure 7

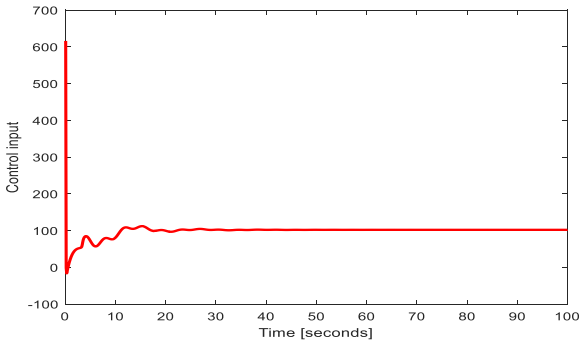


Fig. 7 Control Input

C. Kalman Filter-PID Results

Figure 8 shows the dynamic state estimate using the Kalman filter and a constant Weight on Bit (WOB) of 150N. The findings demonstrate the filter's ability to properly estimate drill bit velocity, closely track the reference trajectory, and capture the system's dynamic behavior. While the calculated velocity closely matches the required profile, a little static error is detected, suggesting modest departures from the reference value.

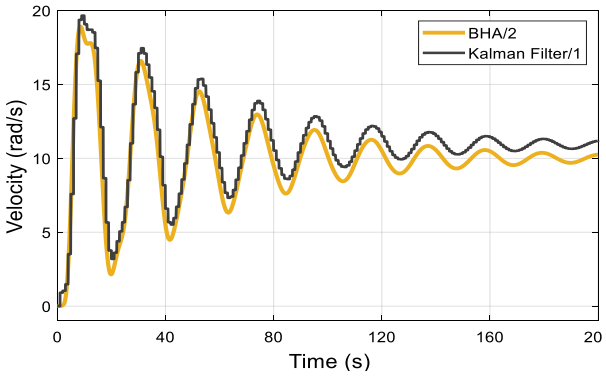


Fig. 8 The bit velocity response using KF with Wob=150N

Figure 9 illustrates how incorporating a PID controller with Kalman filter state estimations improves system performance while dramatically lowering torsional vibrations. While residual oscillations persist, their strength is significantly reduced. Extended observation periods indicate slow system changes, emphasizing the significance of assessing long-term performance to assure stability and optimize control techniques for a wide range of operating situations.

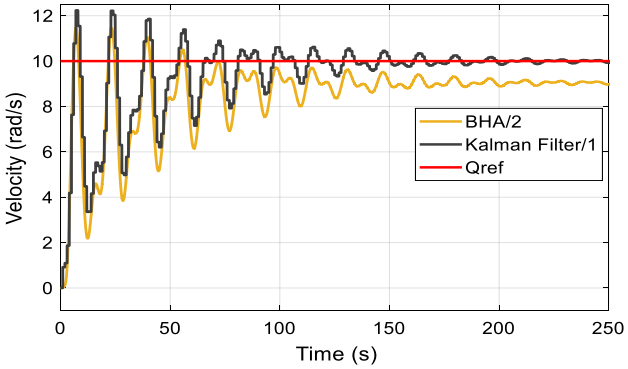


Fig. 9 the bit velocity response using KF-PID with Wob=150N

CONCLUSION

In conclusion, this work effectively investigated advanced control solutions to reduce stick-slip in rotary drilling systems. Both the Kalman filter-based PID controller and the Sliding Mode Controller (SMC) shown promise for minimizing torsional vibrations and enhancing system performance. The Kalman filter produced correct drill bit velocity estimates, demonstrating its usefulness in dynamic state estimation and greatly improving the performance of the PID controller. However, the SMC outperformed the PID controller in terms of quicker stabilization, better handling of nonlinear dynamics, and increased resilience to uncertainties, making it a more efficient option for real-time drilling operations. The Sliding Mode Observer's integration with the SMC was also efficient in predicting drill bit velocity, providing a low-cost alternative to standard sensor-based approaches. Overall, the findings suggest that the SMC-based approach holds significant potential for enhancing drilling efficiency, and future work will focus on further optimization and comparison with other control strategies.

REFERENCES

[1] C. Mendil, M. Kidouche, and M. Z. Doghmane, "Sliding Mode Controller Design for Torsional Vibrations Minimization Under Rock-Bit Interaction Effects," in *Artificial Intelligence and Heuristics for Smart Energy Efficiency in Smart Cities*, vol. 361, M. Hatti, Ed., in Lecture Notes in Networks and Systems, vol. 361. , Cham: Springer International Publishing, 2022, pp. 613–624. doi: 10.1007/978-3-030-92038-8_61.

[2] R. Riane, M. Kidouche, M. Z. Doghmane, and R. Illoul, "Modeling of Torsional Vibrations Dynamic in Drill-String by Using PI-Observer," in *Proceedings of the 4th International Conference on Electrical Engineering and Control Applications*, vol. 682, S. Bououden, M. Chadli, S. Ziani, and I. Zelinka, Eds., in Lecture Notes in Electrical Engineering, vol. 682. , Singapore: Springer Nature Singapore, 2021, pp. 171–185. doi: 10.1007/978-981-15-6403-1_12.

[3] B. Besselink, T. Vromen, N. Kremers, and N. Van De Wouw, "Analysis and Control of Stick-Slip Oscillations in Drilling Systems," *IEEE Trans. Contr. Syst. Technol.*, vol. 24, no. 5, pp. 1582–1593, Sep. 2016, doi: 10.1109/TCST.2015.2502898.

[4] M. Z. Doghmane, A. Bacetti, and M. Kidouche, "Stick-Slip Vibrations Control Strategy Design for Smart Rotary Drilling Systems," in *Artificial Intelligence and Renewables Towards an Energy Transition*, vol. 174, M. Hatti, Ed., in Lecture Notes in Networks and Systems, vol. 174. , Cham: Springer International Publishing, 2021, pp. 197–209. doi: 10.1007/978-3-030-63846-7_20.

[5] I. Kessai, S. Benammar, M. Z. Doghmane, and K. F. Tee, "Drill Bit Deformations in Rotary Drilling Systems under Large-Amplitude Stick-Slip Vibrations," *Applied Sciences*, vol. 10, no. 18, p. 6523, Sep. 2020, doi: 10.3390/app10186523.

[6] M. Wu *et al.*, "Disturbance estimator and smith predictor-based active rejection of stick–slip vibrations in drill-string systems," *International Journal of Systems Science*, vol. 51, no. 5, pp. 826–838, Apr. 2020, doi: 10.1080/00207721.2020.1744046.

[7] C. Mendil, M. Kidouche, and M. Z. Doghmane, "Hybrid Backstepping Sliding Mode Controller for Stick–slip Vibrations Mitigation in Rotary Drilling Systems," *IETE Journal of Research*, pp. 1–11, May 2021, doi: 10.1080/03772063.2021.1919217.

[8] R. Riane, M. Z. Doghmane, M. Kidouche, K. F. Tee, and S. Djezzar, "Stick-Slip Vibration Suppression in Drill String Using Observer-Based LQG Controller," *Sensors*, vol. 22, no. 16, p. 5979, Aug. 2022, doi: 10.3390/s22165979.

[9] M. Fu, P. Zhang, J. Li, and Y. Wu, "Observer and reference governor based control strategy to suppress stick-slip vibrations in oil well drill-string," *Journal of Sound and Vibration*, vol. 457, pp. 37–50, Sep. 2019, doi: 10.1016/j.jsv.2019.05.050.

[10] R. Mlayeh, S. Toumi, and L. Beji, "Backstepping boundary observer based-control for hyperbolic PDE in rotary drilling system," *Applied Mathematics and Computation*, vol. 322, pp. 66–78, Apr. 2018, doi: 10.1016/j.amc.2017.11.034.

[11] R. Riane, M. Z. Doghmane, M. Kidouche, and S. Djezzar, "Observer-Based H∞ Controller Design for High Frequency Stick-Slip Vibrations Mitigation in Drill-String of Rotary Drilling Systems," *Vibration*, vol. 5, no. 2, pp. 264–289, Apr. 2022, doi: 10.3390/vibration5020016.

[12] Z. A. Hichem, A. Hamza, D. Z. Mohamed, and K. Madjid, "PID-Based Kalman Filter Controller Design Stick-Slip Vibrations Suppression in Drill String Systems," in *2024 2nd International Conference on Electrical Engineering and Automatic Control (ICEEAC)*, Setif, Algeria: IEEE, May 2024, pp. 1–7. doi: 10.1109/ICEEAC61226.2024.10576610.

[13] C. Mendil, M. Kidouche, and M. Z. Doghmane, "Hybrid sliding PID controller for torsional vibrations mitigation in rotary drilling systems," *Indonesian Journal of Electrical Engineering and Computer Science, Vol. 22, No. 1, April 2021, pp. 146~158, DOI: 10.11591/ijeecs.v22.i1.pp146-158*.

[14] M. Z. Doghmane, M. Kidouche, and A. Ahriche, "Decentralized Overlapping Control Design with Application to Rotary Drilling System," *IETE Journal of Research*, vol. 69, no. 4, pp. 2070–2079, May 2023, doi: 10.1080/03772063.2021.1886602.

[15] J. He and M. Luo, "Drilling Rig Hoisting Platform Security Monitoring System Design and Application," *Machines*, vol. 5, no. 3, p. 19, Aug. 2017, doi: 10.3390/machines5030019.

[16] C. Mendil, M. Kidouche, M. Z. Doghmane, S. Benammar, and K. F. Tee, "Rock–bit interaction effects on high-frequency stick-slip vibration severity in rotary drilling systems," *MMMS*, vol. 17, no. 5, pp. 1007–1023, Aug. 2021, doi: 10.1108/MMMS-10-2020-0256.

[17] M. Kidouche and A. Charef, "A CONSTRUCTIVE METHODOLOGY OF LYAPUNOV FUNCTION OF COMPOSITE SYSTEMS WITH NONLINEAR INTERCONNECTION TERM," *International Journal of Robotics and Automation*, vol. 21, no. 1, 2006, doi: 10.2316/Journal.206.2006.1.206-2814.

Nonsingular terminal sliding mode controller for controlling an microgrid's

Rayane Yasmine Mahfoudi
Electrical engineering department
LEER Laboratory Souk Ahras university
Souk Ahras, Algeria
r.mahfoudi@univ-soukahras.dz

Boumous Zouhir
Electrical engineering department
LEER Laboratory Souk Ahras university
Souk Ahras, Algeria
zohir.boumous@univ-soukahras.dz

Boumous Samira
Electrical engineering department
LEER Laboratory Souk Ahras university
Souk Ahras, Algeria
samira.boumous@univ-soukahras.dz

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

***Abstract*—The paper discusses a control system for a microgrid that uses renewable energy and energy storage, connecting them with voltage source inverters. A master-slave control strategy is used to manage data sharing through CAN and ZigBee networks. Challenges like external disturbances and delays affect voltage and current control, making a strong control method necessary. The study uses a binary iteration algorithm to set limits on system delays and employs a special controller to ensure reliability.**

The proposed control method was tested in MATLAB/Simulink and showed good results, keeping the microgrid stable and controlling voltage and current accurately, even with delays and other issues.

Keywords— renewable energy, master-slave control strategy, CAN/ZigBee.

I. INTRODUCTION

The global energy landscape is undergoing a significant transformation driven by the ever-increasing demand for energy and the persistent price instability of traditional energy sources. As populations grow and economies expand, the need for reliable, affordable, and sustainable energy solutions has become more pressing than ever. Conventional energy sources such as oil, coal, and natural gas, while historically dominant, are subject to volatile market dynamics, geopolitical tensions, and environmental concerns. These challenges have highlighted the critical need to diversify energy portfolios and embrace innovative alternatives.This growing urgency has spurred a global shift toward renewable energy sources, energy efficiency measures, and advancements in technology aimed at enhancing energy security and sustainability. Addressing these issues requires a holistic approach that balances economic, environmental, and social considerations, ensuring a resilient energy future for all.

NCS has become an impressive approach in the era of communication Gener ally, a control system which controls the process through a feedback loop connected with a data network is known as a NCS. NCS in MGs may cause communication delays, packet dropout, channel shading, and quantization errors. The advantages of NCSs with wireless communication compared to the wired type are flexibility, Electrical parameters of the batteries.

Various control strategies have been created to stabilize islanded microgrids, with hierarchical control being a key approach that includes power sharing techniques, which can operate with or without communication networks.Droop control is a widely used method that functions without communication, offering benefits like flexibility and reliability, but it has limitations such as dependency on voltage control and challenges with non-linear loads.Solutions to droop control's shortcomings include adding external loops to manage harmonic components, AC signal injection, and implementing harmonic droop coefficients with filtering strategies to improve performance.

II. SYSTEM INVESTIGATED

A. *Hybrid microgrid configuration*

The hybrid microgrid (HMMG) configuration consists of two interconnected distributed generation (DG) units, operating according to a master-slave strategy based on a communication network. Renewable energy sources include a photovoltaic (PV) system and energy storage systems (ESS), such as batteries and supercapacitors, connected to a DC link via a voltage source converter (VSC). Resistive-inductive filters are used to extract the fundamental frequency from the inverter output voltages. The DG units feed an RLC load via a resistive-inductive transmission line.

The master DG unit regulates the load voltage to a constant setpoint. Using Park transformation, the load voltage is converted into a synchronized rotating frame dq. The direct and quadratic components of the charge voltage are tested simultaneously against their respective references, and the resulting error is transmitted to the NTSM controller to generate the control signal. Space vector pulse width modulation (SVPWM) is used to produce the modulating signals transmitted to the inverter.
The DG slave unit operates identically to the master unit, regulating the direct and quadratic components of the supplied current. The quadratic component of the load voltage is assumed to be zero, enabling active and reactive power to be calculated. The slave unit has a power capacity equivalent to half that of the master unit. For adequate power sharing between DG units, a load flow division technique is used, assigning appropriate values for current references. These values are transmitted via the communication network, forming a networked control system. What's more, the number of slave DG units is expandable, and each unit can regulate its active and reactive power independently.

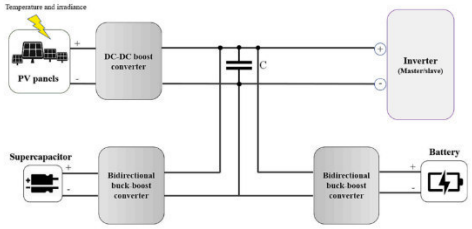


Fig 1: Hybrid system model

B. *Communication network configuration*

Networked control systems (NCS) have become an important approach to communication exchange in various processes, including microgrids (MGs). An NCS is a control system that uses a feedback loop connected to a data network. However, the use of NCSs in MGs can lead to problems such as communication delays, packet loss, channel interference and quantization errors, which can significantly reduce process performance or render control systems unstable. Although various challenges and solutions related to NCSs have been studied, wireless communication technologies offer advantages in terms of flexibility over wired systems.Over the past two decades, many researchers have explored two key technologies in this field: control area networks (CAN) and ZigBee.

III. CONTROL STRATEGY

The NTSMC (Non-Singular Terminal Sliding Mode Controller) is an advanced sliding mode control method, designed to ensure rapid convergence and avoid the singularity problems inherent in some traditional controllers.

A. *NTSMC for master DG unit*

he controller design for the master distributed generation (DG) unit aims to maintain the nominal value of the load voltage despite external disturbances and parametric uncertainties due to unmodeled loads. To achieve this, the direct and quadratic components of the load voltage, denoted V_d and V_q, must correspond to their respective reference values. Sliding variables are therefore defined for this control.

$$e_d = V_{d-ref} - V_d(t)$$

$$e_q = V_{q-ref} - V_q(t)$$

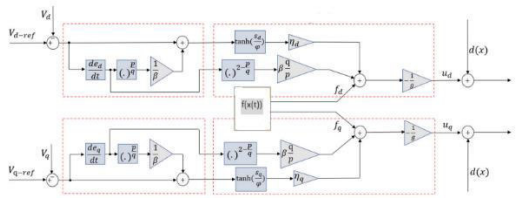


Fig 2: The NTSMC block diagram for the master DG unit.

The objective of the control system is to generate the control signal inside the system, stabilize the system, and ensure zero tracking errors. In other words, the control signal drives the master DG unit to adjust and maintain the load voltage near the reference value. This is to be accomplished despite the existence of external disturbances, unmodeled load fluctuation, and communication delays.

B. *NTSMC for slave DG unit*

The slave DG unit provides the required active and reactive powers by regulating its supplied currents, it,abcs. The reference currents, Ids− ref and Iqs− ref are defined by load flow division according to the capacity of the slave DG units and transfer through CAN/ZigBee networks. Sliding variables are therefore defined for this control.

$$e_{ds} = I_{ds-ref}(t - ţ) - I_{ds}(t)$$

$$e_{qs} = I_{qs-ref}(t - ţ) - I_{qs}(t)$$

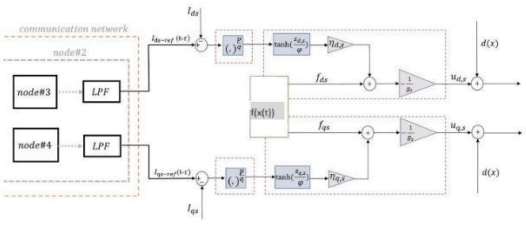


Fig 3: The NTSMC block diagram for the slave DG unit.

IV. RESULTS AND SIMULATIONS

This section presents the simulation results of the suggested control scheme utilizing Matlab/Simulink.

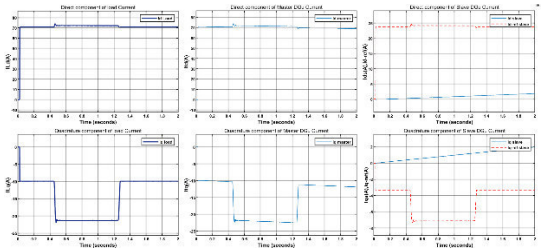


Fig 4: dq-Components of the voltages in the microgrid

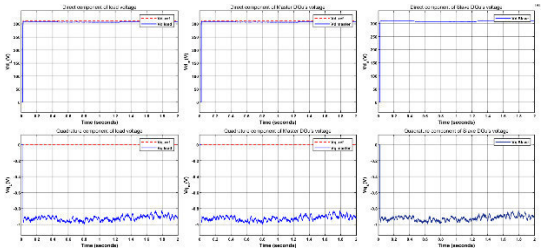


Fig 5: dq-Components of the current in the microgrid.

The dq-components of currents supplied by the slave DG unit immediately rise and follow their references, which have been obtained through the communication network, as shown in Fig 4 when the load increases. Because the master DG unit has double the capacity of the slave DG unit, it can automatically increase its currents to cover the remaining loads. As a result, there is minimal oscillation in the dq-components of load voltage at their reference values Fig 5.

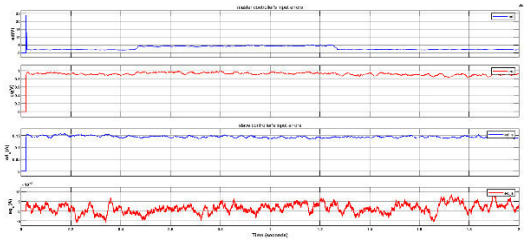


Fig 6: dq-Components of the error in the microgrid.

Unknown balanced load

A three-phase resistive-inductive load with unknown dynamics, that absorbs active and reactive power of 30 kW, 5 kvar, and a three-phase unknown resistive load with 10 kW are applied to the three phases at t = [1 2]s and unloaded on t = [4 3]s, respectively. As mentioned in the previous sections, the power-sharing block can send the reference values to the salve controller through the network connection. Therefore, passing the reference signal through its maximum boundary limitation (MADB) will affect the MG stability

the values of both DGs' active and reactive power injection in Fig 7 show a suitable power quality by NTSMC, along with communication issues.

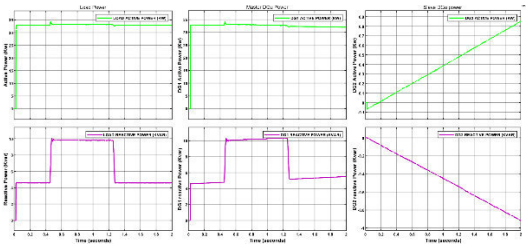


Fig 7: Active and reactive power sharing between DG units by adding unknown balanced loads.

.

CONCLUSION

In summary, this article presents a sophisticated control method that helps manage electricity in a small, independent power system. By using a master-slave communication strategy, the system can effectively share power among different energy sources while maintaining stable voltage and current levels. This is crucial for ensuring that all devices connected to the microgrid operate efficiently and reliably.

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a

REFERENCES

[1] M. Pirmohammad Talatape, AA. Afzalian,"A non-singular terminal sliding mode controller for a communication-based hybrid microgrid," Journal of Energy Storage, 2024.

[2] W. Zhou, Y. Wang, Y. Liang, "Sliding mode control for networked control systems," ISA Trans, 2022.

[3] Z. Man, Y. Feng, X. Yu, "Non-singular terminal sliding mode control of rigid manipulator," Automatica 38,2002.

[4] M.T. Rizi, H. Eliasi, "Nonsingular terminal sliding mode controller for voltage and current control of an islanded microgrid," Electr. Pow. Syst. Res, 2020.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

State space observer design for high frequency stick slip vibrations in drill-string of rotary drilling system

Bengana Chahrazad
Department of Automation and Process Electrification
Laboratory of Applied Automation (LAA)
Faculty of Hydrocarbons and Chemistry
Universityof M'hamed Bougara
Boumerdès, 35000, Algeria
c.bengana@univ-boumerdes.dz

Aibeche Abderezzak
Department of Automation and Process Electrification
Laboratory of Applied Automation (LAA)
Faculty of Hydrocarbons and Chemistry
Universityof M'hamed Bougara
Boumerdès, 35000, Algeria
aibeche_umbb@univ-boumerdes.dz

Mohamed Zinelabidine Doghmane
Departmentof Tank Assessment
Direction of Operations Exploration
Division of Exploration Sonatrach
Hassi Messaoud 30500, Ouergla, Algeria
Bob L. Herd
Department of Petroleum Engineering
Texas Tech University
Lubbock, 79401, Texas,USA
doghmane_m@yahoo.com

***Abstract*— Drilling process can be affected by several types of vibrations ,These vibrations significantly impede the reliability and efficiency of drilling operations, necessitating the need for effective solutions. The core problem tackled in the research is the identification, diagnosis, and reduction of high stick-slip-whirling vibrations that can be observed in rotating oil drilling systems. These vibrations induce wear and tear that causes drilling efficiency decreases, and safety risks appear. In this work, a three-degree-of-freedom model was created to illustrate the severity of the stick-slip phenomenon as consequence of torsional vibrations. The main objectif of this study is to design a state space observer by pole placement method for estimating unmeasured states, enhancing control performance, and improving system reliability. By providing insights into the internal state of a system, they facilitate advanced control strategies, fault detection, and optimization, even in systems with limited or noisy sensor data.**

Keywords—Rotary drilling process, Drill string, Stick slip vibrations, State space obserever, Torque on bit

I. INTRODUCTION

Drilling is a crucial procedure in industries like mining, construction, and oil & gas where accurate operations are necessary to assure effectiveness, safety, and optimal efficiency.

However, vibrations, which can come from a variety of sources and negatively impact the process, are one of the main difficulties addressed during drilling operations. Axial, lateral, and torsional vibrations are among these vibrations.

which are continuous rotational oscillations of the drill string. Each of these vibration types can negatively affect drilling performance, lead to equipment wear, and even cause damage to the wellbore or surrounding structures.

As such, diagnostic and monitoring drilling vibrations have become critical parts of current drilling operations. In order to identify, assess, and minimize the impacts of these vibrations in real time, vibration monitoring systems make use of sensors and data analytic tools. Early detection of severe vibration patterns allows operators to improve the drilling process, adjust drilling parameters, and avoid significant expenses or downtime, all of which contribute to the drilling operation's success and longevity.

In the last few years, many researches about the vibrations in rotary drilling systems have been conducted in the literature the research explores the potential of digital transformation and automation techniques in real-time monitoring and analysis. Through the integration of cutting-edge sensors, data acquisition systems, and intelligent algorithms, continuous monitoring of drilling system health becomes achievable, enabling the early detection of anomalous behavior and potential faults. The major outcomes of this research showcase the remarkable effectiveness of the developed diagnostic and monitoring approaches in identifying and mitigating severe stick-slip-whirling vibrations. The application of advanced signal processing techniques enables the early detection of anomalies and potential faults, empowering timely interventions and preventive maintenance measures.

Moreover, the integration of digital transformation and automation techniques facilitates real-time monitoring and adaptive control strategies, thereby enhancing drilling system performance, minimizing downtime, and bolstering operational safety . the aim of this study is to design a state-space observer which aimed at improving the monitoring, control, and performance of the system by estimating the internal states of the system that cannot be directly measured. Rotary drilling involves complex dynamics, including interactions between the drill bit, the drill string, the formation, and the rig equipment, which may not be fully observable or measurable. A state-space observer can estimate the dynamics of the drill bit and the surrounding environment, allowing the system to better understand and mitigate undesirable vibrations. By predicting and responding to vibration patterns, the observer helps in adjusting parameters like weight on the bit or rotational speed to reduce vibrations and improve drilling performance.

II. ROTARY DRILLING PROCESS

The drilling process is the set of operations allowing the digging of holes in the subsurface to arrive at the targeted reservoirs. Thus, the performance of hydrocarbon exploration and production depends on the drilling process systems, for which the most common in the oil industry is called the rotary drilling system [1]. The drilling rig is composed of several elements, all of them are important and each of them has a specific role so that the safety is ensured by blow out preventers (BOP)[2]. In the rotary drilling process, the drill string is rotated by a top drive system or a rotary table at a constant speed. However, the rotary speed at the lower end of the drill string that is at the bit fluctuates because of the oscillation that occurs due to low torsional stiffness of the drilling pipes. The magnitude of oscillation depends on three parameters:

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

- the properties of the driving system at the surface
- the friction that exists between the wellbore and drill string
- bit/rock interaction [3]

III. TORSIONAL VIBRATIONS SEVERITY

Torsional vibrations lead to irregular downhole rotations. Non-uniform rotation is developed when the bit becomes temporary stationary, causing the string to periodically torque up and then spin free. The severity of stick-slip will affect how long the bit stays stationary and consequently the rotational acceleration speed when the bit breaks free. The downhole RPM can become several times larger than the RPM applied at surface[4]. The highly nonlinear rock-bit interaction dynamic is the main cause of such type of vibrations [5] ,When stick-slip phenomenon occurs on the bit, stick and slip motions alternate. In this study, for the convenience of analysis, the drill bit is assumed to be in a critical state where the stick phase ends and the slip phase is incited. That is to say, the driven torque transmitted by the drillstring overcomes the frictional torque acted on the bit and thus the bit starts to rotate at the time t =0. According to the field drilling applications, we assume that the rotary table rotates clockwise with a constant angular velocity. Consequently, the drill bit rotates clockwise at the initial stage of the slip phase.[6]

Figure.1 represents the downhole measurement of rotational speed during stick slip vibration

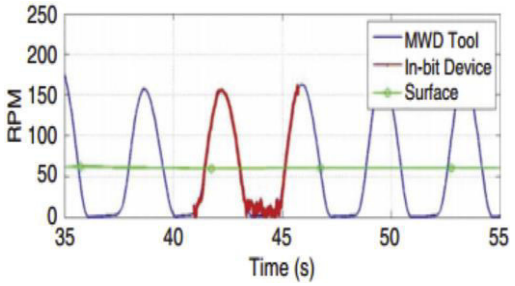


Fig .1.Downhole measurement of rotational speed during stick-slip [6]

IV. DRILL STRING MODEL

'Mass-spring-damper' approach or lumped pendulum vibration model indicates a mechanism that takes the design of a proxy system with masses coupled to springs (dampers can also be included). Several DOFs may be present in the resulting proxy system, dependent on the quantity of springs and masses employed. By dividing the system into two sections, this strategy enables us to generate a simplified and discrete model of the examined system, which makes it easier to solve the nonlinear equations of motion that will be obtained at the end of this section. A mathematical model's primary goal is to express and depict the different external forces acting on the dynamic system as equations [7] . The drill string, in this model, is considered as a torsion pendulum, It has three degrees of freedom 3DOF where the upper disc represents the rotary table, the middle disc is considered as the tool string, and the lower disc represents the drill bit[2] as shown in the figure 2.

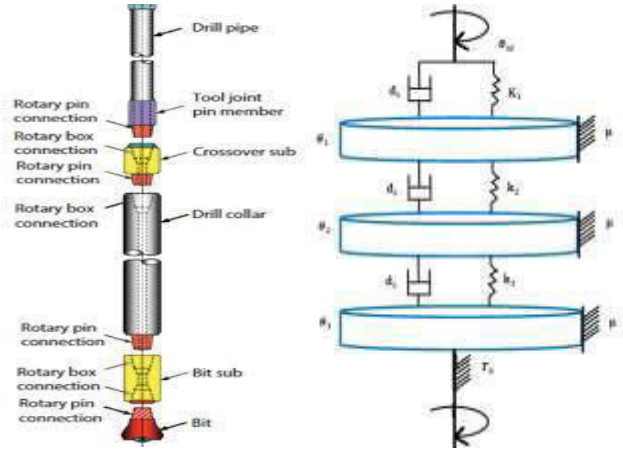


Fig .2. Drilling rig diagram and its equivalent model with 3DOF

The 3DOF proxy model is represented by figure.3

$$\begin{cases} \ddot{\theta}_1 = \frac{1}{j_1}(d_1(\dot{\theta}_{td} - \dot{\theta}_1) - k_1(\theta_{td} - \theta_1) - d_2(\dot{\theta}_1 - \dot{\theta}_2) - k_2(\theta_1 - \theta_2) - \mu\dot{\theta}_1) \\ \ddot{\theta}_2 = \frac{1}{j_2}(d_2(\dot{\theta}_1 - \dot{\theta}_2) + k_2(\theta_1 - \theta_2) - d_3(\dot{\theta}_2 - \dot{\theta}_3) - k_3(\theta_2 - \theta_3) - \mu\dot{\theta}_2) \\ \ddot{\theta}_3 = \frac{1}{j_3}(d_3(\dot{\theta}_2 - \dot{\theta}_3) + k_3(\theta_2 - \theta_3) - \mu\dot{\theta}_3 - T_b) \end{cases}$$

Fig .3. 3DOF model representation

The table shown in figure .4 represents Parameters description for the designed rotary drilling model

Parameter	Description	Unit
θ_{td}	The top drive angular displacement	[rad]
$\theta_{i=1,2,3}$	The angular displacements of the rope section i	[rad]
$k_{i=1,2,3}$	Torsional stiffness coefficient of the rope section i	[N.m/rad]
$d_{i=1,2,3}$	Internal damping coefficient of the rope section i	[N.m.S/rad]
μ	Wall friction coefficient	[N.m]
$j_{i=1,2,3}$	The inertia of the rope section i	[kg.m^2]

Fig .4. Parameters description for the designed rotary drilling model

Figure.5 represents the variable changes which were applied To use the standard notation for states and inputs in the mathematical model

$$\dot{\theta}_1 = x_1 \quad \theta_{td} - \theta_1 = x_4$$
$$\dot{\theta}_2 = x_2 \quad \theta_1 - \theta_2 = x_5 \quad \dot{\theta}_{td} = u$$
$$\dot{\theta}_3 = x_3 \quad \theta_2 - \theta_3 = x_6$$

Fig .5. Variable changes

By substituting the variable. The state equations of the system were then rewritten in the form given in the figure.6

$$\begin{cases} \dot{x}_1 = A_1x_1 + A_2x_2 + A_3x_3 + A_4x_5 + A_5u \\ \dot{x}_2 = B_1x_1 + B_2x_2 + B_3x_3 + B_4x_5 + B_5x_6 \\ \dot{x}_3 = C_1x_2 + C_2x_3 + C_3x_6 + C_4T_b \\ \dot{x}_4 = \dot{\theta}_{td} - \dot{\theta}_1 = u - x_1 \\ \dot{x}_5 = \dot{\theta}_1 - \dot{\theta}_2 = x_1 - x_2 \\ \dot{x}_6 = \dot{\theta}_2 - \dot{\theta}_3 = x_2 - x_3 \end{cases}$$

Fig .6. State equations after variable changes

Where the figure.7 represents the coefficients

$$\begin{cases} A_1=\frac{-d_1-d_2-\mu}{j_1} & B_1=\frac{d_2}{j_2} & C_1=\frac{d_3}{j_3} \\ A_2=\frac{d_2}{j_1} & B_2=\frac{-d_2-d_3-\mu}{j_2} & C_2=\frac{-d_3-\mu}{j_3} \\ A_3=\frac{-K_1}{j_1} & B_3=\frac{d_3}{j_2} & C_3=\frac{K_3}{j_3} \\ A_4=\frac{-K_2}{j_1} & B_4=\frac{K_2}{j_2} & C_4=\frac{-1}{j_3} \\ A_5=\frac{d_1}{j_1} & B_5=\frac{-K_3}{j_2} & \end{cases}$$

Fig .7. The coefficients variables

The state space matrix form is given by figure. 8

$$\dot{x}(t)=A(x(t))+B(x)\,u(t)$$
$$Y(t)=C(x(t))$$
$$x(t)=[\,x_1\;\;x_2\;\;x_3\;\;x_4\;\;x_5\;\;x_6\,]$$

Fig .8. State space form of the system

Y(t) is the measured output variable which is drill bit speed and x(t) is the state vector

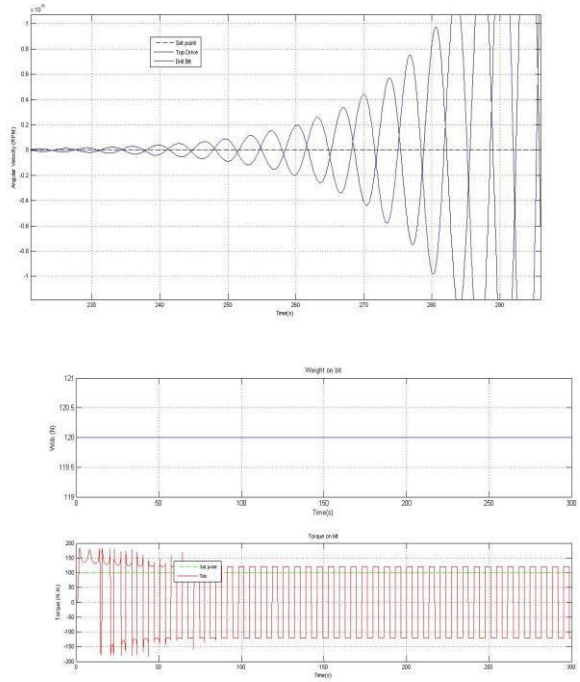


Fig .9. Drill string stick slip simulation

V. STATE SPACE OBSERVER DESIGN

A. *Basics of observers of linear dynamical systems in the state space form*

We consider the following linear state space model shown in figure.8 A, B, C are the system matrices.

Here, it should be emphasized that only the output vector Y of the system can be observed. However , in order to design a state feedback algorithm, we need to have full access and information about the state vector 'x' , this motivates us to develop an algorithm that will estimate the state vector from the observed output vectors Y, such an algorithm is called an observer in the control engineering literature.

Consequently, the main goal of an observer is to estimate the state of the system from the measurement of the output vector.

Also, it should be emphasized that when designing the observer we assume that the system matrices A, B, and C of the state space model are known.

The observer of this system is given by the following state equations in the figure .10

$$\dot{\hat{x}}=A\hat{x}+Bu+L\,(y-C\hat{x})$$

Fig .10. Observer state space

x 'hate' is the state of the observer and L is the observer gain matrix , here several important things need to be kept in mind , the inputs of the observer are: u 'the control input of the system', and Y ' the output of the system'.

Using the pole placement method in Matlab in the design of state space observer.

Our goal is to compute the matrix L of the state space observer, let us define the estimation error in figure.11

$$e(t)=x(t)-\hat{x}(t)$$
$$\dot{e}(t)=\dot{x}(t)-\dot{\hat{x}}(t)$$

Fig .11. The estimation error

By substituting the state equation of the dynamical system and the state space of the observer we obtain the closed loop observer dynamic equation shown in the figure.12

$$\dot{e}(t)=(A-LC)e(t)$$

Fig .12. closed loop observer dynamic equation

The goal of the observer design is to select the matrix L such that the closed loop system is asymptotically stable.

The closed loop stability of the system is achieved by using the pole placement method.

The idea is to design the matrix L such that the eigenvalues of the matrix A-LC are at prescribed locations.

Matlab has a function called 'place (.)' to perform pole placement, where the goal is to compute the feedback control matrix K such that the poles of the closed loop system 'A-BK', are at prescribed locations, however our observer design problems look a bit different , our closed loop matrix is 'A-LC' , we can still use the place function by observing the following, it is a very known fact that the matrix and its transpose have the same eingenvalues, this means that the transpose of the matrix 'A-LC' has the same structure as the matrix 'A-BK' . So we can compute the matrix L

B. *state space observer simulation*

To simulate the design observer, we need to transform the state space model of the observer in the appropriate form. The observer is given by the equation shown in the figure 10

This equation can be written as shown in the following figure.13

$$\dot{\hat{x}} = A\hat{x} + Bu + L\,(y - C\hat{x})$$

$$\dot{\hat{x}} = A\hat{x} + Bu + Ly - LC\hat{x}$$

$$\dot{\hat{x}} = (A - LC)\,\hat{x} + Bu + Ly$$

$$\dot{\hat{x}} = (A - LC)\,\hat{x} + [B \;\; L]\begin{bmatrix} u \\ y \end{bmatrix}$$

Fig .13. closed loop observer dynamic equation(2)

As we can see in this equation we can obtain the state space observer matrices

Aob=A-LC Bob=[B L] Cob=eye(6) D=[ ]

From the observer state space and its matrices we can simulate Xobserver as shown in the figure 14

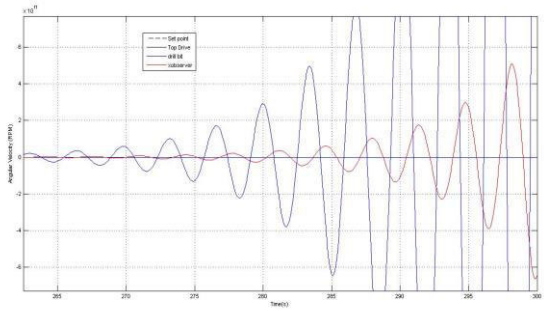


Fig .14. closed loop observer dynamic simulation

C. result and discussion

We can observe that the blue line is the state of the system and the red line represents the state of the observer, the state of the observer sometime approaches the state or we can say it asymptotically tracks the state of the system , this means that our observer is properly designed

VI. CONCLUSION

An efficient approach for optimizing monitoring and control in rotary drilling systems is the design of a state-space observer. The observer improves efficiency and safety by estimating unmeasured states, which helps the system react to changing operational situations in real time. However, the selections of model, observer gain design, noise rejection, and real-time performance need to be carefully considered because of the drilling system's basic complexity and nonlinearity. Better performance, increased resilience to unanticipated field difficulties, and more precise control are all possible with the appropriate observer design and implementation for a rotary drilling system. The simulation of a state-space observer provides valuable insights into the performance of the observer and its ability to estimate the states of a system from available measurements. Key aspects to evaluate include the convergence of the estimation error, the robustness to noise and disturbances, and how well the observer tracks the system's output. Tuning the observer's gain matrix and testing with various initial conditions and noise levels help optimize its design for real-world applications

REFERENCES.

[1] R. Riane, M. Z. Doghmane, M. Kidouche, K. F. Tee, et S. Djezzar, « Stick-Slip Vibration Suppression in Drill String Using Observer-Based LQG Controller », *Sensors*, vol. 22, nº 16, p. 5979, août 2022, doi: 10.3390/s22165979.

[2] C. Mendil, M. Kidouche, et M. Z. Doghmane, « Hybrid sliding PID controller for torsional vibrations mitigation in rotary drilling systems », *IJEECS*, vol. 22, nº 1, p. 146, avr. 2021, doi: 10.11591/ijeecs.v22.i1.pp146-158.

[3] A. Sharma, K. Abid, S. Srivastava, A. F. Baena Velasquez, et C. Teodoriu, « A review of torsional vibration mitigation techniques using active control and machine learning strategies », *Petroleum*, vol. 10, nº 3, p. 411-426, sept. 2024, doi: 10.1016/j.petlm.2023.09.007.

[4] L. Kyrvestad, « Tools and Techniques to Minimize Shock and Vibration to the Bottom Hole Assembly », 2014.

[5] Y. Xiao, C. Hurich, et S. D. Butt, « Assessment of rock-bit interaction and drilling performance using elastic waves propagated by the drilling system », *International Journal of Rock Mechanics and Mining Sciences*, vol. 105, p. 11-21, mai 2018, doi: 10.1016/j.ijrmms.2018.02.006.

[6] L. Tang, X. Zhu, C. Shi, J. Tang, et D. Xu, « Study of the influences of rotary table speed on stick-slip vibration of the drilling system », *Petroleum*, vol. 1, nº 4, p. 382-387, déc. 2015, doi: 10.1016/j.petlm.2015.10.004.

[7] I. Kessai, S. Benammar, M. Z. Doghmane, et K. F. Tee, « Drill Bit Deformations in Rotary Drilling Systems under Large-Amplitude Stick-Slip Vibrations », *Applied Sciences*, vol. 10, nº 18, p. 6523, sept. 2020, doi: 10.3390/app10186523

IDENTIFICATION OF HIGH-ORDER SYNCHRONOUS MACHINE PARAMETERS FROM STAND STILL FREQUENCY RESPONSES TESTING

Farid Leguebedj[1], Djamel Boukhetala[2]

[1]*Laboratoire de Recherche en Electrotechnique (LRE),*

[2] *Laboratoire de Commande des Processus (LCP)*

Ecole Nationale Polytechnique, 10 Rue des frères OUDEK, El-Harrach,16200,

Algiers, Algeria

e-mails: {farid.leguebedj/djamel.boukhetala@g.enp.edu.dz}

Abstract : This paper introduces a method based on standstill frequency response (SSFR) tests outlined in IEEE Std 115-A. The idea is to analyze the frequency response of direct axis and quadrature operational inductances, determine transfer function models based on time constants through curve-fitting techniques, and use Matlab's fsolve function to identify parameters of synchronous machine equivalent circuits. This approach is cost-effective, precise, and dependable.

Keywords : synchronous machine, equivalent circuit, parameters identification, SSFR tests.

I INTRODUCTION

The trend in aviation is moving toward "more electric aircraft," where electrical power is increasingly used to replace traditional hydraulic and pneumatic systems for tasks such as actuators, deicing, engine starts, and cabin air-conditioning. This shift is anticipated to raise the electrical power demand to over 1 MW. While standard aircraft generators typically operate at 120 kVA, innovative electrical system designs are expected to be developed [1]. Despite these advancements, the synchronous generator continues to be a key component of the system. Moreover, Park's model remains applicable, incorporating resistor-inductance branches to account for secondary and tertiary effects. The SSFR2 and SSFR3 models support the analysis of voltage regulation, fault analysis, disturbance response, exciter designs, and controller development [2]. Nevertheless, accurately determining circuit parameters remains a significant and persistent challenge.

Parameter identification is divided into offline and online approaches. Offline methods involve testing the machine when it is not in operation, such as through open circuit frequency response, DC excitation methods, and standstill frequency response testing [3]. These techniques are detailed in IEEE Std. 115-A [4] and are integral to the methodology used in this work. Accurately determining sub-transient reactances in both axes under saturation is essential for assessing the machine's electrical and mechanical robustness during a short circuit event [5]. Estimating the machine's flux linkages is challenging due to significant saturation, which introduces cross-saturation effects. Standstill frequency response modeling effectively derives high-order models by separating the d and q axis currents, enabling precise estimation because sinusoidal signals can uniformly excite various frequencies while resisting noise interference [6].

This approach derives time constants using curve-fitting techniques and identifies equivalent circuit parameters by solving non-linear equations. In the literature, these equations are often handled through numerical optimization, which poses significant numerical difficulties [7-9].

Online methods are gaining preference because they have minimal impact on machine operation [10-13], with time domain tests like small perturbation, standstill time domain, and load rejection tests being particularly useful. Among these, the standstill frequency response test is widely favored for its dependability, simplicity, and safety.

Online identification methods can be categorized into three main types: white-box, black-box, and grey-box modeling. White-box modeling is grounded in a comprehensive understanding of the system's structure and parameters, typically utilizing a theoretical model based on established physical principles. Examples include the extended Kalman filter [14], conjugate gradient method, and neural networks [15]. Black-box modeling, on the other hand, develops a model purely from input-output data, without any underlying physical knowledge. Grey-box modeling combines analytical equations with input-output data to determine model parameters. This approach demonstrates how to

derive circuit parameters from a given topology using Matlab software for educational purposes.

The paper is organized into six sections: Section II discusses the operational parameters that define the synchronous machine. Section III outlines the necessary tests as per IEEE Std. 115-A. The methodology for determining machine parameters is detailed in Section IV. Section V covers the presentation and validation of the results, followed by the conclusion in Section VI.

II OPERATIONAL QUANTITIES

Parameters can be expressed as transfer functions that illustrate the connection between the rotor and stator terminals when the machine is modeled as a two-port network, as depicted in Figure 1. This method allows the reactive behavior of the machine to be accurately represented while considering its circuits as a black box.

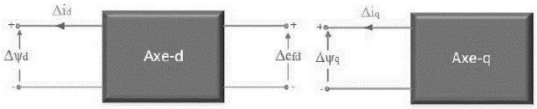


Fig. 1 D-axis quadrupole and q-axis dipole.

The measurable quantities at the stator and rotor are connected through equations that rely on operational parameters. These equations also include additional formulas to determine the resistances and reactances of the machine's windings. Here are the equations for stator flux:

$$\Delta\psi_d(s) = G(s)v_f(s) - L_d(s)i_d(s) \quad (1)$$

$$\Delta\psi_q(s) = -L_q(s)i_q(s) \quad (2)$$

Related to this, s refers to the Laplace operator. $G(s)$ represents the transfer function from the armature to the field. $L_d(s)$ indicates the operational inductance along the d-axis, while $L_d(s)$ denotes the operational inductance along the q-axis. The currents i_d and i_q correspond to the d-axis and q-axis, respectively. Additionally, V_f represents the voltage across the field winding.

III STAND STILL FREQUENCY RESPONSES TESTING

The Stand-Still Frequency Response (SSFR) test measures the frequency characteristics of stationary rotating machines. Recently, it has become a standard technique for evaluating the dynamic behavior of synchronous machines and gathering operational characteristics based on frequency. The SSFR method involves conducting individual tests for the direct and quadrature axes, which require specific rotor alignments with respect to the stator. Figure 2 illustrates the experimental procedure for the SSFR test, following the IEEE-Std115 guidelines for the d- and q-axis.

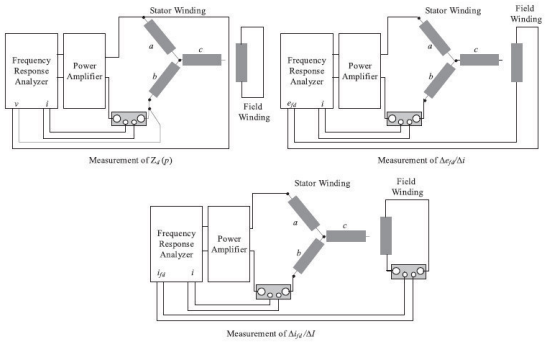


Fig. 2 Protocol of SSFR tests according to the IEEE standard 115 A

Calculating the parameters of a synchronous machine can be done using curve-fitting techniques that are straightforward and intuitive in Matlab.

The operational impedances of the d-axis and q-axis are as follows.

$$Z_d(s) = R_a + sL_d(s) \quad (3)$$

$$Z_q(s) = R_a + sL_q(s) \quad (4)$$

Both can be evaluated by implementing minor alterations to the armature voltage and current while simultaneously short-circuiting the field winding, positioning the rotor for d-axis and q-axis assessments, respectively.

$$Z_d(s) = -\frac{\Delta e_d(s)}{\Delta i_d(s)}\Big|_{\Delta e_{fd}=0} \quad (5)$$

$$Z_q(s) = -\frac{\Delta e_q(s)}{\Delta i_q(s)}\Big|_{\Delta e_{fd}} \quad (6)$$

To achieve accurate simulation of the transient behaviour of synchronous machines, additional measures must consider the d- and q-axis equivalent circuits. These parameters include the armature-to-field transfer function, $G(s)$, and the armature-to-field transfer impedance, $Z_{af0}(s)$. Evaluation of $G(s)$ can be conducted through the use of small variations in field and armature currents when the field winding is short-circuited.

$$sG(s) = \frac{\Delta i_{fd}(s)}{\Delta i_d(s)}\Big|_{\Delta e_{fd}=0} \quad (7)$$

The armature-to-field transfer impedance can be evaluated through minor fluctuations in field voltage

and armature current when the field winding is open-circuited.

$$Z_{af0}(s) = -\frac{\Delta e_{fd}(s)}{\Delta i_d(s)}\Big|_{\Delta i_{fd}=0} \quad (8)$$

To illustrate the Standstill Frequency Response methodology, a third-order d-axis equivalent circuit is employed. However, this approach is also applicable to other equivalent circuits as outlined in the IEEE Std 115-A standard.

IV METHODOLOGY

The proposed design methodology begins with the establishment of the topology of d-q axis circuits.

IV.1 Third-order model (SSFR3)

The synchronous machine is modelled with a field winding and two damper windings on the d-axis (see Figure 3) and three on the q-axis (see Figure 4). This configuration (2x3 model) is typically used for stability analysis.

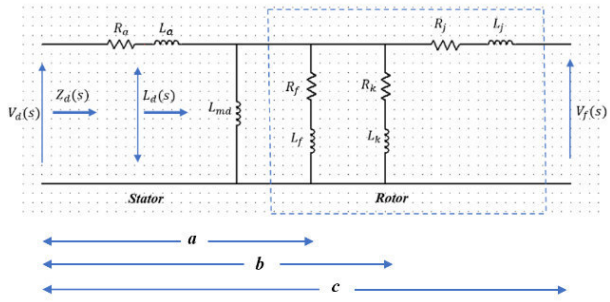


Fig. 3 Equivalent Circuit of a third order model for d-axis (SSFR3)

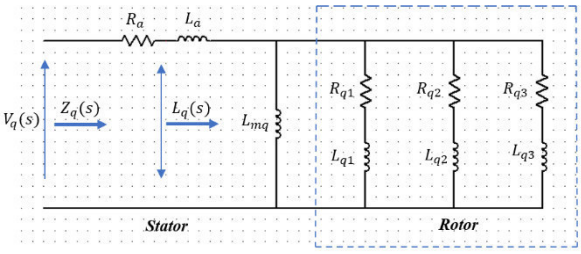


Fig. 4 Equivalent Circuit of a third order model for q-axis (SSFR3)

The equations pertaining to the operational inductance of the d-axis are presented below [16].

$$L_{md} = L_d - L_a \quad (9)$$

$$L_{amd} = \frac{L_a L_{md}}{L_a + L_{md}} \quad (10)$$

$$N_{um} = [1 + s\{(L_f + L_{amd})/R_f + (L_k + L_{amd})/R_k + (L_j + L_{amd})/R_j\} + s^2\left\{\frac{L_f L_j + L_j L_{amd} + L_f L_{amd}}{R_f R_j} + \frac{L_j L_k + L_k L_{amd} + L_j L_{amd}}{Rb Rk} + \frac{L_f L_k + L_f L_{amd} + L_k L_{amd})}{R_k R_f}\right\} + \frac{s^3(L_f L_k L_j + L_f L_k L_{amd} + L_k L_j L_{amd} + L_f L_j L_{amd})}{R_f R_k R_j}](L_a + L_{md}) \quad (11)$$

The denominator is structured in a manner that is analogous to that of the numerator, with the substitution of L_{amd} with L_{md} in the coefficients of s, s^2, and s^3.

$$D_{en} = [1 + s\{(L_f + L_{md})/R_f + (L_k + L_{md})/R_k + (L_j + L_{md})/R_j\} + s^2\left\{\frac{L_f L_j + L_j L_{md} + L_f L_{md}}{R_f R_j} + \frac{L_j L_k + L_k L_{md} + L_j L_{md}}{Rb Rk} + \frac{L_f L_k + L_f L_{md} + L_k L_{md})}{R_k R_f}\right\} + \frac{s^3(L_f L_k L_j + L_f L_k L_{md} + L_k L_j L_{md} + L_f L_j L_{md})}{R_f R_k R_j}] \quad (12)$$

Then,

$$L_d(s) = \frac{N_{um}}{D_{en}} \quad (13)$$

Thus, the operational inductance for a third-order model can be expressed as follows :

$$L_d(s) = L_d \frac{(1+sT'_d)(1+sT''_d)(1+sT'''_d)}{(1+sT'_{do})(1+sT''_{do})(1+sT'''_{do})} \quad (14)$$

Furthermore, equations (13) and (4) are also applicable to the q-axis.

IV.2 Experimental Tests on a High Power Generator

These tests are designed to identify the parameters of the generator, which will then be utilized in stability studies. The results of the SSFR for a 277.8 MVA power machine operating at 16.5 kV have been published in the EPRI [17]. Additionally, the d-axis impedance has been employed in determining the synchronous machine parameters.
Figure 5 illustrates the measurements of the impedance $Z_d(s)$.

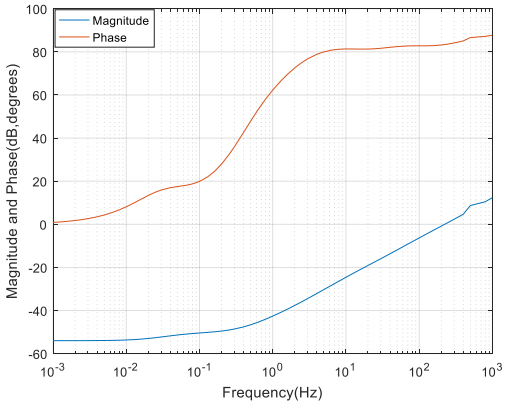


Fig. 5 Frequency response of the impedance $Z_d(s)$ (measured data)

IV.2 The operational inductance

To determine the frequency response of $L_d(s)$, we must consider the impedance measurements $Z_d(s)$ illustrated in Figure 5.
The expression for $L_d(s)$ is provided by Equation 15.

$$L_d(s) = \frac{Z_d(s) - R_a}{s} \quad (15)$$

The stator resistance R_a is defined as follows :

$$R_a = \lim_{s \to 0} |Z_d(s)| \quad (16)$$

By extrapolating the real component of the operational impedance at zero frequency, we can determine that the stator resistance value is $R_a = 0.00200\ \Omega$, as illustrated in Figure 6.

Equation (15) allows for the monitoring of operational inductance magnitude in dB and its phase in degrees, as illustrated in Figure 7. Based on the operational inductance $L_d(s)$ at zero frequency, we can extrapolate a synchronous inductance of $L_d = 0.004898$ H, as illustrated in Figure 8.

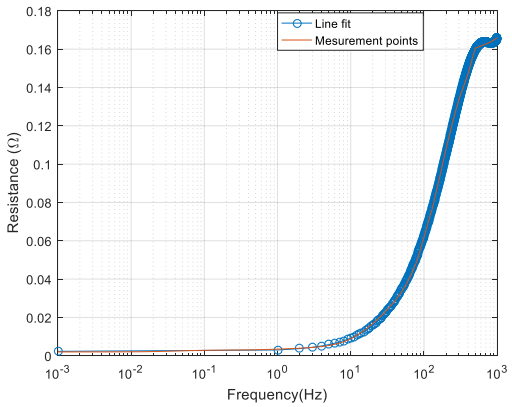


Fig. 6 Variation of the resistance versus frequency

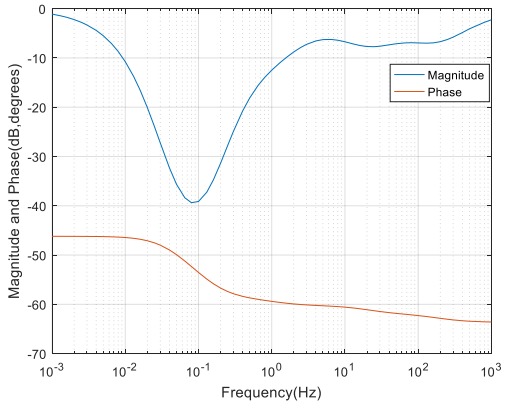


Fig. 7 Operational inductance frequency response: magnitude (dB) and phase (degrees)

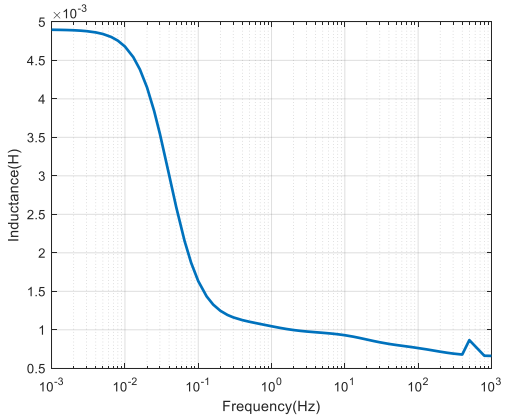


Fig. 8 Variation of the operational inductance (in H) versus frequency

IV.3 D-axis equations set

This section examines the relationship between the time constants of operational inductance and the equivalent circuit parameters of the third-order model (SSFR3). By expanding equation (14) and aligning it with equations (11) and (12), we derive a system of non-linear equations in which the time constants serve as inputs and the equivalent circuit parameters serve as outputs.

The system of equations is expressed as follows:

$$L_{amd} = \frac{L_a * L_{md}}{L_a + L_{md}} \quad (17)$$

$$T'_d + T''_d + T'''_d = \frac{L_f + L_{md}}{R_f} + \frac{L_k + L_{md}}{R_k} + \frac{L_j + L_{md}}{R_j} \quad (18)$$

$$T'_d * T''_d + T'_d * T'''_d + T''_d * T'''_d = \frac{L_f * L_k + L_f * L_{md} + L_j * L_{md}}{R_f * R_j} + \frac{L_k * L_j + L_k * L_{md} + L_j * L_{md}}{R_k * R_j} + \frac{L_f * L_k + L_f * L_{md} + L_k * L_{md}}{R_f * R_k} \quad (19)$$

$$T'_d * T''_d * T'''_d = \frac{L_f * L_k * L_j + L_k * L_j * L_{md} + L_f * L_k * L_{md} + L_f * L_j * L_{md}}{R_f * R_k * R_j} \quad (20)$$

$$T'_{d0} + T''_{d0} + T'''_{d0} = \frac{L_f + L_{amd}}{R_f} + \frac{L_k + L_{amd}}{R_k} + \frac{L_j + L_{amd}}{R_j} \quad (21)$$

$$T'_{d0} * T''_{d0} + T'_{d0} * T'''_{d0} + T''_{d0} * T'''_{d0} = \frac{L_f * L_k + L_f * L_{amd} + L_k * L_{amd}}{R_f * R_k} + \frac{L_k * L_j + L_k * L_{amd} + L_j * L_{amd}}{R_k * R_j} + \frac{L_f * L_j + L_f * L_{amd} + L_j * L_{amd}}{R_f * R_j} \quad (22)$$

$$T'_{d0} * T''_{d0} * T'''_{d0} = \frac{L_f * L_k * L_j + L_f * L_k * L_{amd} + L_f * L_j * L_{amd} + L_k * L_j * L_{amd}}{R_f * R_k * R_j} \quad (23)$$

In accordance with the recommendations set forth in the technical literature [18], the value of L_a, representing the leakage inductance, has been selected as 8% L_d.

IV.4 Identification of time constants using numerical curve fitting techniques

In order to calculate the optimal time constants from the frequency response data, numerical curve fitting techniques were employed, utilising both the magnitude and phase. The process is comprised of several steps :

- Define the model order.
- Perform curve fitting with initial parameter estimates.
- Propose a cost function for optimization.

By employing the Matlab functions “freqs” and “invfreqs,” we transformed the measured frequency response data into a transfer function for the third-order model (SSFR3), resulting in the following outcomes within the Matlab environment.

$$L_d(s) = 0.004899\frac{(1+0.896976s)(1+0.084855s)(1+0.002473s)}{(1+3.944719s)(1+0.101208s)(1+0.003354s)} \quad (24)$$

IV.5 Equivalent circuit parameter identification

Nonlinear equations from subsection IV.3 were simulated using Matlab's fsolve function, which employs a combination of Newton's method and trust-region-dogleg algorithms. Table 1 provides a summary of the parameters for the SSFR3 model's equivalent circuit.

Table. 1 Estimation of equivalent circuits parameters

parametrs	SSFR3 model
$Rf\,(\Omega)$	0.001692
Lf (H)	0.000859
$Rk\,(\Omega)$	0.017233
Lk (H)	0.000859
$Rj\,(\Omega)$	0.031553
Lj (H)	0.002174

V RESULTS AND VALIDATION

To validate the accuracy of the results, the estimated parameters of the SSFR3 model, as detailed in Table 1, are integrated into the transfer function outlined in Equation (13). The results are presented in Figures 9 and 10. A comparison of the frequency responses of the SSFR3 model with the measured data provides clear evidence of the effectiveness of the proposed approach.

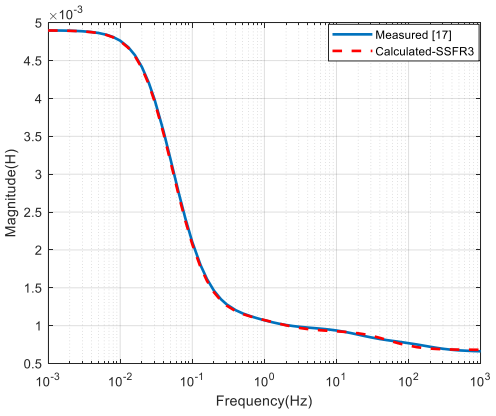


Fig. 9 Operational inductance magnitude versus frequency

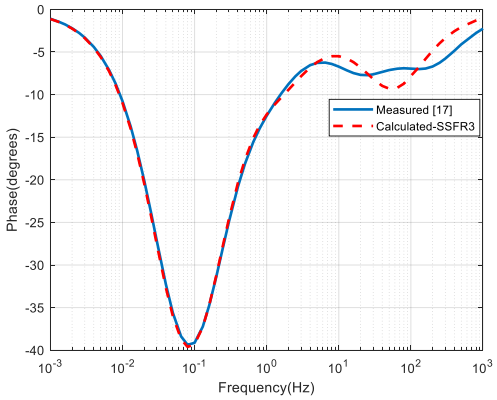


Fig. 10 Operational inductance phase versus frequency

VI CONCLUSION

This work presents a methodology for calculating the d-q equivalent circuit parameters of synchronous generators in accordance with the standards set forth in the IEEE Standard 115-A. Through this method, values are obtained that align well with those reported in experimental studies. This novel and educational approach is adaptable to any circuit order and custom topology, leveraging a general, reproducible, and comprehensible framework that facilitates rapid implementation.

REFERENCES

[1] A. Griffo, R. Wrobel, P. H. Mellor and J. M. Yon, “Design and Characterization of a Three-Phase Brushless Exciter for Aircraft Starter/Generator,” in *IEEE Transactions on Industry Applications*, vol. 49, no. 5, pp. 2106-2115, Sept.-Oct. 2013

[2] R. Wamkeue, C. Jolette, A. B. M. Mabwe, and I. Kamwa, “Cross identification of synchronous generator parameters from RTDR test timedomain analytical responses, ”*IEEE Trans. Energy Convers*.,vol. 26, no. 3, pp. 776–786, Sep. 2011.

[3] B. Zaker, G. B. Gharehpetian, M. Karrari and N. Moaddabi, “Simultaneous Parameter Identification of Synchronous Generator and Excitation System Using Online Measurements,” in *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1230-1238, May 2016.

[4] “IEEE guide for synchronous generator modeling practices and applications in power system stability analyses,” IEEE Std 1110-2002 Revis. IEEE Std 1110-1991, 1-72, 2003.

[5] F. Maurer, M. T. Xuan and J. J. Simond, “Two Full Parameter Identification Methods for Synchronous Machine Applying DC-Decay Tests for a Rotor in Arbitrary Position,” in *IEEE Transactions on Industry Applications*, vol. 53, no. 4, pp. 3505-3518, July-Aug. 2017.

[6] H. Radjeai, A. Barakat, S. Tnani and G. Champenois, “Identification of synchronous machine by Standstill Frequency Response (SSFR)

method - influence of the stator resistance -," *The XIX International Conference on Electrical Machines - ICEM 2010*, Rome, 2010, pp. 1-5.
[7] R. M. Saunders, "Synchronous-machine standstill frequency-response test data analysis," in *IEEE Transactions on Energy Conversion*, vol. 6, no. 3, pp. 564-571, Sep 1991.
[8] J. Bladh, M. Wallin, L. Saarinen and U. Lundin, "Standstill Frequency Response Test on a Synchronous Machine Extended With Damper Bar Measurements," in *IEEE Transactions on Energy Conversion*, vol. 31, no. 1, pp. 46-56, March 2016.
[9] E. C. Bortoni and J. A. Jardini, "A standstill frequency response method for large salient pole synchronous machines," in *IEEE Transactions on Energy Conversion*, vol. 19, no. 4, pp. 687-691, Dec. 2004.
[10] T. L. Vandoorn, F. M. De Belie, T. J. Vyncke, J. A. Melkebeek, and P. Lataire, "Generation of multi sinusoidal test signals for the identification of synchronous machine parameters by using a voltage-source inverter, "*IEEE Trans. Ind. Electron.*, vol. 57, no. 1, pp. 430–439, Jan. 2010.
[11] S. Pillutla and A. Keyhani, "Neural network observers for on-line tracking of synchronous generator parameters, "*IEEE Trans. Energy Conversion*, vol. 14, pp. 23–31, Mar. 1999
[12] Melgoza, R. Jesus, G. T. Heydt, and A. Keyhani, "An algebraic approach for identifying operating point dependent parameters of synchronous machine using orthogonal series expansions," *IEEE Trans. Energy Conversion*, vol. 16, pp. 92–98, Mar. 2001.
[13] H. B. Karayaka, A. Keyhani, G. T. Heydt, B. L. Agrawal and D. A. Selin, "Synchronous generator model identification and parameter estimation from operating data," in *IEEE Transactions on Energy Conversion*, vol. 18, no. 1, pp. 121-126, Mar 2003.
[14] B. Aubert, J. Régnier, S. Caux and D. Alejo, "Kalman-Filter-Based Indicator for Online Interturn Short Circuits Detection in Permanent-Magnet Synchronous Generators," in *IEEE Transactions on Industrial Electronic*s, vol. 62, no. 3, pp. 1921-1930, March 2015.
[15] M. Karrari and O. P. Malik, "Identification of physical parameters of a synchronous Generator from online measurements," in *IEEE Transactions on Energy Conversion*, vol. 19, no. 2, pp. 407-415, June 2004.
[16] F. Leguebedj, and D. Boukhetala, "Estimation of Synchronous Machine Parameters by Stand Still Frequency Responses Testing," *ENP Engineering Science Journal*, Vol. 3, No. 2, pp. 55-62, December, 2023.
[17] Electric Power Research Institute, "Compendium of the EPRI Workshop on determination of synchronous machine stability study constants," *by N. E. I Parsons*, Aug. 1980.
[18] A. Walton, "A systematic method for the determination of the parameters of synchronous machines from the results of frequency response tests," *IEEE Trans. Energy Convers.*, Vol. 15, No. 2, pp. 218–223, Jun. 2000.

FOC with LDIW-PSO for OEWIM Supplied by SVM Multi-Level Dual Inverter

Mohamed Lamine Ladjel
Department of Electrical Engineering,
University of Biskra
I3CUB Research Laboratory
Biskra, Algeria
mohamedlamine.ladjal@univ-biskra.dz

Abdelhakim Dendouga
Department of Electrical Engineering,
University of Biskra
I3CUB Research Laboratory
Biskra, Algeria
a.dendouga@univ-biskra.dz

***Abstract*— In order to improve the performance as well as efficiency of the variable speed control for Open-Ended Stator Winding Induction Motor (OEWIM) drives, the field-oriented control (FOC) technique is applied. The control performance is being improved in recent studies by employing optimization approaches as much as possible. The method that is most widely used is the particle swarm optimization algorithm. The PSO approach is used to determine the ideal values of the proportional-integral controller's parameters corresponding to the speed, flux, and currents controller in order to improve the speed response for OEWIM. This study presents a novel method to improve the variable speed control of FOC in OEWIM, based on PSO with a linearly decreasing inertia weight. The FOC is coupled with dual space vector modulation (SVM) to maintain a consistent switching frequency, reduced harmonic distortion and common mode voltage (CMV). In this case, two-level inverters are supplied to both ends of an open-end stator winding induction motor in order to perform a space vector modulation technique for a multi-level inverter. Through the use of simulated results from many tests of the entire system, the benefits of the proposed control system have been verified.**

Keywords*— *Space vector modulation, Four-level dual inverter, Open-end winding induction motor, Field-oriented control, PSO algorithm.

I. INTRODUCTION

In recent years, there's been an increase in studies aimed at creating new structures and control methods for multilevel inverters used in various industrial applications. These inverters are popular for medium and high-power variable speed machines due to their many advantages, including lower switch voltage stress, reduced dV/dt, and decreased voltage harmonic distortion [1]. Inverter control typically employs PWM techniques, which can be categorized as multi-carrier PWM (MCPWM), selective harmonic elimination (SHE), and space vector modulation (SVM). As MCPWM methods struggled to reduce output harmonics, researchers turned to SHE PWM to selectively remove harmonics [3]. However, SHE PWM has its drawbacks, such as slower reactions during transient events, limited switching angle range, and the need for pre-calculated values [2]. Multilevel PWM inverters face issues with common mode voltages, which affect output voltage quality and limit switching capabilities. CMV leads to motor shaft voltage, potentially causing bearing currents and damage when exceeding the grease's dielectric strength. Another concern is common mode leakage current flowing through grounded stator windings and frames, returning to the source via ground and causing electromagnetic interference. SVM is proposed as a solution, offering benefits like harmonic reduction, DC-link voltage balancing, and CMV reduction. Compared to sinusoidal PWM, SVM provides higher output voltage and more effective harmonic elimination [2].

The open-end winding (OEW) concept allows for the benefits of multi-level inverters using dual two level inverters on each end of the stator windings, creating a multi-level waveform [4]. The multi-level inverter setup for an OEWIM in Simulink testing demonstrates how two level inverters can achieve three level inversion [5]. OESW structures are gaining attention for their advantages over traditional star or delta connections, including lower switching losses, reduced common-mode voltage, and improved fault tolerance [4]. This paper focuses on controlling an OESWIM with two separate DC sources.

Various control strategies have been applied to OESWIMs, such as FOC [4], backstepping control (BSC) [6], direct torque control (DTC) [7], and sliding mode control (SMC) [8]. field-oriented control is popular due to its low complexity and stability, offering suitable dynamic performance. Within FOC, the proportional-integral (PI) controller is commonly used for its straightforward structure and low cost. It manages speed, flux, and currents to minimize overshoot and steady-state error as well as the speed oscillations. However, determining optimal PI parameters can be challenging, as they significantly impact control performance. Traditional methods like Ziegler-Nichols and Cohen-Coon often yield suboptimal results and are time-consuming. The particle swarm optimization (PSO) algorithm has emerged as a popular optimization technique, known for its robustness and ease of implementation [9].

The original PSO algorithm lacked an Inertia Weight, which was introduced by Shi and Eberhart in 1998. They found that high Inertia Weight supports global search, while low Inertia Weight enhances local search. Subsequent research explored dynamic inertia weight adjustment to improve PSO capabilities. Eberhart and Shi proposed a Random Inertia Weight approach, demonstrating improved early-stage convergence. A linearly decreasing inertia weight strategy has been shown to enhance PSO performance and effectiveness [10]. This study applies PSO to optimize OEWIM performance using IFOC by fine-tuning the PI controller settings for controlling speed, flux, and currents. Results indicate robust PI controller performance in reducing steady-state error, overshoot, settling time, and integral time absolute error (ITAE), even under unexpected changes in speed, mechanical load, and system parameters.

II. SPACE VECTOR MODULATION

A. *Two Level SVM*

In modern OESWIM control, SVM is a popular method for generating sinusoidal waveforms at inverter outputs. The core principle involves reconstructing the reference voltage vector V_{ref} using eight voltage vectors, each representing a

specific switching state in a three-phase voltage inverter. The reference voltage vector V_{ref} is analyzed over a modulation period T_m. This vector is identified within a sector using adjacent voltage vectors and zero vectors (V_0 and V_7). Both the reference vector and the eight voltage vectors are depicted on the αβ-plane (as shown in Fig. 1). The ON (1) or OFF (0) switch states are determined by the reference vector's position on this αβ-plane.

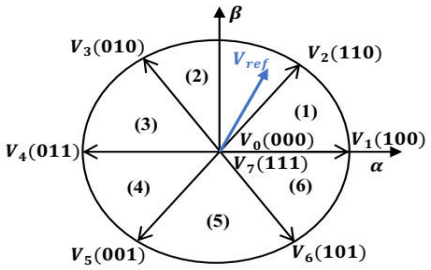


Fig. 1. Eight vectors of voltage in the αβ-plane.

The implementation of two-level SVM involves the following steps [9-13]:

- Determining of V_α, V_β, V_{ref}, and the rotation angle (θ).
- Calculating the application duration (x, y, and z) for neighboring vectors.
- Establishing the switching intervals for each switch (S_1, S_3, S_5, S_1', S_3', and S_5').

During a modulation period Tm, the reference vector V_{ref} can be estimated by generating an average vector. This is achieved by applying the inverter's adjacent non-zero vectors and zero vectors.

B. *Dual Inverter with asymmetric voltage sources*

A typical dual inverter configuration consists of two standard two-level output inverters, each powering one end of the IM terminals. As depicted in Fig. 2, the neutral point is electrically isolated from the squirrel-cage induction motor.

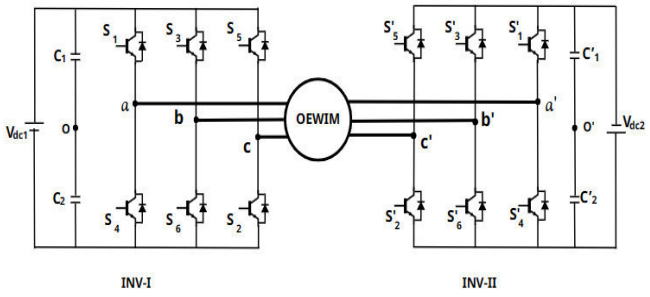


Fig. 2. OESWIM is fed by two inverters.

Two-level inverters are connected to both ends of the induction motor windings to create a multi-level inverter structure for an OESWIM [5]. Independent DC voltage sources provide electricity to each inverter. A space vector diagram for a dual two-level inverter system with symmetrical voltage sources is shown in Fig. 3. This inverter arrangement produces 48 voltage vectors, distributed uniformly and symmetrically across the α-β plane.

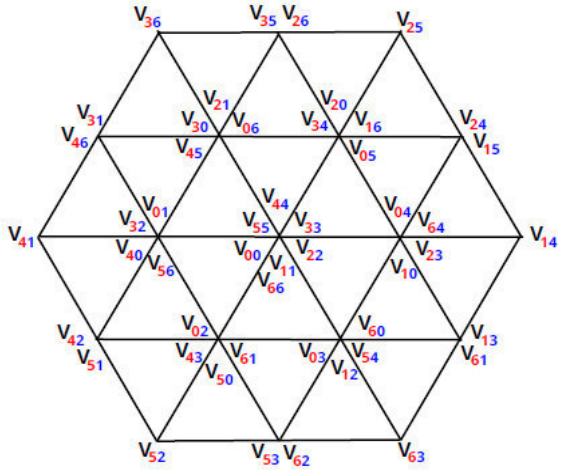


Fig. 3. Dual two-level inverter space vector diagram with asymmetric voltage sources.

C. *Common Mode Voltage*

Common mode voltage refers to the potential difference between the neutral point of a three-phase power supply and that of a three-phase load. In three-phase AC motors, the neutral point typically corresponds to the star point where the three phases intersect. For an open-end stator winding (OESW) configuration, the total output voltage across each motor winding is calculated as follows:

$$\begin{cases} V_{aa'} = V_{ao} - V_{a'o'} \\ V_{bb'} = V_{bo} - V_{b'o'} \\ V_{cc'} = V_{co} - V_{c'o'} \end{cases} \quad (1)$$

Where $V_{aa'}$, $V_{bb'}$, and $V_{cc'}$ are phase voltages, and V_{CMD} is the common mode voltage.

$$V_{CMV} = \frac{V_{aa'} + V_{bb'} + V_{cc'}}{3} \quad (2)$$

III. ROTOR FIELD-ORIENTED CONTROL

Field-Oriented Control (FOC) stands out as the leading drive method in industrial settings, especially for applications demanding high performance. When using the synchronous reference coordinate system, the dynamic equations for an Induction Motor (IM) can be expressed as [11]:

$$\frac{d}{dt}\begin{bmatrix} i_{ds} \\ i_{qs} \\ \varphi_{dr} \\ \varphi_{qr} \end{bmatrix} = \begin{bmatrix} -\left(\frac{R_s}{\sigma L_s} + \frac{1-\sigma}{\sigma\tau_s}\right) & w_e & \frac{1-\sigma}{\sigma L_m \tau_r} & \frac{(1-\sigma)w_e}{\sigma L_m} \\ -w_e & -\frac{R_s}{\sigma L_s} & -\frac{(1-\sigma)w_e}{\sigma L_m} & \frac{1-\sigma}{\sigma L_m \tau_r} \\ \frac{L_m}{\tau_r} & 0 & -\frac{1}{\tau_r} & w_{sl} \\ 0 & \frac{L_m}{\tau_r} & w_{sl} & -\frac{1}{\tau_r} \end{bmatrix} \begin{bmatrix} i_{ds} \\ i_{qs} \\ \varphi_{dr} \\ \varphi_{qr} \end{bmatrix} + \frac{1}{\sigma L_s}\begin{bmatrix} V_{ds} \\ V_{qs} \end{bmatrix} \quad (3)$$

In this context, V_{ds} and V_{qs} represent stator voltage, i_{ds} and i_{qs} denote stator current, while φ_{ds} and φ_{qs} indicate rotor flux linkage. R_s and R_r stand for stator and rotor resistance, respectively. L_s and L_r are the stator and rotor inductance, with L_m being the mutual inductance. *we* signifies the synchronous reference frame speed, and w_r is the rotor's electrical speed. The slip speed is defined as $w_{sl} = w_e - w_r$, and $\sigma = 1 - L_m^2 / L_r L_s$. The motor's mechanical equation is:

$$J_m \frac{dw_{rm}}{dt} = T_{em} - T_L - B_m w_{rm} \tag{4}$$

Here, J_m represents motor inertia, w_{rm} is the rotor-shaft's mechanical speed. T_{em} and T_L denotes the electromagnetic and loading torque, respectively. B_m represents the friction coefficient. For rotor-field oriented control, we set $\varphi_{qr} = 0$. The estimated slip speed can be calculated as:

$$w_{sl} = \frac{L_m i_{qs}}{\tau_r \varphi_{dr}} \tag{5}$$

the rotor time constant is represented by $\tau_r = R_r / L_r$. Rotor flux linkage is obtained through:

$$\varphi_{dr} = \frac{L_m}{\tau_r s + 1} i_{qs} \tag{6}$$

With S representing the Laplace operator. Under Field-Oriented Control conditions, the IM's electromagnetic torque is determined by:

$$T_e = \frac{3 n_p L_m}{2 L_r} \varphi_{dr} i_{qs} \tag{7}$$

Figure 4 illustrates the FOC technique's structure, highlighting the calculation block and control loops, all utilizing PI controllers. The terms V_{ed} and V_{eq} shown in Figure 4 are expressed as:

$$\begin{cases} V_{ed} = w_e \sigma L_s i_{qs} \\ V_{ed} = w_e (\sigma L_s i_{qs} + \frac{Lm}{Lr} \varphi_{dr}) \end{cases} \tag{8}$$

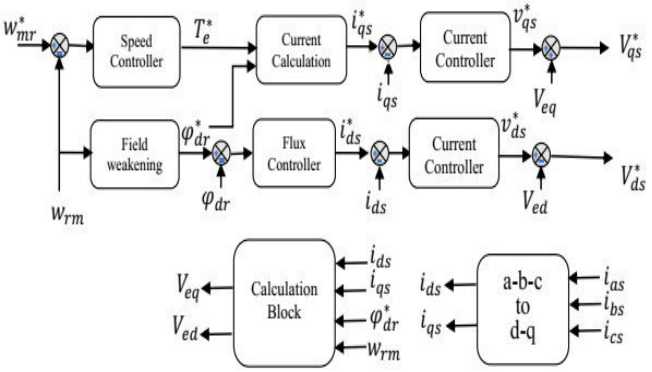


Fig. 4. FOC technique scheme

A. PI Controller Design

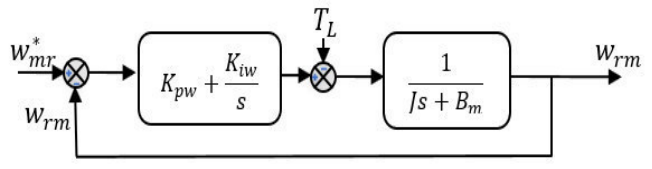


Fig. 5. The PI speed controller diagram.

The open-loop transfer function can be derived from Fig. 5, as indicated in reference [12-15].

$$F_{OL} = \frac{K_{pw}(s + \frac{K_{iw}}{K_{pw}})}{s} \frac{1}{J(s + \frac{B_m}{J})} \tag{9}$$

The parameters for the PI controller can be calculated to counteract the poles of equation (9).

$$\begin{cases} K_{pw} = \frac{3J}{T_{rw}} \\ K_{iw} = \frac{3B_m}{T_{rw}} \end{cases} \tag{10}$$

A similar approach is used to determine the parameters for both the flux and current controllers.

IV. PSO-PI Controller Design

PSO, introduced by Kennedy and Eberhart [9], draws inspiration from the collective behavior of bird flocks. This optimization method has gained popularity among researchers due to its global exploration capabilities, proven reliability, and ease of implementation across various applications. In the PSO algorithm, particles explore the search space using two reference points: the local best position identified by the swarm in the current iteration, and the global best position determined from all previous iterations. The conventional PSO algorithm is prone to rapid convergence to local solutions. To address this, several modified versions have been proposed, including the inertia weight approach, the constriction factors technique, and the tracking dynamic system method. This study employs the linearly decreasing strategy for inertia weight, enabling PSO to swiftly locate the optimal solution initially while ensuring effective convergence towards the end [16]. The following are the equations for updating location, velocity, and weight due to inertia:

$$W_i = W_{\max} - \frac{W_{\max} - W_{\min}}{Itr} * i \tag{11}$$

$$\begin{aligned} V_i(k+1) = W_i * V_i(k) + C * Rand * (P_{best} - x_i) \\ + C * Rand * (G_{best} - x_i) \end{aligned} \tag{12}$$

$$x_i(k+1) = x_i(k) + V_i(k+1) \tag{13}$$

In these equations, Wi represents the inertia weight factor, $Vi(k)$ and $xi(k)$ denote the particle's velocity and position respectively, Pbest and Gbest signify the best solutions for individual particles and the entire swarm, C is learning constant, and Rand is a random variable uniformly distributed between 0 and 1. The PI controller is a preferred choice for OEWIM control due to its simple structure, straightforward design process, and cost-effective hardware implementation. However, it can be challenging to figure out the proper values for the coefficients (ki, kp) of the PI controller, though. The PSO approach is used to determine the optimal values for the PI parameters in FOC. ITAE is used as an objective function to minimize the speed response error. ITAE is defined as:

$$ITAE = \int_0^t t |error| \, dt \tag{14}$$

where $error = w_{mref} - w_{rm}$ represents the difference between the reference and actual speed. Figure 6 illustrates the pseudo-code for the enhanced PSO technique used to optimize

the proposed control system, with P representing the number of particles in the swarm.

- Initialization

set the values of: C_1, C_2, W_{min}, W_{max}, Itr, P;
Generate random positions and velocities for each particle for starting;
Evaluate them using the objective function;
Select the best position;
Update P_{best}, G_{best};

- Search loop

Update W_i by (11);
Calculate $x_i(k+1)$ and $V_i(k+1)$ for each particle by (12) and (13), respectively;
Check the boundary control mechanism for each new position;
Evaluate it using the objective function;
Check P_{best}, G_{best} and update them if needed;
Repeat it for all iterations;

- Termination

Terminate the algorithm if the stopping criteria are achieved;
Present G_{best};

Fig. 6. The modified PSO algorithm's structure.

V. Results and Discussion

The effectiveness of the proposed optimization control for the OEWIM with a multi-level dual inverter is evaluated using the MATLAB Simulink platform. To assess the robustness and examine the dynamic performance under various operating conditions, numerical simulations compare the PSO-optimized PI controller to a conventional PI controller in the FOC for the OEWIM (Fig. 7). The simulations use the following drive system parameters: Rotor flux reference: 0.9 Wb; Motor specifications: P=1.5kW, Rs=4.85Ω, Rr=3.8Ω, Ls=274mH, Lr=274mH, Lm=258mH, J=0.031Kg.m2, Bm=0.00114 N.m.s, 1 pole pair; Inverter switching frequency: 1 kHz; Sampling time: 1 microsecond.

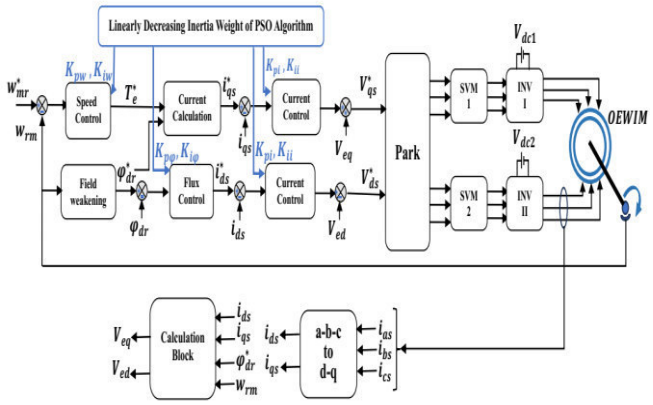


Fig. 7. an illustration of the complete driving mechanism.

In order to achieve optimal performance and reduce ITAE for rapid reaction, the suggested controller runs 20 iterations (Fig. 10).

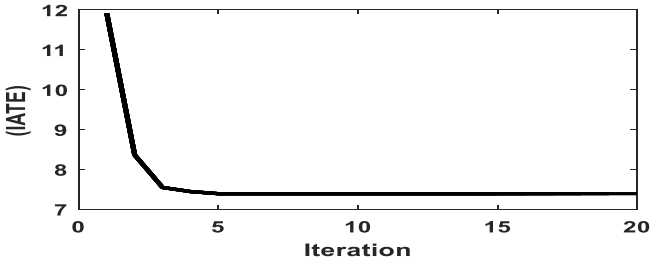


Fig. 8. The relation between iterations and ITAE..

Fig. 9 shows the multi-level phase voltages obtained through the SVM-controlled dual inverter.

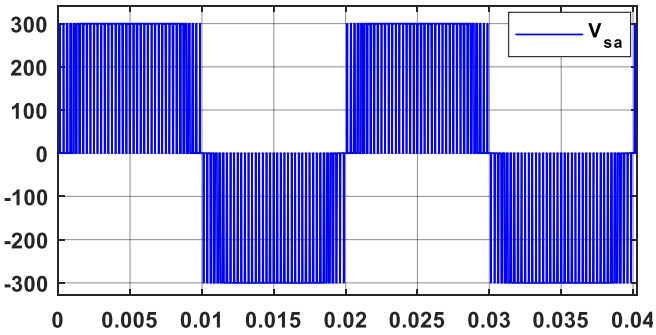


Fig. 9. The phase voltages for three levels.

Figure 4 depicts the CMV, maintained at Vdc/6=100V, which outperforms most PWM techniques with five and six-level inverters mentioned in [1]..

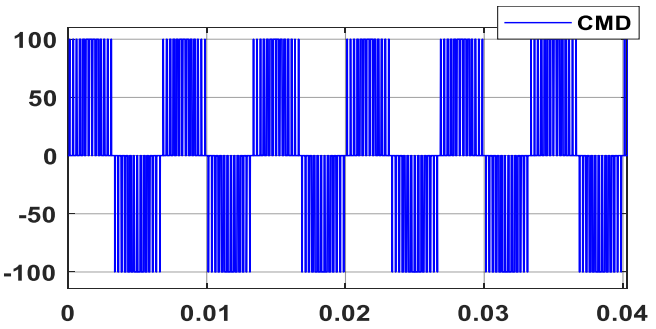


Fig. 10. The CMV.

A. *Comparative test*

The test follows these operational sequences:

- Motor starts unloaded.
- Load torque applied at 0.5s.
- Load torque inverted at 1s (demonstrating bidirectional converter capability).
- Motor unloaded again at 1.5s.
- Reference speed inverted after 3s.

Figure 11 clearly demonstrates the FOC's excellent performance in terms of decoupling, reference trajectory tracking, and rapid dynamic response. The comparative test results are as follows:

1) Conventional PI controller:

- Poor speed reference tracking.
- Long response time with significant overshoot.
- Substantial static errors during step load torque application (0.5-1.5s) and reference speed reversal.

2) PI controller with PSO algorithm:

- Perfect speed tracking without overshoot.
- Improved response time.

These results highlight the superior performance of the PSO-optimized PI controller compared to the conventional PI controller in the given OEWIM control scenario.

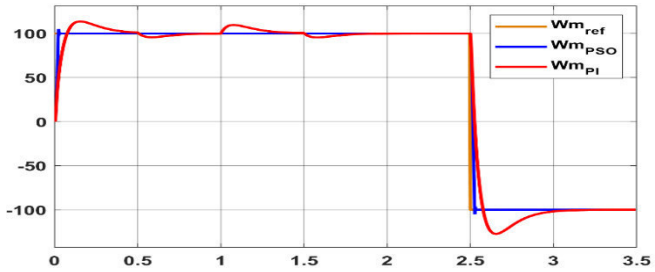


Fig. 11. The PSO-PI and conventional PI's rotating speeds and its references.

Figure 12 illustrates the electromagnetic torque and load torque disturbance. The optimal controller demonstrates rapid and precise torque dynamics during load changes, with acceptable torque ripples. This contrasts sharply with the traditional PI controller's performance. Additionally, a negative torque is generated on the rotor shaft.

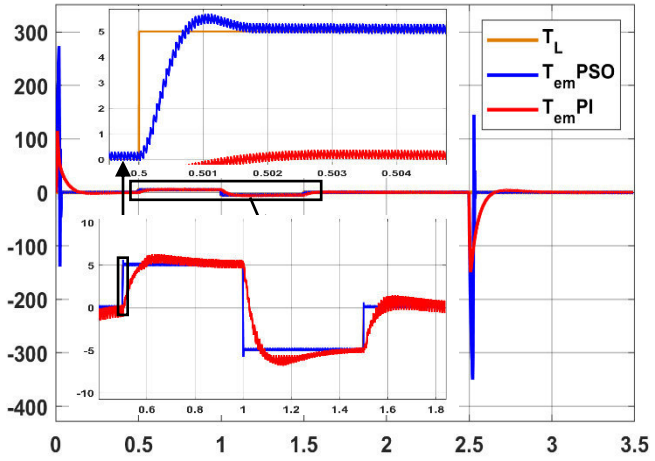


Fig. 12. The PI and PSO-PI electromagnetic torque under applied load.

Figure 13 shows the rotor flux components in the d-q rotating frame. The d-axis rotor flux maintains stability at the reference value (0.9Wb), while the q-axis flux remains near zero. This indicates preserved decoupling between rotor flux and electromagnetic torque. During speed reversal at 2.5s, the conventional PI experiences significant perturbation in the rotor flux (Phrd) curve, leading to decoupling loss. However, the PSO technique maintains perfect decoupling during speed reversal.

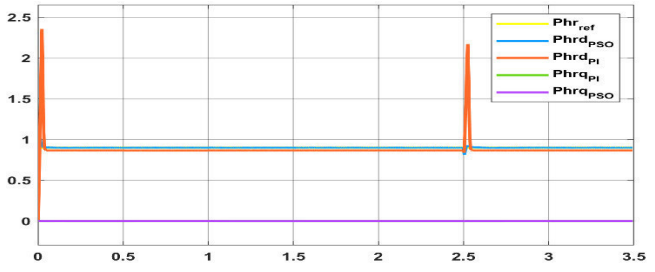
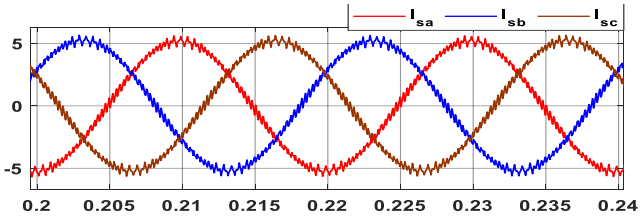
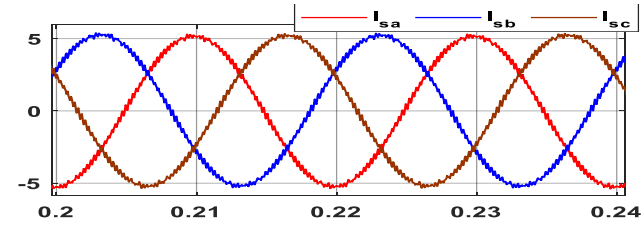


Fig. 13. The flux elements for the d-q rotor within a rotating frame.

Figure 14 compares the phase stator current for both controllers. The PSO-optimized PI controller (Fig. 14 (b)) produces a more regulated sinusoidal waveform with less chattering compared to the conventional PI (Fig. 14 (a)).



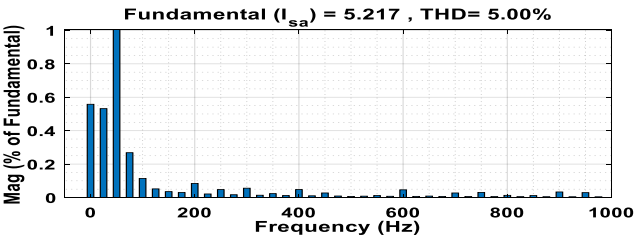
(a)



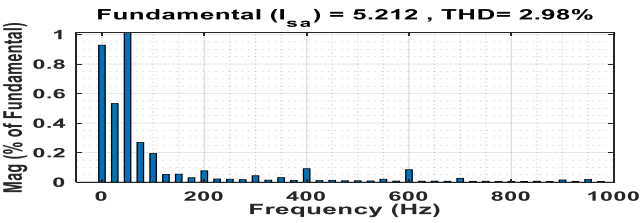
(b)

Fig. 14. Phase stator current: (a) traditional PI, (b) PI controllers using PSO.

Figure 15 displays the harmonic spectrums. The Total Harmonic Distortion (THD) of the current with the PSO algorithm is 2.98%, significantly lower than the 5% THD obtained with the conventional PI.



(a)



(b)

Fig. 15. Harmonic spectrums of: (a) the traditional PI, (b) PI with PSO

B. Robustness test

The drive system is tested under a fixed load torque at a low speed of 10 rad/s. Initially, the stator resistance Rs is 4.85 Ω. At 1.5s, Rs is increased to 7.25 Ω, while Rr is simultaneously changed from 3.805 Ω to 5.7075 Ω. This test evaluates the impact of motor parameter variations on the PSO controller with linearly decreasing inertia weight method. Figure 16 shows that even at lower speeds, the rotational speed closely follows its reference, demonstrating the control technique's excellent performance at low operating speeds. When actual motor parameters differ from those used in the controllers, the suggested control initially becomes unstable. However, this instability is quickly resolved, and the rotational speed closely tracks the reference speed. Figure 17 demonstrates the electromagnetic torque response under parameter variations. When the stator and rotor resistances increase at t=1.5s, the electromagnetic torque shows

momentary oscillation but quickly stabilizes to match the load torque demand. This rapid recovery demonstrates the robustness of the PSO-optimized controller under significant parameter variations. The sustained tracking performance under these conditions validates the controller's ability to maintain accurate torque control even with 50% parameter variations.

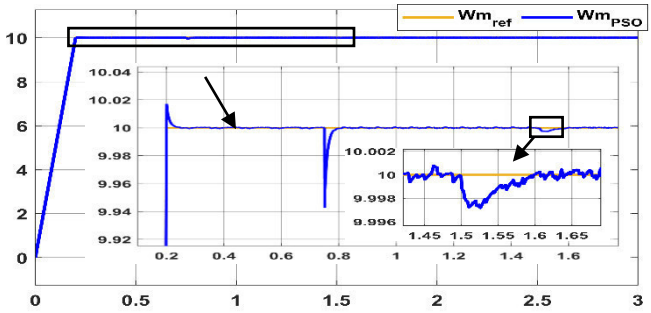


Fig. 16. Rotational and Reference speeds.

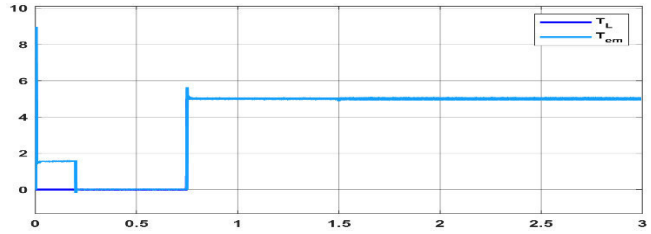


Fig. 17. Electromagnetic torque developed concerning load torque.

Figure 18 demonstrates that the three-phase stator currents maintain sinusoidal waveforms even at low speeds (below 10 rad/s). The stator current amplitudes change when Rs and Rr vary. These results confirm that the optimal FOC technique based on PSO-optimized provides improved performance and high accuracy even with Rs and Rr variations, validating the effectiveness of the proposed control method.

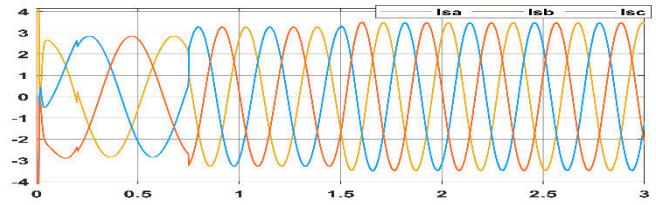


Fig. 18. The stator's three-phase current.

VI. Conclusion

This research examined the effectiveness of a proposed rotor field-oriented control system for an Open-End Winding Induction Motor (OESWIM), utilizing a Linearly Decreasing inertia weight strategy of particle swarm optimization (LDIW-PSO) Algorithm. The motor was powered by a Four-Level Dual Inverter controlled through Space Vector Modulation (SVM). The control scheme was designed to dynamically adjust the parameters of four PI controllers (managing speed, flux, and d-q axis currents). By employing PI controllers, the drive system demonstrated stability, reliability, and robustness under various external disturbances, including changes in load torque, reference speed, and parameter fluctuations. Implementing the SVM technique enabled the generation of near-perfect sinusoidal input currents and facilitated efficient operation of the OESWIM across all four quadrants. This control strategy is suitable for applications requiring high dynamic performance, offering reduced Total Harmonic Distortion (THD) and mitigated Common Mode Distortion (CMD). After evaluating these findings, it is thought that there is potential for this study to continue with hybrid optimization techniques combining PSO with other meta-heuristic algorithms. A key advantage of the proposed control method is its low computational complexity, making it implementable on cost-effective microcontrollers. These characteristics make it an attractive option for industrial applications, particularly those demanding high performance and efficiency.

References

[1] M. Ranjit, V. RameshBabu, S. R. Jalluri, Performance Analysis of Multicarrier PWM Techniques for Dual Inverter Fed Open End Winding Induction Motor Drive, in: 2nd International Conference on Power and Embedded Drive Control (ICPEDC), 2019, pp.18-23.

[2] V. Jayakumar, B. Chokkalingam, and J. L. Munda, A comprehensive review on space vector modulation techniques for neutral point clamped multi-level inverters. IEEE Access, vol.9 (2021) pp.112104-112144.

[3] A. El-Hosainy, H. A. Hamed, H. Z. Azazi, and E. E. El-Kholy, A review of multilevel inverter topologies, control techniques, and applications,'' In Nineteenth International Middle East Power Systems Conference (MEPCON), 2017, pp.1265-1275.

[4] S. Khadar, H. Abu-Rub, and A. Kouzou, Sensorless field-oriented control for open-end winding five-phase induction motor with parameters estimation. IEEE Open J. Ind. Electron. Soc, vol.2 (2021) pp.266-279.

[5] B. R. Vinod, G. Shiny, Direct Torque Control Scheme for a Four-Level-Inverter Fed Open-End-Winding Induction Motor. IEEE Trans. Energy Convers, vol.34 N°4(2019) pp.2209-2217.

[6] K. Saad, K. Abdellah, H. Ahmed, and A. Iqbal, Investigation on SVM-Backstepping sensorless control of five-phase open-end winding induction motor based on model reference adaptive system and parameter estimation Engineering Science and Technology, Eng. Sci. Technol. Int. J, vol. 22 N°4(2019), pp.1013-1026.

[7] K. V. Praveen Kumar and T. V. Kumar, Experimental implementation of direct torque control of open end winding induction motor, in: IEEE Region 10 Conference (TENCON), Singapore, 2016, pp.3318-3323

[8] S. Khadar. A. Kouzou. M. M Rezaoui, and A. Hafaifa, Fault Tolerant Sensorless Sliding Mode Control by Parameters Estimation of an Open-End Winding Five-Phase Induction Motor, Modelling, Meas. Control, vol.92 N°(2-4)(2019), pp.6–15.

[9] Ali, Jamal & Hannan, M. A. & Mohamed, Azah. (2016). Improved indirect field-oriented control of induction motor drive based PSO algorithm. Jurnal Teknologi. 78. 10.11113/jt.v78.8888.

[10] M. L. Ladjel and A. Dendouga, "Incorporation of Nonlinear Control and SOSM Controller with Improved PSO Algorithm for OEWIM Fed by SVM Seven-Level Dual Inverter". Arab J Sci Eng (2024).

[11] A. Bendaikha, S. Saad, Comparative Study of Five-Level and Seven-Level Inverter Controlled by Space Vector Pulse Width Modulation. International Journal of Power Electronics and Drive Systems, vol.8 N°2(2017), pp.755.

[12] A. DENDOUGA and A. DENDOUGA, "High Performance Vector Control of Permanent Magnet Synchronous Motor Fed by Direct SVM Matrix Converter," 2019 4th International Conference on Power Electronics and their Applications (ICPEA), Elazig, Turkey, 2019, pp. 1-6

[13] A. DENDOUGA, "A Comparative Study Between the PI and SM Controllers Used by Nonlinear Control of Induction Motor Fed by SVM Matrix Converter", IETE Research Journal, Vol.63 N°4, pp. 3019-3029, 2020.

[14] A. DENDOUGA, Feedback linearization associated to a sliding mode controller for an operating with unity power factor of induction motor fed by matrix converter. SN Appl. Sci. 2, 879 (2020).

[15] Y. C. Luo, Z. S. Ke, and Y. P. Kuo, Sensorless Rotor-Field Oriented Controlled Induction Motor Drive with Particle Swarm Optimization Algorithm Speed Controller Design Strategy. Math. Probl. Eng, vol.2014 (2014).

[16] M. L. Ladjel and A. Dendouga, "Field Oriented Control with Improved PSO Algorithm for Open End Winding Induction Motor Fed by SVM Four Level Dual Inverter," in Proc. 2nd Int. Conf. Electronics, Energy and Measurement (IC2EM), Nov. 2023, pp. 1-8.

Neural Network Modeling of Manipulator Robot

Chafea Stiti
Laboratoire des systèmes électriques
et télécommande,
Blida1 University
Blida, Algeria
stiti_chafea@univ-blida.dz

Mohamed Benrabah
Laboratoire de robotique Parallélisme
et systemes embarqués
USTHB University
Alger, Algeria
mbenrabah@usthb.dz

Abdelhadi Aouaichia
Laboratoire des systèmes électriques
et télécommande,
Blida1 University
Blida, Algeria
aouaichia.abdelhadi@gmail.com

Kamel Kara
Laboratoire des systèmes électriques
et télécommande,
Blida1 University
Blida, Algeria
km_kara@yahoo.fr

***Abstract*— This paper proposes a neural network modeling technique based on the use of two multilayer perceptron working in both. The suggested model is evaluated by modeling a two degree of freedom manipulator robot. To demonstrate the level of precision of this modeling technique, the values of determination coefficient and root mean square error criteria are computed and analyzed to evaluate the obtained neural network model.**

Keywords— MPC, Neural Networks, Multilayer Perceptron, Manipulator Robot

I. INTRODUCTION

The employment of manipulator robots has grown significantly in the last few years across a range of fields, including the military, industry, and medicine. The ability of the manipulators to complete laborious, precise tasks explains this improvement. Because manipulator robots can execute precise operations, they are especially well-suited for applications that need precision trajectory monitoring, such as welding specific parts or painting specific areas while following preset trajectories [1]. The coupled and complex dynamics of manipulator robots with substantial nonlinearities make it challenging to develop control algorithms to achieve high-precision tracking capabilities [2].In response to this challenge, researchers have developed sophisticated control algorithms to guarantee accurate trajectory tracking. These methods include Funnel Control (FC) [3], Model Predictive Control (MPC) [4] and Non-linear Model Predictive Control (NMPC) [5].

The efficiency of the NMPC control technique is generally due to the choice of a model that accurately represents the system to be controlled [16]. Neural networks are a modelling method that has enjoyed great success thanks to their ability to accurately represent complex dynamics (non-linearities and uncertainties) [6] [7], for this issue it's very adopted method in the robotic domain.

Among the many noteworthy benefits of neural network modeling of manipulator robots are its accuracy in simulating nonlinear dynamics, as well as its versatility and effectiveness in handling uncertainties and disturbances [6]. These benefits help neural networks perform better and be more dependable when modeling and controlling manipulator robots. Numerous neural network types, including Multilayer Perceptrons (MLP) [8], Convolutional Neural Networks (CNN) [9], Recurrent Neural Networks (RNN) [10], and Feedforward Neural Networks (FNN) [11], can be employed for this purpose.

To model the complex nonlinear relationships between the inputs and outputs of a Two Degree Of Freedom (2-DOF) manipulator robot, which are described in the literature by a highly nonlinear time-varying dynamic model, two MLPs operating in both were proposed in this study, and the simulation tests were used to prove the simplicity and efficiency of this kind of neural network model, which led to its selection.

This document is organized as follows: Section 2 contains the mathematical model and the description of the dynamic model of a 2-DOF robotic arm; Section 3 presents the neural network model used to model this manipulator robot; Section 4 contains the interpretation of the simulation results; the last section provides a conclusion and some perspectives.

II. DYNAMIC MODELING OF TWO-DEGREE-OF-FREEDOM MANIPULATOR ROBOT

Figure 1 shows the representation of the 2-DOF manipulator robot that was employed in this study.

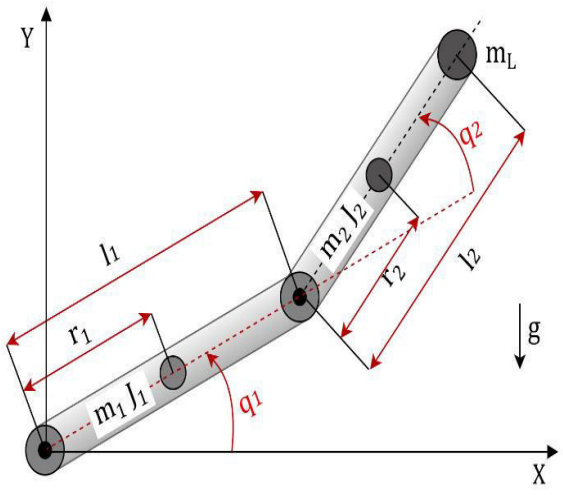


Figure 1 The 2-DOF arm robot.

The mathematical model of the system under study is obtained by applying Lagrange's equation, and it is represented by the equations (1) [12].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \quad (1)$$

Where, the torques of joints 1 and 2, denoted by τ_1 and τ_2, respectively, are the input variables of this arm robot. The angular positions q_1 and q_2 , which correspond to joints 1 and 2, respectively, are the output variables.

$$M_{11} = I_1 + I_2 + m_1 r_1^2 + (m_2 + m_L) l_1^2 + m_2 + (2 m_2 l_1 r_2 + 2 m_L l_1 l_2) \cos(q_2) + m_L l_2^2 \quad (2)$$

$$M_{12} = I_2 + m_2 r_2^2 + (m_2 l_1 r_2 + m_L l_1 l_2) \cos(q_2) + m_L l_2^2 \quad (3)$$

$$M_{21} = M_{12} \quad (4)$$

$$M_{22} = I_2 + m_2 r_2^2 + m_L l_2^2 \quad (5)$$

$$V_1 = -(m_2 l_1 r_2 + m_L l_1 l_2)(2\dot{q}_1 + \dot{q}_2)\dot{q}_2 \sin(q_2) \quad (6)$$

$$V_2 = (m_2 l_1 r_2 + m_L l_1 l_2)\dot{q}_1^2 \sin(q_2) \quad (7)$$

$$F_1 = \mu_1 \dot{q}_1 \quad (8)$$

$$F_2 = \mu_2 \dot{q}_2 \quad (9)$$

$$G_1 = (m_1 r_1 + m_2 l_1 + m_L l_1) g cos(q_1) + (m_2 r_2 + m_L l_2) g cos(q_1 + q_2) \quad (10)$$

$$G_2 = (m_2 + m_L) g r_2 \cos(q_1 + q_2) \quad (11)$$

The nomenclature listed in Table 1 is used to create the mathematical model of the 2-DOF arm robot presented above.

Table 1 The nomenclature of 2-DOF arm robot.

Parameters	Definition
$m_{1,2}$	Mass of link 1 and 2 of the robot.
$l_{1,2}$	Length of each link of the robot.
$r_{1,2}$	Length to center of mass of each link.
m_L	Load mass.
$I_{1,2}$	Inertia of each link.
$\mu_{1,2}$	Friction coefficients.
g	Gravitational acceleration.
$M_{11,22}$	Mass and inertia matrix.
$V_{1,2}$	Centrifugal and Coriolis matrix
$F_{1,2}$	Friction vector.
$G_{1,2}$	Gravity vector

The Table 2 gathers physical parameters [12] of the 2-DOF manipulator robot described previously.

Table 2 The physical parameters of the 2-DOF arm robot.

Parameters	Values	Parameters	Values
m_1	0.3929243 kg	m_L	0.2 kg
m_2	0.0944039 kg	I_1	0.00114111 m^2kg
l_1	0.2032 m	I_2	0.00202470 m^2kg
l_2	0.1524 m	μ_1	0.141231 N m
r_1	0.104648 m	μ_2	0.353078 N m
r_2	0.081788 m	g	9.81 m/s^2

III. NEURAL NETWORK MODELING OF THE 2DOF ARM ROBOT

To model the 2-DOF manipulator robot, two different neural networks type of Multi-Layer Perceptron (MLP) working in both are proposed to predict each output of this robot with a sample time of 10 ms. The first MLP is used to predict the angular position of joint 1 of the 2-DOF manipulator robot, and the second MLP is utilized to predict the angular position of joint 2. This two MLPs has the same architecture, which can be represented with the following figure:

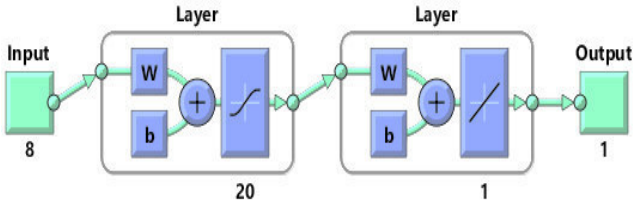


Figure 2 The configuration of one MLP.

The configuration of each MLPs is as follow:

- One input layer containing eight neurons corresponding to the following vector: $[\tau_1(k), \tau_1(k-1), \tau_2(k), \tau_2(k-1), q_1(k), q_1(k-1), q_2(k), q_2(k-1)]$
- Twenty neurons with a sigmoid activation function make up the single hidden layer of each MLP.
- One neuron with a linear activation function is present in the output layer in each MLP, to give the predicted angular positions of joint 1 ($q_1(k+1)$) and 2 ($q_2(k+1)$).

IV. SIMULATION RESULTS

To test the proposed neural network model. Two group of dataset were generated utilizing the state model provided in equation (1), where the random inputs within the system's operational range are used to train each MLP in this neural network model with the Levenberg-Marquardt algorithm.

The obtained test results for joint1 and 2 of the 2-DOF manipulator robot are illustrated in Figure 3 and Figure 4, respectively, and the correlation coefficient and root mean square error are used as a criteria to evaluate this results and its values are gathered in Table 3.

Table 3 The values of RMSE and correlation coefficient of each MLPs.

Joints	Criteria	Values
q_1	$RMSE$	0.00097
	R^2	0.99999
q_2	$RMSE$	0.00074
	R^2	0.99999

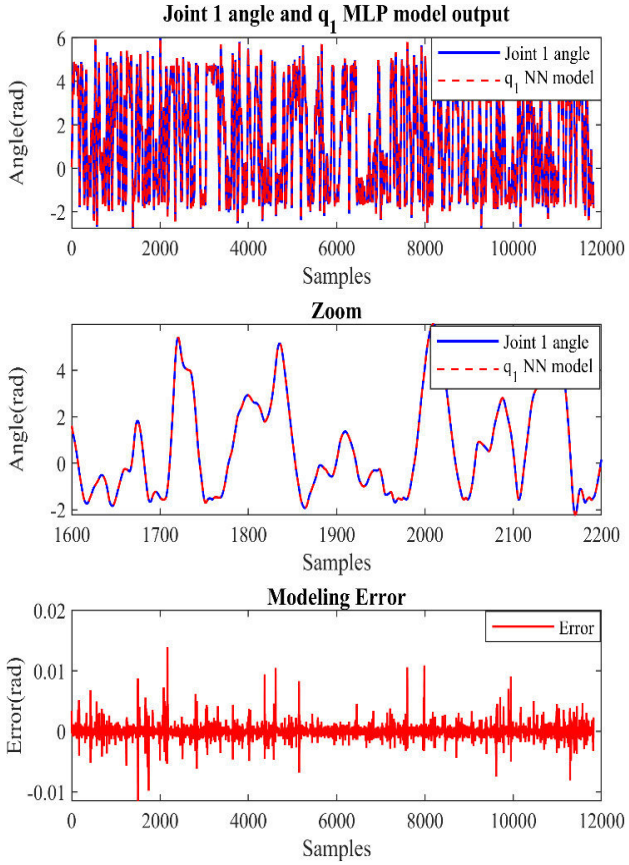


Figure 3 The test results of the neural network model of the angle1.

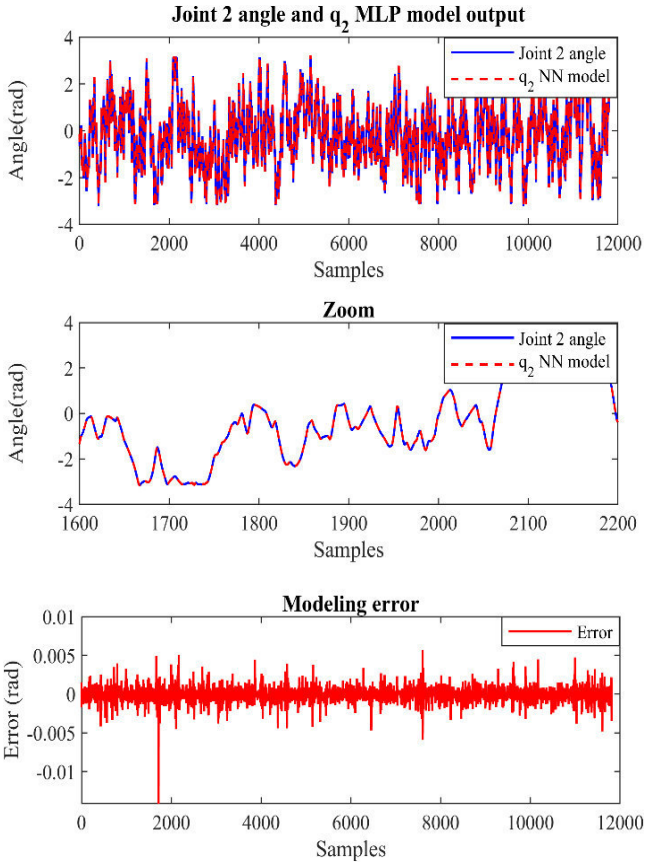


Figure 4 The test results of the neural network model of the angle2.

The simulation results of the proposed neural network model are displayed in **Error! Reference source not found.**, **Error! Reference source not found.**, where it is evident that the modeling error is quite low for the two MLPs corresponding to the q_1 and q_2 models. Due to the values of $RMSE$ and R^2 gathered in Table 3 for each MLPs, we can say that the proposed neural network model show good accuracy, proved by the values of R^2 and $RMSE$ which are nearly equal to one and to zero, respectively.

V. CONCLUSION AND PERSPECTIVES

In this paper, two multilayer perceptron working in both have been proposed as neural network modeling approach, to model the 2-DOF manipulator robot. This modeling approach resulted in a nonlinear model of a simple structure that proved to be an accurate approximation of the arm robot system, and this accuracy was approved by the simulation results obtained and presented.

As future work, we plan to use this obtained neural network model of the arm robot as a prediction model to design a neural network model predictive controller to control the angular positions of the manipulator robot.

REFERENCES

[1] Y. Gao and S. Chien, "Review on space robotics: Toward top-level science through space exploration," *Sci. Robot.*, vol. 2, no. 7, 2017, doi: 10.1126/scirobotics.aan5074.

[2] J. Yuh, "Design and control of autonomous underwater robots: A survey," *Auton. Robots*, vol. 8, no. 1, pp. 7–24, 2000, doi: 10.1023/A:1008984701078.

[3] S. S. M. B. O. G. R. F. S. K. M. Bettayeb, "Tracking Funnel Control Applied on a 4-DOF Robot Manipulator: A Comparative Study," *2023 Adv. Sci. Eng. Technol. Int. Conf.*, 2023.

[4] J. S. Terry, L. Rupert, and M. D. Killpack, "Comparison of linearized dynamic robot manipulator models for model predictive control," *IEEE-RAS Int. Conf. Humanoid Robot.*, pp. 205–212, 2017, doi: 10.1109/HUMANOIDS.2017.8246876.

[5] R. Grandia, A. J. Taylor, A. Singletary, M. Hutter, and A. D. Ames, "Nonlinear Model Predictive Control of Robotic Systems with Control Lyapunov Functions," *Robot. Sci. Syst.*, 2020, doi: 10.15607/RSS.2020.XVI.098.

[6] H. Kanchev, N. Hinov, B. Gilev, and B. Francois, "Modelling and control by neural network of electric vehicle traction system," *Elektron. ir Elektrotechnika*, vol. 24, no. 3, pp. 23–28, 2018, doi: 10.5755/j01.eie.24.3.20974.

[7] J. Topić, B. Škugor, and J. Deur, "Neural network-based modeling of electric vehicle energy demand and all electric range," *Energies*, vol. 12, no. 7, 2019, doi: 10.3390/en12071396.

[8] R. Kruse, C. Borgelt, F. Klawonn, C. Moewes, M. Steinbrecher, and P. Held, *Computational Intelligence: A Methodological Introduction*, no. x. 2013. doi: 10.1007/978-1-4471-5013-8.

[9] A. Saxena, "An Introduction to Convolutional Neural Networks," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 12, pp. 943–947, 2022, doi: 10.22214/ijraset.2022.47789.

[10] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, no. March, pp. 1–43, 2020, doi: 10.1016/j.physd.2019.132306.

[11] Y. Liang, "Deep Learning Basics Lecture 1: feedforward", [Online]. Available: https://www.cs.princeton.edu/courses/archive/spring16/cos495/slides/DL_lecture1_feedforward.pdf

[12] F. Lin, *F. Lin, Robust control design: an optimal control approach*. John Wiley & Sons, 2007.

A Second Order Sliding Mode Control For an Overhead Crane System

Oussama Boutalbi*[1], Khier Benmahammed[1], Mabrouk Boubezoula[1]
Elrachid Bendaoud [2] and Boualem Boukezata[3,4]

***Abstract*—In this paper, we propose an efficient PD-Twisting control algorithm to solve the challenging problem of robust manipulating of heavy loads using a multi-degrees of freedom overhead crane system. The proposed approach provides a unified way to solve both problems of reference tracking and parametric load swings attenuation. An appropriate sliding manifold that ensures the stabilization of the tracking error and the boundedness of the load swings is firstly defined. Then, a discontinuous compensator based on the Twisting Control Algorithm – a Second Order Sliding Mode algorithm – and the feedback linearizing control approach has been introduced to guarantee best performances of the system, with respect to tracking accuracy, load swings attenuation and system robustness. The effectiveness of the proposed approach has been proven through theoretical study and numerical tests.**

***Index Terms*—Second Order Sliding Mode Control, Overhead Crane, Under-actuated Systems, Robust Cpntrol**

I. Introduction

Automated crane systems have attracted great interests of control researchers due the growing applications of cranes in many fields and the theoretical challenges of the control problem [1]–[6]. In industrial/construction sites (e.g., nuclear waste-handling facilities, shipyards, airports, automobile plants, new buildings, etc.) cranes are used to transport large, heavy payloads via a cable attached to a motor-actuated platform. The flexible structure of the crane systems makes them useful to handle different tasks in large work-space environments. However, this flexibility property may be shown as a source of some challenging problems during crane motion, mainly, the parametric excitation of load oscillations. Load oscillations behavior is a natural result of (i) the flexibility in the mechanical structure (ii) the unmodeled dynamics of the plant, like friction forces and external disturbances, like wind and (iii) the time varying parameters of the system such as cable length [7], [8] .

Thus, the immediate concern of any control design is the development of advanced control techniques that take into account load oscillations behavior to reach a high level of motion accuracy and safety despite the uncertain and the under-actuated nature of the system.

*Corresponding Author: botalbioussama@gmail.com
[1]Laboratoire des Systèmes Intelligents, Faculté de Technologie, Université Ferhat Abbas –Sétif 1-
[2]Le Laboratoire d'Automatique de Sétif (L.A.S), Faculté de Technologie, Université Ferhat Abbas –Sétif 1-
[3]Université Mohamed El Bachir El Ibrahimi de Bordj Bou Arréridj
[4]Laboratoire de qualité d'énergie dans les réseaux électriques, Université Ferhat Abbas-Sétif 1

From a general control point of view, several approaches, including intelligent artificial techniques [9] and Sliding Mode Control (SMC) algorithms [10], have been developed to deal with robust control problem of uncertain under-actuated systems. Mainly, conventional algorithms of SMC theory are applied to control the behavior of crane systems [11]–[13]. Conventional SMC techniques are characterized by their simplicity of design and their robustness against a class matched disturbances [14]. However, this class of SM controllers are applied only to systems with relative degree one which limits the degrees of the freedom of the control design and degrades the systems' performances. Recent developments in SMC theory prove that higher order SM is more sufficient than conventional algorithms in terms of system stability, robustness and accuracy [15], [16].

Motivated by the success applications of Second Order Sliding Modes (SOSMC) in many engineering systems, this paper addresses the design of a SOSMC to achieve high tracking/positioning performances for the overhead crane. Using the fact that the overhead crane dynamics is partially feedback linearizable, a composite sliding manifold was defined such that its regulation ensures the boundedness of the tracking error as well as the attenuation of the load swings. A Twisting algorithm, one of the most studied algorithms in SOMC theory [17]–[19], is then introduced in a second step as a discontinuous compensator to improve the system performances in terms of motion accuracy, robustness against structured and unstructured uncertainties and swings attenuation. In addition to its simplicity of design and robustness characteristics, this algorithm provides finite time convergence of the state to the origin. Moreover, under discretization, such controller guarantees a quadratic precision of regulation with respect to the sampling period [16], [20]. A formal proof of the asymptotic stability of the proposed control approach is established by using Lyapunov theory.

The outline of this paper is as follows. Some basic concepts about the robust stabilization of Single Input Single Output (SISO) non-autonomous dynamics are highlights in section2. Section 3 gives the detail of the proposed approach to overcome the challenges of the control design for the multi-degree of freedom crane system. Simulation results illustrate the advantages and the effectiveness of the proposed strategy are given in section 4. Section 5 concludes the paper.

II. Robust Stabilization of a Class of Nonlinear Systems

To introduce the basic concepts of the robust stabilization of non-autonomous systems subjected to unknown perturbation by using sliding mode control theory, let us consider the perturbed output of the form:

$$\begin{cases} \dot{\sigma}_1 = \sigma_2 \\ \dot{\sigma}_2 = u + d(t) \end{cases} \tag{1}$$

The control input u has relative degree two with respect to the output σ. Therefore, a Second Order Sliding Mode control (SOSMC) for the control input may be defined as follows [18], [19]:

$$u = -k_p\sigma_1 - k_d\sigma_2 - r_1 sign(\sigma_1) - r_1 sign(\sigma_2) \tag{2}$$

which guarantees that the system is globally uniformly asymptotically stable if condition of **theorem3** and **assumption.1** are respected.

Note also that the dynamics eq.1 under the control law eq.2 may be rewritten in a state space form as:

$$\dot{\sigma} = A\sigma + Bsign(\sigma) + C \tag{3}$$

with:
$A = \begin{bmatrix} 0 & 1 \\ -k_p & -k_d \end{bmatrix}$,
$B = \begin{bmatrix} 0 & 0 \\ -r_1 & -r_2 \end{bmatrix}$,
$C = \begin{bmatrix} 0 & c \end{bmatrix}^T$,
$\sigma = \begin{bmatrix} \sigma_1 & \sigma_2 \end{bmatrix}^T$, $sign(\sigma) = \begin{bmatrix} sign(\sigma_1) & sign(\sigma_2) \end{bmatrix}^T$,

Assumption 1: It is considered that $|d(t)| \leq c$ for a positive constant c.

Theorem 1: The introduced tracking control algorithm 2 makes the system (eq.1/3) **globally uniformly asymptotically stable** if the following conditions are respected:

$$\begin{cases} r_2 \geq c, \\ r_1 \geq r_2 + c, \\ k_p \geq \frac{1}{\epsilon^2} \\ k_d \geq \frac{1}{\epsilon} \end{cases} \tag{4}$$

for a sufficiently small positive constant ϵ.

III. Control Design for an Overhead Crane System

This section develops a Second Order Sliding mode Control algorithm for an overhead crane system with five degrees of freedom.

A. System Modeling

The dynamics of an overhead crane system with n control inputs may be expressed in the Lagrangian form:

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = B\tau \tag{5}$$

with $q(t) = [q_u\ q_a]^T$ and $\dot{q}(t) = [\dot{q}_u\ \dot{q}_a]^T$ are generalized vectors of coordinates and velocities respectively, $q_a \in \mathbb{R}^n$ represents the actuated part of the system , $q_u \in \mathbb{R}^m$ represents the unactuated part of the system. $M(q)$ is the Inertia matrix, $C(q,\dot{q})$ is the centripetal and Coriolis matrix, G(q) is the gravity vector. . The control input vector is denoted by $\tau \in \mathbb{R}^n$, $B = [b_1\ b_2]$, $b_1 = 0_{[n\times(n-m)]}$ and $b_2 = I_{[n\times n]}$.

The dynamical model (eq.5) preserves some standard properties of mechanical Lagrangian dynamics, including the symmetric positive definite property of the inertia matrix M, the skew-symmetry of the matrix $\dot{M} - C(q,\dot{q})$.

For systems that are trivially under-actuated such as an overhead crane, we can reorganize the generalized coordinates described by eq.5 into the following form:

$$\begin{cases} m_{11}\ddot{q}_u + m_{12}\ddot{q}_a + f_1(q,\dot{q}) = 0 \\ m_{21}\ddot{q}_u + m_{22}\ddot{q}_a + f_2(q,\dot{q}) + \delta d = \tau \end{cases} \tag{6}$$

with $\delta d \in \mathbb{R}^n$ represents an unknown bounded vector introduced here to describe all plant uncertainties, including external disturbances and unstructured unmodeled dynamics.

As a consequence of the uniform definitiveness of the inertia matrix M; the term m_{11} is 2*2 invertible matrix. Therefore, we may solve $\ddot{q}_u$ the first equation in 6 as:

$$\ddot{q}_u = -m_{11}^{-1}(m_{12}\ddot{q}_a + f_1(q,\dot{q})) \tag{7}$$

And substitute the resulting expression eq.7 into the second equation of 6 to obtain:

$$\bar{m}_{22}\ddot{q}_a + \bar{f}_2 = \tau \tag{8}$$

where: $\bar{m}_{22} = m_{22} - m_{21}m_{12}m_{11}^{-1}$ and $\bar{f}_2 = f_2 - m_{21}m_{11}^{-1}f_1$

A feedback linearizing controller can therefore be defined for equation 8 according to

$$\tau = \bar{m}_{22}u + \bar{f}_2 \tag{9}$$

where $u \in \mathbb{R}^n$ is an auxiliary controller to be determined. Then the complete system up to this point may be written as:

$$\begin{cases} m_{11}\ddot{q}_u + f_1 = -m_{12}u \\ \ddot{q}_a = u + \delta d \end{cases} \tag{10}$$

Since the input-output relationship from q_a to u in eq.10 is linear, the active part of eq.6 has been completely linearized. However, considering the full state vector, only partial linearization has been achieved.

B. Feedback control design

Given a desired trajectory vector $q_{ad}(t) = [x_d(t)\ y_d(t)\ l_d(t)]$, the tracking error is

$$\begin{cases} \eta_1 = q_u, \\ \eta_2 = \dot{q}_u, \\ e_1 = q_a - q_{ad}(t) \\ e_2 = \dot{q}_a - \dot{q}_{ad}(t) \end{cases} \tag{11}$$

The main goal of the control design is to define the control inputs that ensure the stabilization of the tracking errors $e_1(t)$ and $e_2(t)$ as well as the load swings $\eta_1,\ \eta_2$ to zero. We may write the system eq.11 in state space as:

$$\begin{cases} \dot{e}_1 = e_2 \\ \dot{e}_2 = u + d_0 \\ \eta_1 = q_u, \\ \eta_2 = \dot{q}_u, \end{cases} \tag{12}$$

where $d_0 = -\ddot{q}_{ad}(t) + \delta d$

For the stabilization process of the system eq.12, let us consider the coupled sliding manifold σ of the form:

$$\sigma = e_1 + k[\eta_1 + h\int_0^t \eta_1 dx] \tag{13}$$

with $\sigma \in \mathbb{R}^3$, h positive tunning parameter. The matrix $k \in \mathbb{R}^{3\times 2}$ is defined as: $\begin{bmatrix} k_\theta & 0 \\ 0 & k_\beta \\ 0 & 0 \end{bmatrix}$ with k_θ and k_β are positive coupling factors and they have an important role in the dynamic controller.

The system has relative degree two with respect to the output σ . Taking its second derivative, one gets:

$$\ddot{\sigma} = (I + km_{11}^{-1}m_{12})u + km_{11}^{-1}(f_1 + m_{12}\eta_2) + d_0 \tag{14}$$

whith $I_{(3\times 3)}$ is the unit matrix. Since the matrix $(I + km_{11}^{-1}m_{12})$ is invertible, if a new control (a feedback linearizing controller) is introduced as:

$$u = (I + km_{11}^{-1}m_{12})^{-1}[v + km_{11}^{-1}(f_1 + m_{12}\eta_2)] \tag{15}$$

it reduces the system eq.14 to

$$\ddot{\sigma} = v + d_0 \tag{16}$$

We may rewrite system (eq.16) as a system of three decoupled double-integrators, like:

$$\begin{cases} \ddot{\sigma}_x = v_x + d_{0x} \\ \ddot{\sigma}_y = v_y + d_{0y} \\ \ddot{\sigma}_l = v_l + d_{0l} \end{cases} \tag{17}$$

where v_x , v_x and v_x are new control inputs should designed to stabilize the resulting independent dynamics $\ddot{\sigma}_x$, $\ddot{\sigma}_y$ and $\ddot{\sigma}_l$, receptively, despite model uncertainties and external disturbances. A proper choice of the control input v_α (where $\alpha = x, y, l$) is the PD-Twisting control algorithm discussed above. Thus, three controllers are defined as follows:

$$\begin{cases} u_x = -k_{px}\sigma_{1x} - k_{dx}\sigma_{2x} - r_{1x}sign(\sigma_{1x}) - r_{2x}sign(\sigma_x) \\ u_y = -k_{py}\sigma_{1y} - k_{dy}\sigma_{2y} - r_{1y}sign(\sigma_{1y}) - r_{2y}sign(\sigma_{2y}) \\ u_l = -k_{pl}\sigma_{1l} - k_{dl}\sigma_{2l} - r_{1l}sign(\sigma_{1l}) - r_{2l}sign(\sigma_{2l}) \end{cases} \tag{18}$$

The control parameters $k_{p\alpha}, k_{d\alpha}, r_{1\alpha}$ and $r_{2\alpha}$, $\alpha = x, y, l$ are chosen in order to satisfy the stability conditions listed in $theorem.3$, inequalities 4.

As seen above, controllers in eq.18 allows the designer to stabilize the output eq.13 to zero, which, in its turn ensures the convergence of the tracking error $e(t)$ as well as the swing angles vector η to zero if the positive tunning parameters k_θ, k_β and h are suitably chosen as we shown in [21].

IV. Numerical tests

Computer simulations; were performed in Matlab/Simulink environment; have been carried out to asses the performances of the proposed approach, using the 5-DOF crane model depicted in Fig.1. The system parameters are taken as:

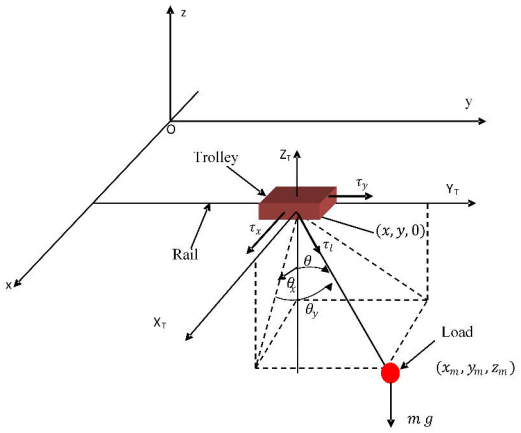


Fig. 1. 3D Overhead Crane System

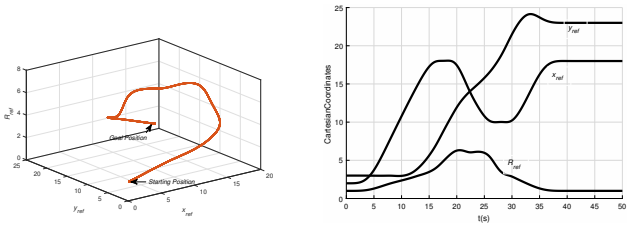
Mass of the cart : $m_c = 10 \ \ kg$
Mass of moving rail: $m_w = 20 \ \ kg$
Cable with zero mass and variable length.
Mass of the payload: $m_p = 10 \ \ kg$
The acceleration gravity was taken as: $g = 10m/s^{-2}$.

As seen above, the system dynamics may be arranged to take the form defined in eq.5. In this case, the system is driven with three control inputs f_x, f_y and f_l. Thus the vector of driving control inputs is defined as $\tau = [f_x \ f_y \ f_l]$. The vector of actuated degrees of freedom is given as $q_a = [x \ y \ l]^T$, where x and y are the Cartesian co-ordinates of the lifted load in fixed frame $OXYZ$, l

Unstructured uncertainties are simulated by considering unknown external disturbances and mechanical friction forces in the moving cart, the moving rail and the hoisting mechanism which are modeled using the LuGre friction model, because it can capture most of the friction behavior that has been, in general, observed experimentally in control systems. Full representation of the crane system in the form 1 may be shown in [21].

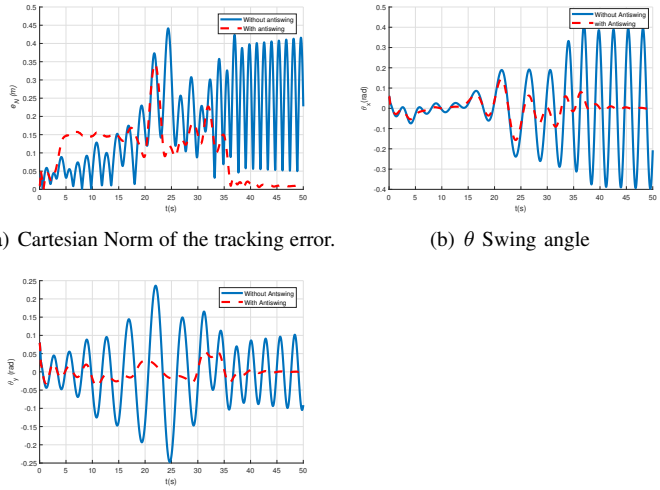
A. Task Specification

In the proposed tests, we consider the task of moving the load from its starting position $r_A = [x_A \ \ y_A \ \ l_A]$ to the target position $r_B = [x_B \ \ y_B \ \ l_B]$ along the specified path shown in Fig.2.a as fast as possible, without oscillations and within the physical limits of the system (velocity and acceleration limits). The corresponding 3D vector of reference trajectories is defined as a combination of the free-collision path that interpolates, through a sequence of waypoints, the starting and the goal positions and the time-dependent motion profile that describes the load dynamics along the prescribed path. The interpolation algorithm, as well as the motion profile, should ensure a sufficient continuity, at least C^2, to achieve feasible motion of the crane system. For more details about this planning approach, the reader is referred to [21], [22]. The total length of the planned path shown in Fig.2.a is $L = 47.3158 \ meters$. The whole execution time of the task



(a) Spesified task (b) References trajectories

Fig. 2. Specified task



(a) Cartesian Norm of the tracking error. (b) θ Swing angle

(c) β Swing angle

Fig. 3. Performances of the anti-swing controller

was fixed as $T = 50s$ which leads to the time-dependent reference trajectories seen in Fig.2.b.

B. Anti-swing controller analysis

The aim of this test is to highlight the importance of the proposed anti-swing controller in the overall behavior of the system. Therefore, two controllers were considered for a comparative study:

A SOSMC controller based on the sliding manifold defined in eq.13, with the following design parameters:

$k_\theta = 3, \quad k_\beta = 1, \quad h = 1$

A SOSMC controller based on a sliding manifold without anti-swing part, i.e.:

$k_\theta = 0, \quad k_\beta = 0, \quad h = 0, \sigma = e$

In both tests, the feedback control parameters are fixed as follows:

$k_{px} = 9, \quad k_{dx} = 6, \quad r_{1x} = 10, \quad r_{2x} = 5$
$k_{py} = 9, \quad k_{dy} = 6, \quad r_{1y} = 10, \quad r_{2y} = 5$
$k_{pl} = 20, \quad k_{dl} = 9, \quad r_{1l} = 4, \quad r_{2l} = 2$

The Cartesian norm of the tracking and swing angles characterizing the motion of crane by applying the above controllers are shown in Fig.3. Here, note that the Cartesian norm of the tracking error is obtained as:

$$e_N = \sqrt{(x - x_d)^2 + (y - y_d)^2 + (l - l_d)^2} \qquad (19)$$

As seen, the depicted results in Fig.3.b and Fig.3.c clarify the importance of the antiswing part in the motion of the system. Without antiswing part, undamped and growth load swings are presented during the load motion. Such load swings are naturally induced from the accelerations and declarations phases that characterize the system motion along the x and y axis as well as the parametric resonance along the $z - axis$. The introduction of the antiswing part allows the designer to attenuate significantly the load swings and eliminates the excitation of the parametric resonance of the system and minimizes the load swings during motion, as highlights in Fig.3.a and Fig.3.b with dashed lines. Moreover, load swings attenuation via the anti swing part improves significantly the tracking accuracy and mainly the final positioning of the load in its target position, Fig.3.a.

C. Tracking control analysis

The aim of this test is to evaluate the performances of the proposed SOSMC, where tracking accuracy, load swings attenuation and smoothness of the control signals are tacking as matter characteristics to evaluate the performances of the system.

Considering the sliding manifold eq.13, with the following design parameters:

$k_\theta = 3, \quad k_\beta = 1, \quad h = 1$

Then, for a comparative study, three approaches are used for the stabilization of the dynamics eq.17:

1) A pure PD controller with the following parameters:

$k_{px} = 9, \quad k_{dx} = 6$
$k_{py} = 9, \quad k_{dy} = 6$
$k_{px} = 20, \quad k_{dx} = 9$

2) A First Order Sliding Mode Controller designed as [23]:

$v_\alpha = -M_\alpha sign(\sigma_{1\alpha} + \lambda_\alpha \sigma_{2\alpha})$
$M_\alpha = \psi_\alpha |\sigma_{1\alpha}| + \Psi_\alpha |\sigma_{2\alpha}| + \delta_{0\alpha}$
$\alpha = [x, y, l]$

where $\lambda_\alpha, \psi_\alpha, \Psi_\alpha$ and $\delta_{0\alpha}$ are positive design parameters were chosen in this paper as follows:

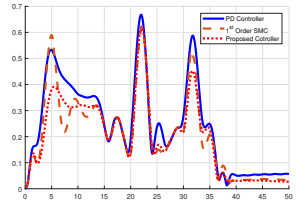
$\lambda_x = 1, \psi_x = 10, \Psi_x = 6, \delta_x = 4$
$\lambda_y = 1, \psi_y = 10, \Psi_y = 7, \delta_y = 4$
$\lambda_l = 1, \psi_l = 20, \Psi_l = 5, \delta_l = 4$

3) The SOSMC controller discussed above, where the control parameters are fixed as follows:

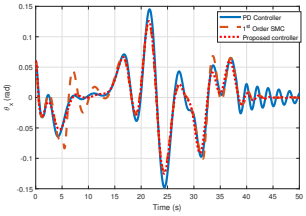
$k_{px} = 9, \quad k_{dx} = 6, \quad r_{1x} = 10, \quad r_{2x} = 5$
$k_{py} = 9, \quad k_{dy} = 6, \quad r_{1y} = 10, \quad r_{2y} = 5$
$k_{px} = 20, \quad k_{dx} = 9, \quad r_{1x} = 4, \quad r_{2x} = 2$

Obtained results are depicted in Fig.4. The corresponding Cartesian Norms of tracking errors, as shown in Fig.4.a, highlights the high level accuracy of the proposed PD-T control approach compared to the PD controller, where the proposed method reduces the RMS, Root Mean Square, of the tracking error to $RMS_{PD-T} = 0.222m$, whereas the RMS values of the tracking errors are $RMS_{1^{st}SM} = 0.2463m$ and $RMS_{PD} = 0.2967$ for the 1^{st} Order SMC and the PD controller, receptively. Moreover, Fig.4.b and Fig.4.c illustrate

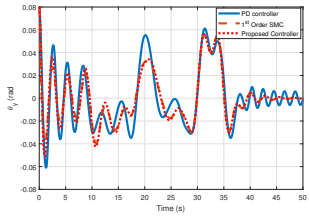
the positive performances of the proposed PD-Twisting algorithm in the attenuation of undesirable load swings. The



(a) Cartesian Norm of the tracking error.



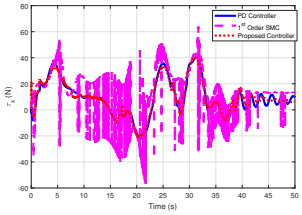
(b) Swing angle



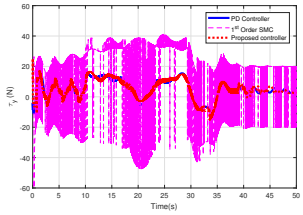
(c) Swing angle

Fig. 4. Tracking performances

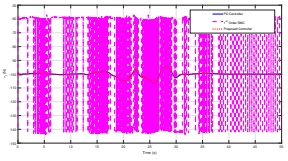
corresponding driving control inputs are depicted in figure 5. In addition to the above results related to the high level of accuracy for the proposed PD-T algorithm, one can clearly note the smoothness of the control signal for the proposed approach compared to the classical $1^{st}OSMC$.



(a) Travelling control input



(b) Traverse control input



(c) Lifiting control input

Fig. 5. Driving Control inputs

V. Conclusion

In this paper, we propose a Second order Sliding Mode controller for the robust and the accurate manipulation of heavy loads using multi-degrees of freedom overhead crane system. The key idea of this paper consists on (*i*) defining the sliding manifold as coupled output that ensures both processes of accurate tracking of predefined trajectories, the attenuation of any load swings may be faced during the transfer sequence and eliminates parametric excitation of load swings (*ii*) and then designing second order discontinuous controller based on the twisting algorithm that forces the stabilization of the specified output to zero despite mode uncertainties and external disturbances. Numerical tests are provided to illustrate the performances of the proposed control approach.

References

[1] H.-h. Lee, "Modeling and Control of a Three-Dimensional Overhead," *Journal of Dynamic Systems, Measurement, and Control, ASME*, vol. 120, no. December, pp. 471–476, 1998.

[2] O. Sawodny, H. Aschemann, and S. Lahres, "An automated gantry crane as a large workspace robot," *Control Engineering Practice*, vol. 10, pp. 1323–1338, 2002.

[3] E. M.Abdel-Rahman, A. H.Nayfeh., and M. N.Ziyad, "Dynamics and Control of Cranes : A Review," *Journal of Vibration and Control*, vol. 9, pp. 863–908, 2003.

[4] F. Omar, F. KARRAY, O. BASIR, and L. YU, "Autonomous Overhead Crane System Using a Fuzzy Logic Controller," *Journal of Vibration and Control*, vol. 10, pp. 1255–1270, 2004.

[5] L. Ramli, Z. Mohamed, A. M. Abdullahi, H. I. Jaafar, and I. M. Lazim, "Control strategies for crane systems : A comprehensive review," *Mechanical Systems and Signal Processing*, vol. 95, pp. 1–23, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.ymssp.2017.03.015

[6] Z. You, W. Xu, X. Xu, X. Zhu, and Q. Hu, "An adaptive integral sliding mode control method for overhead cranes with monocular visual feedback," in *2021 40th Chinese Control Conference (CCC)*, 2021, pp. 379–384.

[7] M. P. Cartmell, "On the Need for Control of Nonlinear Oscillations in Machine Systems," *Meccanica*, vol. 38, pp. 185–212, 2003.

[8] J. Peña, R. H. B. Fey, and H. Nijmeijer, "An Introduction to Parametric Resonance," in *Parametric Resonance in Dynamical Systems*, eds ed. Springer, New York, NY, 2012, pp. 1–13.

[9] F. Lewis, J. Campos, and R. Selmic, *Neuro-Fuzzy Control of Industrial Systems with Actuator Nonlinearities*. Society for Industrial and Applied Mathematics, 2002. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9780898717563

[10] V. Utkin, G. Jurgen, and J. Shi, *Sliding Mode Control in Electro-Mechanical Systems*, second edi ed. CRC Press Taylor & Francis Group, 2009.

[11] D. T. Liu and W. P. Guo, "Tracking Control for an Underactuated Two-Dimensional Overhead Crane," *Journal of Applied Research and Technology*, vol. 10, no. 4, pp. 597–606, 2012.

[12] J. Collado, L. Fridman, and V. Carlos, "Author ' s Accepted Manuscript Super Twisting Control of a Parametrically Excited Overhead Crane $," *Journal of the Franklin Institute*, p. 9, 2013.

[13] G. Bartolini, A. Pisano, and E. Usai, "OUTPUT-FEEDBACK CONTROL OF CONTAINER CRANES : A COMPARATIVE ANALYSIS," *Asian Journal of Control*, vol. 5, no. 4, pp. 578–593, 2003.

[14] V. I. Utkin, "Sliding Mode Control Design Principles and Applications to Electric Drives," *IEEE Transactions on Industrial Electronics*, vol. 40, no. 1, pp. 23–36, 1993.

[15] S. V. Emeryanov and S. K. Korovin, "High-Order Sliding Modes in Control Systems," *Computational Mathematics and Modelefing,*, vol. 7, no. 3, pp. 294–318, 1996.

[16] A. Levant, "Sliding order and sliding accuracy in sliding mode control," *International Journal of Control*, vol. 58, N°.6, no. October 2012, pp. 37–41, 2007.

[17] Y. Orlov, "Finite time stability and robust control synthesis of uncertain switched systems ," *SIAM JOurnal of Control and Optimization*, vol. 43, no. 4, pp. 1253–1271, 2005.

[18] Y. V. Orlov, *Discontinuous Systems Lyapunov Analysis and Robust Synthesis under Uncertainty Conditions*. Springer, 2009.

[19] C. Vázquez, J. Collado, and L. Fridman, "Control of a Parametrically Excited Crane : A Vector Lyapunov Approach," *IEEE Transaction On Control Systems Technology*, pp. 1–9, 2013.

[20] V. Torres-González, T. Sanchez, L. M. Fridman, and J. A. Moreno, "Design of Continuous Twisting Algorithm," *Automatica journal*, vol. 80, pp. 119–126, 2017.

[21] O. Boutalbi, K. Benmahammed, and M. Boubezoula, "Robust feedback linearizing controller design for an overhead crane system," in *2017 6th International Conference on Systems and Control (ICSC)*, May 2017, pp. 640–645.
[22] W. Blajer and K. Kołodziejczyk, "Motion planning and control of gantry cranes in cluttered work environment," *IET Control Theory Appl*, no. 5, pp. 1370–1379, 2007.
[23] M. Boubezoula, A. Hassam, and O. Boutalbi, "Robust-flatness Controller Design for a Differentially Driven Wheeled Mobile Robot Mabrouk Boubezoula ," *International Journal of Control, Automation and Systems*, vol. 16, no. 4, pp. 1895–1904, 2018.

VI. Appendix

A. Dynamical model for an overhead crane system

Considering the crane system 1, its dynamical model may be written in the Lagrangian forms (eq.5,6), where the corresponding matrices are given as follows:

$$M_x = m_c + m_w$$
$$M_y = m_c$$
$$M_l = m$$
$$m_{11} = \begin{bmatrix} ml^2cos^2(\theta_y) & 0 \\ 0 & ml^2 \end{bmatrix}$$
$$m_{12} = \begin{bmatrix} mlcos(\theta_x)cos(\theta_y) & 0 & 0 \\ -mlsin(\theta_x)(\theta_y) & mlcos(\theta_y) & 0 \end{bmatrix}$$
$$m_{21} = \begin{bmatrix} mlcos(\theta_x)cos(\theta_y) & -mlsin(\theta_x)sin(\theta_y) \\ 0 & mlcos(\theta_y) \\ 0 & 0 \end{bmatrix}$$
$$m_{22} = \begin{bmatrix} M_x + m & 0 & msin(\theta_x)cos(\theta_y) \\ 0 & M_y + m & mlsin(\theta_y) \\ msin(\theta_x)cos(\theta_y) & mlsin(\theta_y) & M_l + m \end{bmatrix}$$
$$c_{11} = \begin{bmatrix} -ml^2\dot{\theta}_x sin(\theta_y)cos(\theta_y) & -ml^2\dot{\theta}_x sin(\theta_y)cos(\theta_y) \\ ml^2\dot{\theta}_x sin(\theta_y)cos(\theta_y) & ml\dot{l} \end{bmatrix}$$
$$c_{12} = \begin{bmatrix} 0 & 0 & ml\dot{\theta}_x cos^2(\theta_y) \\ 0 & 0 & ml\dot{\theta}_y \end{bmatrix}$$
$$c_{21} = \begin{bmatrix} A & B \\ 0 & m\dot{l}cos(\theta_y) - ml\dot{\theta}_y sin(\theta_y) \\ -ml\dot{\theta}_x cos^2(\theta_y) & ml\dot{\theta}_y \end{bmatrix}$$
$$A = m\dot{l}cos(\theta_x)cos(\theta_y) - ml\dot{\theta}_x sin(\theta_x)cos(\theta_y) - ml\dot{\theta}_y cos(\theta_x)sin(\theta_y)$$
$$B = -m\dot{l}sin(\theta_x)sin(\theta_y) - ml\dot{\theta}_x cos(\theta_x)sin(\theta_y) - ml\dot{\theta}_y sin(\theta_x)cos(\theta_y)$$
$$c_{22} = \begin{bmatrix} 0 & 0 & m\dot{\theta}_x cos(\theta_x)cos(\theta_y) - m\dot{\theta}_y sin(\theta_x)sin(\theta_y) \\ 0 & 0 & m\dot{\theta}_y cos(\theta_y) \\ 0 & 0 & 0 \end{bmatrix}$$
$$G_1 = \begin{bmatrix} mglsin(\theta_x)cos(\theta_y) \\ mglcos(\theta_x)sin(\theta_y) \end{bmatrix}$$
$$G_2 = \begin{bmatrix} 0 \\ 0 \\ mgcos(\theta_x)cos(\theta_y) \end{bmatrix}$$

Open-Circuit and Short-Circuit Faults in Three-Phase Three Level Inverter.

Fatma Khater
Laboratory of Applied Automatic, dept. of Automation
University M'hamed Bougara
Boumerdes, Algeria
f.khater@univ-boumerdes.dz

Abderrezak Aibeche
Laboratory of Applied Automatic, dept. of Automation
University M'hamed Bougara
Boumerdes, Algeria
aibeche_umbb@univ-boumerdes.dz

***Abstract*— Multilevel-Multicellular inverters have become essential components in various industrial applications. Fault diagnosis, reliability and security of multicellular converters are crucial tools for safeguarding electrical power conversion and preventing damage and minimizing maintenance costs, ensuring the continuity of electrical drives. The reliability, continuity, and robustness of a three-phase multicellular inverter heavily rely on the precise diagnosis of faults in insulated gate bipolar transistor (IGBT) switches. This study focuses on analyzing the operational behavior of a three-phase multicellular inverter under both normal and faulty conditions, with particular emphasis on open-circuit and short-circuit faults in the converter switches. To achieve this objective, the paper introduces a fault diagnosis technique examining the impact of a faulty three-phase multilevel-multicellular inverter on the overall performance of the electrical system.**

Keywords—Three-phases multicellular inverter, faults diagnosis, open-circuit faults, short-circuit faults.

I. INTRODUCTION

Power electronics have significantly advanced over the past few decades, driven by the development of semiconductor power components for high-speed switching converter systems and the emergence of new converter designs. Some systems prioritize high switching frequencies, while others are optimized for transferring high power levels, such as multi-level, multi-cell, and multi-cell stepped structures [1]. Multicellular converters display nonlinear characteristics resulting from their switching behavior, creating challenges for control design and robust stability analysis [2].

The concept of multicell structures is a relatively recent development, emerging in the 1990s. It involves a series of commutation cells interconnected by floating voltage sources, which are implemented using capacitors [3], [4].

These converters are commonly used in both industry and research, prompting the development of specialized control strategies to ensure their proper operation. Various approaches have been explored in the literature to develop control and observation methods for multicellular converters.

The modeling of the multicellular converter constitutes a crucial step in the implementation of control laws [5]. Previous research suggests that modeling multilevel-multicell converters presents considerable difficulty; indeed, the latter contains continuous variables (currents and voltage) and discrete variables (switches). In the literature, there are three types of models: the average, harmonic and the instantaneous models [6].

This type of converters requires complex control. While various control techniques, such as Pulse Width Modulation (PWM), closed-loop control strategies like the Sliding Mode Controller (SMC) have been developed to enhance stability and robustness [7], [8].

The reliability of multilevel-multicellular converters is crucial, as a single component failure can result in system malfunction. Therefore, fault diagnosis is essential for implementing fault-tolerant control strategies in these converters. More than 30% of converter faults are caused by power semiconductor switches, particularly insulated gate bipolar transistors (IGBTs) and metal-oxide semiconductor field-effect transistors (MOSFETs).

Multicellular converters are highly susceptible to power semiconductor failures, which can be categorized as open-circuit (OC) and short-circuit (SC) faults [9].

In general, there are mainly two categories of fault diagnosis techniques: model-based techniques and signal-based techniques [10]. In this study, the output voltage and current are used as monitored variables for fault diagnosis.

The rest of the manuscript is organized as follows, after a brief presentation of the subject in the introduction, section 2 provides an overview of the three-phase multilevel-multicellular inverter model. Subsequently, section 3 introduces and tests the sliding mode controller. Following this, the focus shifts to investigating faults within the converter, specifically concentrating on the diagnosis of open circuit and short circuit faults in section 4. Finally, in section 5, concluding remarks are presented.

II. THREE-PHASE MULTILEVEL-MULTICELL INVERTER MODEL

A three-phase multilevel-multicellular inverter consists of three single-phase multilevel-multicellular inverters connected in parallel and controlled to produce output voltages that are shifted by 120 degrees. Figure 1 presents a representative diagram of a floating capacitor three-phase multilevel-multicellular inverter. It generates a three-phase load through resistance R and inductance L. E represents the DC input voltage. This is a hybrid system, which both continuous variables (voltages and currents) and discrete variables (binary switches) evolve.

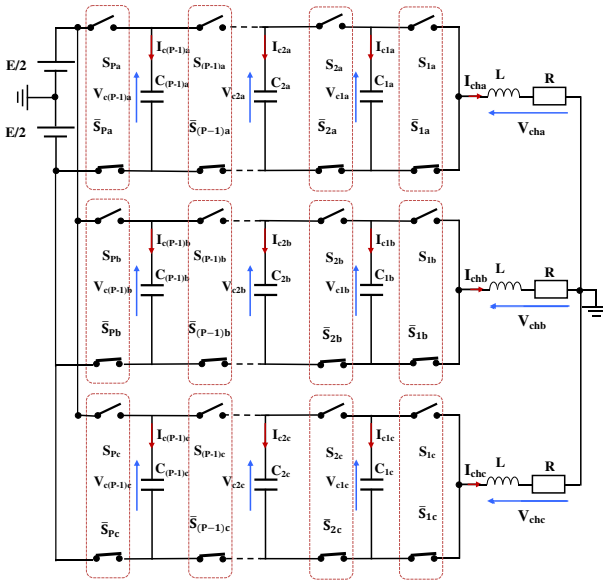


Fig. 1. Representative diagram of a three-phase multilevel-multicellular inverter.

The following definitions are provided with reference to phase (a), but they apply to all phases. Each phase of the inverter is composed of p elementary switching cells connected in series, with each cell containing two complementary switches. It is controlled by a binary signal uk; when the cell's upper switch is conducting, this signal equals 1, and 0 when the lower complementary switch of the cell is conducting. These cells are connected to the R-L load in series and separated by floating capacitors. There are p-1 floating voltage sources [11]. There are n_{cells} + 1 voltage levels available at the output of each phase.

Each cell is made up of two adjacent floating capacitors, V_{Ck} and V_{Ck-1}, with V_{C0}=0 and V_{Cp}=E.

$$V_{Cell_k} = V_{Ck} - V_{Ck-1} \tag{1}$$

An equal distribution of voltage constraints must be applied to each floating capacitor, i.e.,

$$V_{Ck} = k.E/p \tag{2}$$

The voltage applied to the load is given by:

$$V_{ch} = \sum_{k=1}^{P} S_k \cdot V_{cell_k} \tag{3}$$

The voltage across capacitor C_k is related to the current C_k by:

$$i_{Ck}(t) = (S_{cell_{k+1}} - S_{cell_k}) \cdot i_{ch}(t) \tag{4}$$

$$\frac{dV_{Ck}}{dt} = \frac{(S_{cell_{k+1}} - S_{cell_k})}{C_k} \cdot i_{ch}(t) \tag{5}$$

In the case of a load of type R L, the equation giving the evolution of the current i_{ch} is obtained from the voltage V_{ch}.

$$V_{ch} = R \cdot i_{ch} + L \cdot \frac{d}{dt} i_{ch} \tag{6}$$

$$\frac{d}{dt} i_{ch} = \frac{V_{ch}}{L} - \frac{R}{L} \cdot i_{ch} \tag{7}$$

By grouping these expressions in the form of state equations, we obtain the instantaneous model describing a multicellular cell p operating as an inverter associated with a load R L:

$$\begin{aligned}
&\frac{d}{dt} V_{C1} = \frac{[S_2 - S_1]}{C_1} \cdot i_{ch} \\
&\frac{d}{dt} V_{C2} = \frac{[S_3 - S_2]}{C_2} \cdot i_{ch} \\
&\vdots \\
&\frac{d}{dt} V_{Cp-1} = \frac{[S_p - S_{p-1}]}{C_{p-1}} \cdot i_{ch} \\
&\frac{d}{dt} i_{ch} = \frac{[S_1 - S_2]}{L} \cdot V_{C1} + \frac{[S_2 - S_3]}{L} \cdot V_{C2} \\
&+ \cdots + \frac{[S_{p-1} - S_p]}{L} \cdot V_{Cp-1} - \frac{R}{L} \cdot i_{ch} + \frac{S_p}{L} \cdot E
\end{aligned} \tag{8}$$

Based on the given differential equations, we construct a state representation with the state variables being the floating voltages V_{ck} and the load current ich, which has the following form:

$$\dot{X} = AX + BU \tag{9}$$

With: $X = \left[V_{C1}, V_{C2} \cdots, V_{Cp-1}, i_{ch}\right]^T$; U: the control vector.

III. Control of Three-Phases Multilevel Multicell Inverter

The control system of a three-phase multicellular inverter must ensure the regulation of output currents and voltages in each phase while maintaining an even distribution of stress across each switch. Various control approaches have been developed and documented in the literature for these types of inverters. Pulse Width Modulation (PWM) control, an open-loop method often referred to as natural control, ensures the correct operation of the inverter and the balancing of capacitor terminal voltages. However, closed-loop control strategies are essential for ensuring stability, robustness, and continuous proper operation of the inverter. One such strategy is the application of sliding mode control, which considers the evolution of output voltages, currents, and capacitor voltages.

The use of the sliding mode control technique on our three-phase multilevel-multicell inverter is the focus of this section.

Sliding mode control (SMC) is a type of nonlinear controller that is a powerful design method for many systems, based on Lyapunov's method. It is considered an effective approach in control systems due to its robustness against model uncertainties and external disturbances. The objective of control through sliding modes can be defined in two key points: first, creating a surface S such that all system trajectories follow the desired tracking behavior, regulation, and stability; and second, determining a control law U that attracts and retains all state trajectories on the sliding surface [12].

The multilevel-multicellular inverter is a nonlinear system with a variable structure and multiple inputs. The synthesis of sliding mode control for such a system involves defining a sliding surface for each switching cell. It is then necessary to associate these surfaces with an appropriate switching law, ensuring the stability and continuous operation of the converter.

We consider the system defined by equation (9), where the dimension of the surface vector S matches that of the control vector U.

To ensure the convergence of a state variable x towards its reference values x_{ref}, various forms of sliding surfaces have been introduced. Typically, the selection of a surface is based on the error in the controlled variable denoted as e(x), where e(x) = x_{ref} - x. Thus, we define p sliding surfaces as follows:

$$\begin{aligned} S_1 &= V_{C\,ref\,1} - V_{C1} \\ S_2 &= V_{Cref} - V_{C2} \\ &\vdots \\ S_p &= i_{Ch_{ref}} - i_{ch} \end{aligned} \tag{10}$$

In order to verify the convergence condition, we use the Lyapunov approach. The Lyapunov's function is chosen as follows:

$$V(x) = \frac{1}{2} S^2(X) \tag{11}$$

To ensure the decrease of the Lyapunov function, it is sufficient to guarantee that its derivative is negative.

$$\dot{V}(x) = \dot{S}(X) \cdot S(X) < 0 \tag{12}$$

The structure of the sliding mode controller consists of two components $U=U_{eq}+U_n$.

U_{eq} enables the maintenance of the controlled variable on the sliding surface S = 0. This command is deduced by considering that the derivative of the surface is zero. Un is determined to satisfy the convergence condition. During the sliding surface and the steady state, the surface is zero, from which we deduce the expression of U_{eq} and Un.

$$U_{eq} = -\left[\frac{\partial S}{\partial X}.B(X)\right]^{-1} \cdot \left[\frac{\partial S}{\partial X} A(X)\right] \tag{13}$$

$$U_n = K_i sign(S(X)) \tag{14}$$

The values of the examined inverter parameters are summarized in Table 1.

TABLE I. OPERATING PARAMETERS OF THE THREE-PHASES THREE-LEVEL INVERTER.

Parameter	Description	Values
Vin [V]	Input DC voltage	2000
Ci, j [F]	Capacitor	3.10^{-4}
Rch, j [Ω]	Load resistance	5
Lch, j [H]	Load inductance	5.10^{-2}
f_{PWM} [kHz]	The switching frequency	5,5
Te [sec]	The switching period	1.10^{-6}

The following figures respectively illustrate the load voltage, load current, and floating capacitor voltage of the three-phase three-level inverter controlled by Sliding Mode Control (SMC).

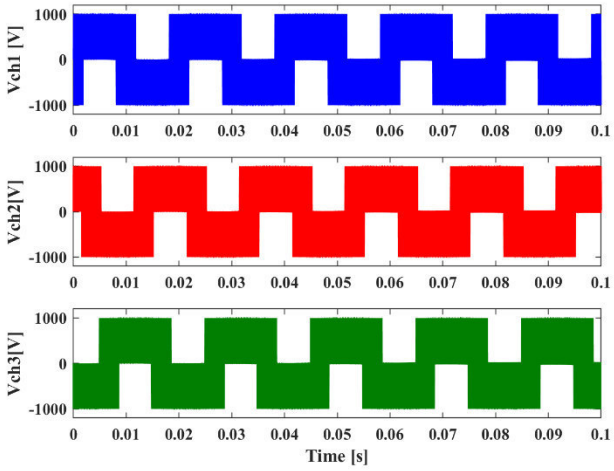


Fig. 2. Load voltage of the three-phase three-level inverter.

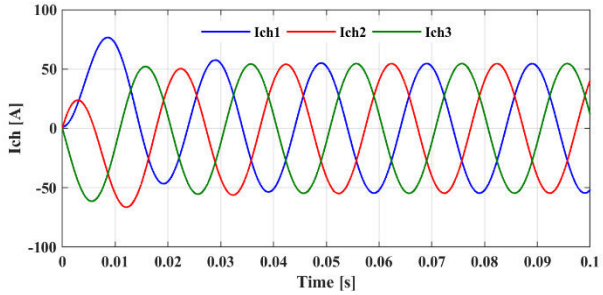


Fig. 3. Load current of the three-phase three-level inverter.

In fig. 3 and 4, the evolution of the output's voltage and current of the three-phase three-level inverter is represented, where the three phases exhibit a 120-degree phase shift, with each phase having three voltage levels.

IV. The Impact of Open-Circuit and Short-Circuit Faults in Three-Phase Three Level Inverter.

Under normal (fault-free) operating conditions, as previously described, the output currents and voltages achieve their reference values. However, when a fault occurs in the IGBTs of the multicellular converter, the system performance deteriorates, resulting in distortions in the output voltages and currents, as well as fluctuations in the capacitor voltages. The primary objective of fault diagnosis is to promptly detect and identify various types of failures at their early stages, enabling timely shutdown and planning of maintenance actions [13].

In this section, we will investigate the impacts of faults applied to the three phase three level inverter. More precisely, we will examine the effects of an open-circuit faults and short-circuit faults on the inverter dynamic.

A. Case one: Open-circuit faults

The following figures illustrate the dynamics of the output's voltage and current as well as the floating capacitor voltage of the three-phase three-level inverter under faults conditions, specifically an open-circuit faults in the IGBTs constituting the inverter, at time t=0.05s.

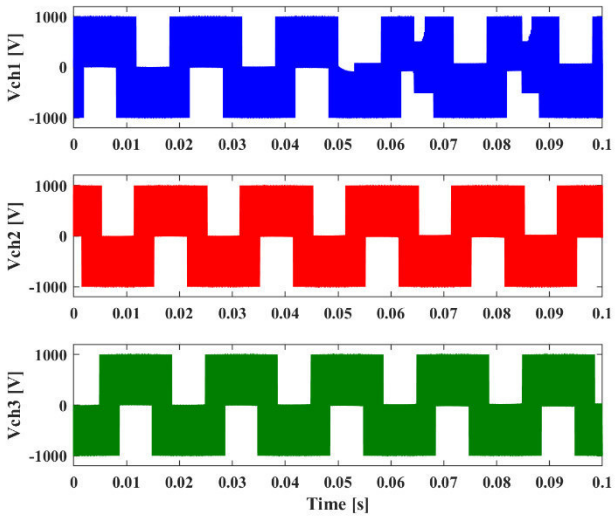


Fig. 4. The variations of Three-phase voltages during the appearance of an OCF on the first phase of the inverter.

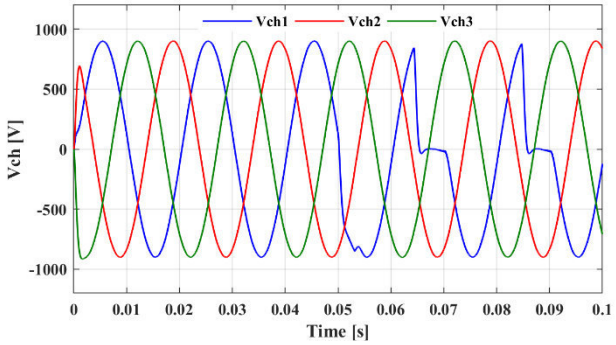


Fig. 5. The variations of Three-phase voltages filtered during the appearance of an OCF on the first phase of the inverter.

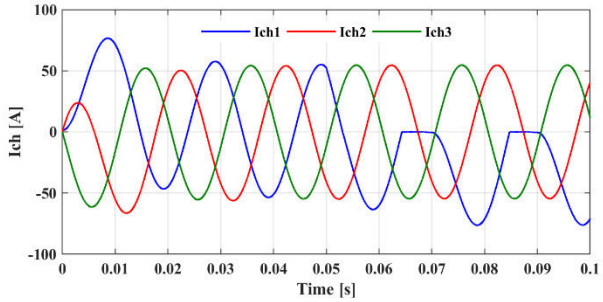


Fig. 6. The variations of Three-phase current filtered during the appearance of an OCF on the first phase of the inverter.

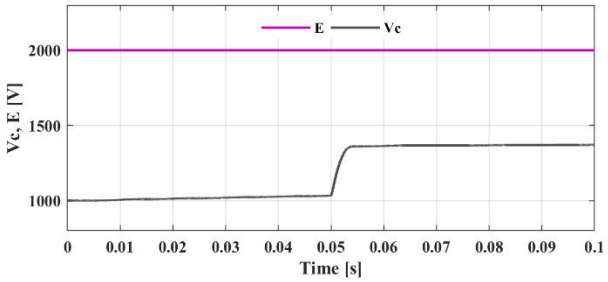


Fig. 7. Evolution of the floating capacitor voltage during the appearance of an OCF on the first phase of the inverter.

The previous figures presents the dynamics of the output voltages and the currents of the faut phase for the case of an open-circuit fault on the IGBT of the upper stage at the time t=0.05 sec, we can see that for an open-circuit fault applied to the IGBT of the upper stage of the 1st phase of the inverter, the degraded regime is manifested by the cancellation of the positive alternation of the current and significant deformations are recorded at the level of the load voltage of the 1st phase of the studied inverter; on the other hand, if the fault applied is in the IGBT of the lower stage, the measured current is canceled on the negative alternation with deformations of the load voltage. we also note that the 1st capacitor voltage corresponding to the 1st cell is distorted when the fault appears.

In order to see the fault impact of an open circuit on the three-phase three level inverter, open-circuit faults were applied to the various IGBT transistors constituting the inverter, we note the same results as above, when open-circuit faults appear on an IGBT of the upper or lower stage of one phase of the three-phase three level inverter, the degraded regime is manifested by a deformation which appears on the load voltage and a cancellation of a positive or negative alternation of the load current of the faulty phase according to the fault stage as presented previously.

B. Case two: Short-circuit faults

The applied fault in this case is a short-circuit fault. The Subsequent figures present the dynamics of the output's voltage and the floating capacitor voltage of the three-phase three-level inverter during the occurrence of a short-circuit fault on the IGBT of the upper stage of the 1st phase of the inverter, at time t=0.05s.

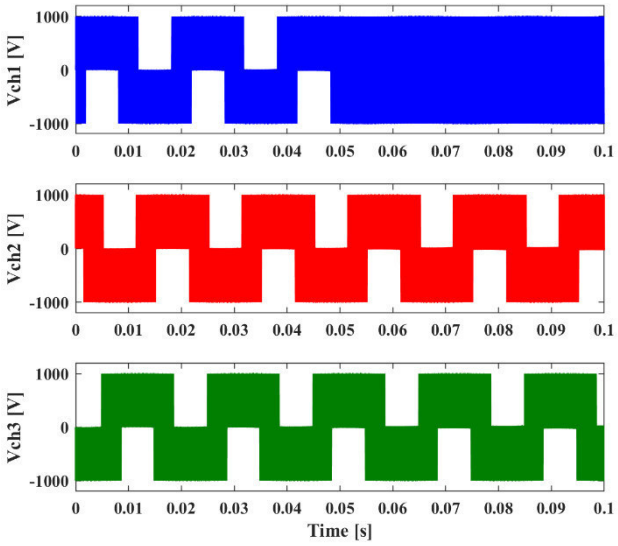


Fig. 8. The variations of Three-phase voltages during the appearance of an SCF on the first phase of the inverter.

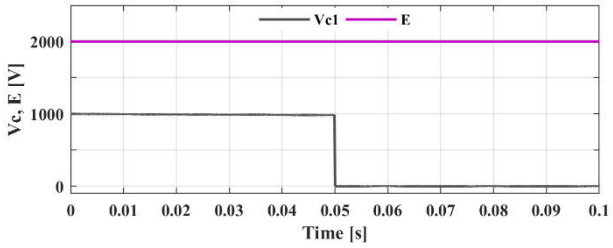


Fig. 9. Evolution of the floating capacitor voltage during the appearance of an SCF on the first phase of the inverter.

The preceding figures shows the dynamic of the output voltages of the faulty phase for the case of a short-circuit fault on the IGBT of the upper stage at the instant t=0.05s. We note that when a short-circuit fault appears to the IGBT of the of the 1st phase of the inverter, the degraded regime is manifested by significant deformations in the voltage load of the 1st phase. it is also noted that the voltage of the 1st capacitor corresponding to the 1st cell where the fault appears suddenly drops to zero during the failure. The impact of a short-circuit fault is usually destructive.

To examine the effect of short-circuit faults on a three-phase three level inverter, faults have been applied to the various IGBT transistors constituting the inverter, the same results as above are observed, when a short-circuit fault appears on an IGBT of the upper or lower stage of one phase of the three-phase inverter, the degraded regime is manifested by a deformation which appears on the load voltage of the faulty phase and the voltage of the floating capacitor which corresponding to the faulty cell is abruptly canceled; if two short-circuits appear on one arm of the inverter, the distortion of the load voltage will increase and the floating capacitor voltages abruptly cancel.

V. Conclusion

Multilevel-multicellular inverters are particularly appealing due to their suitability for high-power and high-voltage applications. However, the occurrence of open-circuit and short-circuit faults in these converters can significantly degrade the performance of the conversion system. This highlights the critical importance of accurate and timely fault diagnosis. Effective diagnostic techniques not only ensure the safety and reliability of electrical drives but also minimize downtime and maintenance costs, thereby improving overall system efficiency and lifespan. Consequently, significant research efforts have been devoted to developing robust fault diagnosis methods tailored to multilevel inverters, emphasizing their role in maintaining optimal performance in demanding applications.

Following a successful fault diagnosis, selecting an appropriate fault-tolerant control strategy becomes crucial to maintaining system functionality. Fault-tolerant control plays a pivotal role in ensuring the reliability and robustness of power systems employing multicellular inverters. This involves designing advanced control strategies capable of detecting, isolating, and compensating for faults within the power converter. By enabling the system to continue operating, even in the presence of faults, such strategies help mitigate performance degradation, minimize disruptions, and enhance overall system resilience. These approaches are indispensable in high-power applications where downtime and inefficiencies can have significant technical and economic impacts.

References

[1] K. Benmansour, A. Benalia, M. Djemaï, J. de Leon, “Hybrid control of a multicellular converter,” Nonlinear Analysis: Hybrid Systems, Volume 1, Issue 1, 2007, Pages 16-29, ISSN 1751-570X.

[2] P. Djondiné “ Overview of Control Techniques for Multicellular Converter,” Journal of Engineering Sciences, 2018, 5 (1), pp.E 10-E 14. 10.21272/jes.2018.5(1). e3.

[3] T.A. Meynard & H. Foch “Multi-Level Choppers for High Voltage” Applications, EPE Journal, (1992) 2:1, 45-50.

[4] R. Bensaid, M. Fadel “Sliding mode observer for multicell converters”, IFAC Nonlinear Control Systems, SI. Petersburg, Russia, 2001.

[5] S. Meradi, K.Benmansour, K.Herizi, M.Tadjine, M.S. Boucherit “Sliding mode and faults tolerant control for multicell converter four quadrants”, Electric Power Systems Research 95 (2013) 128-139.

[6] O. Benzineb, F.Taibi , T.M. Laleg-Kirati, M.S. Boucherit, M.Tadjine, “Control and faults diagnosis based sliding mode observer of a multicellular converter: Hybrid Approach” Journal of ELECTRICAL ENGINEERING. VOL.64, NO. 1,2013,20-30.

[7] K. Taniguchi, Y. Ogino and H. Irie, “PWM technique for power MOSFET inverter,” in IEEE Transactions on Power Electronics, vol. 3, no. 3, pp. 328-334, July 1988.

[8] H. Komurcugil, S. Biricik, S. Bayhan and Z. Zhang, “Sliding Mode Control: Overview of Its Applications in Power Converters,” in IEEE Industrial Electronics Magazine, vol. 15, no. 1, pp. 40-49, March 2021.

[9] A. Aibeche, M. Kidouche, “Sensor fault detection, localization, and system reconfiguration with a sliding mode observer and adaptive threshold of PMSM,” J. Power. Electron, Vol. 16, no. 3, pp. 1012–24, 2016.

[10] F. Khater, A. Aibeche, S A. Fellag, M Z. Doghmane, H. Akroum. “Failure identification and isolation of DC-DC boost converter using a sliding mode controller and adaptive threshold”. Diagnostyka. 2024; 25(3):2024303.

[11] T. A. Meynard, M. Fadel, and N. Aouda " Modeling of Multilevel Converters ", IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL. 44, NO. 3, JUNE 1997.

[12] L. Wu, J. Liu, S. Vazquez and S. K. Mazumder, "Sliding Mode Control in Power Converters and Drives: A Review," in IEEE/CAA Journal of Automatica Sinica, vol. 9, no. 3, pp. 392-406, March 2022.

[13] F. Rojas, C. Jerez, C.M. Hackl, O, Kalmbach, J, Pereda and J. Lillo. “Faults in Modular Multilevel Cascade Converters–Part II: Fault Tolerance, Fault Detection and Diagnosis, and System Reconfiguration.” in IEEE Open Journal of the Industrial Electronics Society, vol. 3, pp. 594-614, 2022.

Design and Implementation of a Automatic Multi-Function System for Car

Ibtissem Adoui
Department of Electronics, Faculty of Sciences and Technology, University Mohammed Elbachir Elibrahimi, Bordj Bou Arreridj, El Anasseur 34625.Algeria.
Laboratory of Materials and Electronic Systems LMSE
ibtissem.adoui@univ-bba.dz

Akram Maouchi
Department of Electronics, Faculty of Sciences and Technology, University Mohammed Elbachir Elibrahimi, Bordj Bou Arreridj, El Anasseur 34625, Algeria

Abdelmounim Bouadi
Department of Electronics, Faculty of Sciences and Technology, University Mohammed Elbachir Elibrahimi, Bordj Bou Arreridj, El Anasseur 34625, Algeria

Abstract— **In a world where car security and safety are growing concerns, we introduce an innovative solution—a smart and automatic system for cars that integrates cutting-edge technology, including the Arduino Nano, GPS module, GSM module, and MPU6050 sensor. This system represents a significant advancement in car tracking, accident detection, and theft prevention. The research is structured into three compelling sections that guide the reader through an exploration of automatic systems to the practical implementation of this device. The results underscore the substantial potential of this automatic system in enhancing car security and safety, providing features such as real-time location tracking, immediate accident alerts, and anti-theft measures. This paper sets the stage for future developments in the exciting field of automatic systems for cars.**

Keywords— Automatic system car, Smart system car · MPU6050 Sensor · GSM Module · GPS Module · Arduino Nano.

I. INTRODUCTION

In today's rapidly evolving technological landscape, the integration of media technologies such as smart phones or navigation systems in automobiles has become increasingly prevalent [1]. Vehicles are no longer mere modes of transportation; they are transforming into intelligent machines that have the potential to revolutionize our driving experiences. The automatic car systems refer to the integration of advanced technologies and features and is a prosthesis that assists the driver, enhancing the original function of the car, and provides a new place environment to the driver by composing a hybrid space where information space and actual space are fused [1]. In the literature many researchers used different smart and automatic car system. In [2] the authors proposed a vehicular network as a way to detect dangerous approaches among vehicles. In [3], [4] a Smart Security System for Vehicles using Internet of Things (IoT) is proposed. Bhavana Patil et al proposed a Smart car Monitoring System using ARDUINO[5]. In [6], [7] [8] and [9] a different smart car parking system is proposed and implemented. Algeria has a relatively high rate of accidents, according to Arab News. In 2021 the country's authorities recorded more than 7,000 traffic accidents which killed at least 2,643 people and left in excess of 11,000 injured as well as more than 5000 recorded private car theft cases in recent years. This motivated us to build Automatic and Smart Multi-Function device to track car location, send alerts in case of theft and accidents, and remotely turn off the car. Pursuing this project will allow us to enhance security by enabling real-time monitoring of the vehicle's location, increasing the chances of recovery and minimizing losses in case of theft. Prompt theft detection is facilitated through motion sensors, ensuring immediate notifications and enabling rapid response to prevent further damage or loss. The hardware and software requirements for constructing an efficient automatic tracking system, incorporating components such as Arduino Nano, NEO-8M GPS module, SIM800A GSM module, and relay control, are thoroughly investigated. The paper culminates with a practical evaluation of the system's capabilities, encompassing functions like remote car control and safety features such as accident detection and anti-theft alerts.

II. THE AUTOMATIC AND SMART MULTI-FUNCTION SYSTEM FOR CARS

A. Definition

The revolution of automatic and smart systems for cars has ushered in a new era of auto motive technology, where vehicles have become intelligent and interconnected entities. These systems integrate advanced technologies such as connectivity, artificial intelligence, and sensor systems to enhance the overall driving experience. Imagine a car that seamlessly connects to your smartphone, allowing you to re motely control various functions such as locking or unlocking doors, starting or stopping the engine, and adjusting climate settings. With the power of artificial intelligence, your car becomes a personal assistant, understanding your voice commands, providing real-time information, and even learning your preferences to offer a tailored driving experience. Sensors embedded in the car's architecture ensure enhanced safety by continuously monitoring the environment, detecting potential hazards, and assisting with collision avoidance. These smart systems have truly revolutionized the way we interact with our cars, making them more than just modes of transportation but intelligent companions that prioritize convenience, safety, and efficiency [10].

An Examples of various Smart Car tracking devices in the marke is shown in Fig.1.

B. Motivation for projet

The automatic system for cars brings numerous benefits that revolutionize the driving experience. It offers convenience through remote control capabilities, enabling functions such as engine start, door unlocking, and climate adjustment from a distance. Enhanced safety features, including accident detection and emergency notifications, contribute to a safer road environment. The system facilitates efficient vehicle management through real-time location tracking and fleet monitoring. Personalization and user experience are enhanced through artificial intelligence

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

algorithms that learn and adapt to individual preferences. Overall, the smart system for cars provides convenience, safety, efficiency, personalization, and connectivity, transforming the driving experience and shaping the future of the automotive industry.

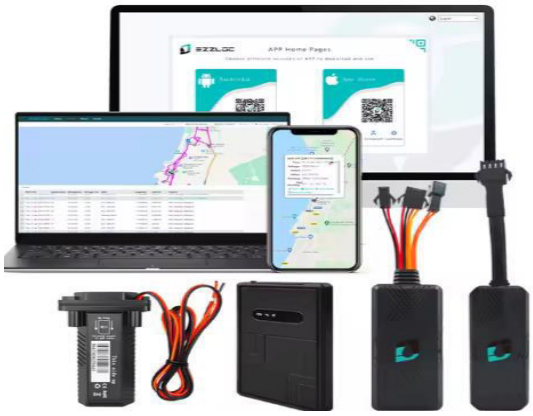


Fig.1: Examples of various Smart Car tracking devices in the marke

III. Study and Conception of the Car Tracker

The primary objective of this automatic tracking system is to provide real-time monitoring of the vehicle's location and ensure enhanced security measures. In case of theft, the system is capable of sending immediate alerts to the user, enabling quick responses and recovery actions. Similarly, in the event of an accident, the system can detect the impact using sensors, trigger emergency and notify relevant authorities or emergency contacts alerts.

A. Methodology for Our Automatic System

The methodology for implementing an Automatic System using components like Arduino Nano, GSM, GPS, and MPU6050 involves initialization of critical components, checking GSM availability, reading sensor values, handling SMS responses, detecting accidents, and continuous testing. The system is initialized with proper communication and functionality. The avail ability of the GSM module is verified for SMS communication. Sensor values from MPU6050 are retrieved for analysis. Incoming SMS messages are analyzed to determine the appropriate response, such as controlling devices or sending location information. An accident detection algorithm continuously monitors sensor data for sudden changes and triggers alert notifications. The system undergoes continuous testing and waits for incoming commands and sensor data. This methodology enables effective car location tracking, theft and accident alerts, and remote control capabilities.

B. Proposed System

Here's the explanation of the circuit diagram we used involving Arduino Nano, GPS module, GSM module and MPU-6050 accelerometer and gyroscope sensor and LM2596 switcher:

1) Arduino Nano: The Arduino Nano serves as the main microcontroller in the circuit. It provides the processing power and controls the various modules.

2) GPS Module: The GPS module communicates with the Arduino Nano via serial communication (UART) to receive location data from GPS satellites. It typically consists of a GPS receiver, antenna, and necessary circuitry.

3) GSM Module: The GSM module allows the Arduino Nano to communicate over the cellular network. It enables sending and receiving SMS messages, as well as transmitting data. The GSM module usually includes a SIM card slot and the necessary communication circuitry.

4) MPU-6050: The MPU-6050 module integrates a 3-axis accelerometer and a 3-axis gyroscope. It measures acceleration and angular velocity, providing information about the vehicle's movement and orientation. It communicates with the Arduino Nano via I2C (Inter-Integrated Circuit) protocol.

5) LM2596 Switcher: The LM2596 is a step-down DC-DC switching regulator. It regulates the voltage input from the power source to a lower and stable voltage required by the circuit components. It ensures a consistent power supply for the Arduino Nano, GPS module, GSM module, and MPU-6050.

Fig.2 shows the block diagram of our project and the different connections between the different components.

Fig.3 shows the circuit implemented by FRITZING for the realization of our global System.

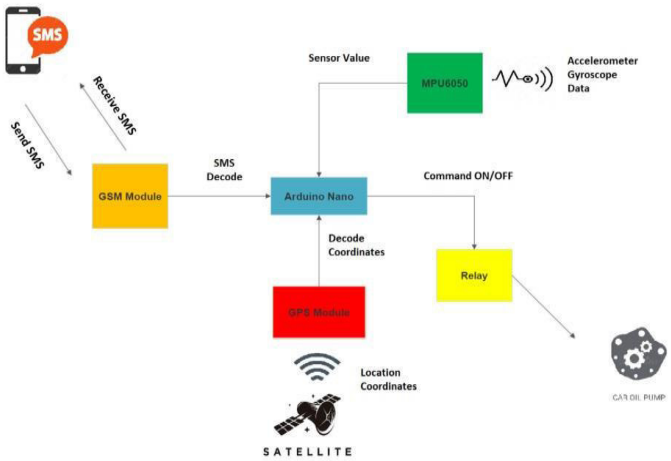


Fig.2: Circuit diagram for AMFS for cars

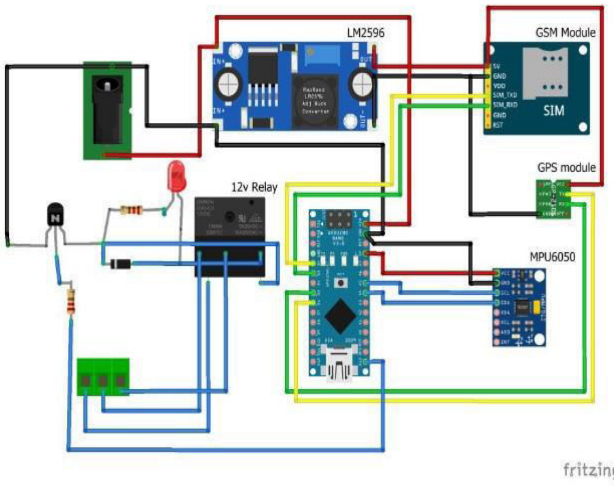


Fig.3: Circuit design AMFS for cars

Where the circuit connections can be summarized as follows:

- The GPS module connects to the Arduino Nano's serial pins (RX and TX) for data communication.
- The GSM module connects to the Arduino Nano's serial pins for communication, and additional pins may be used for control signals (e.g., power on/off).
- The MPU-6050 connects to the Arduino Nano via the I2C interface, typically using the SDA (data line) and SCL (clock line) pins.
- The LM2596 switcher connects to the power source and provides regulated power to the Arduino Nano, GPS module, GSM module, and MPU-6050.

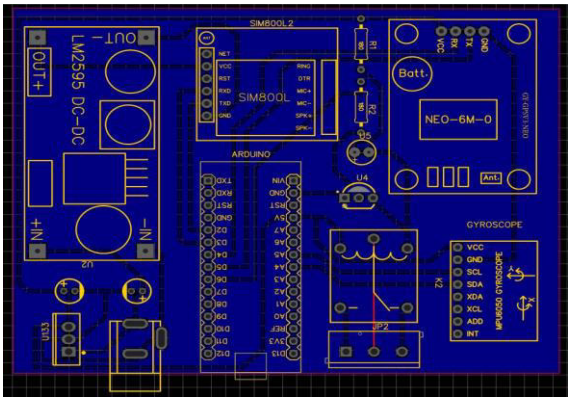
IV. Simulation and Practical Realization

A. *Design Methodology*

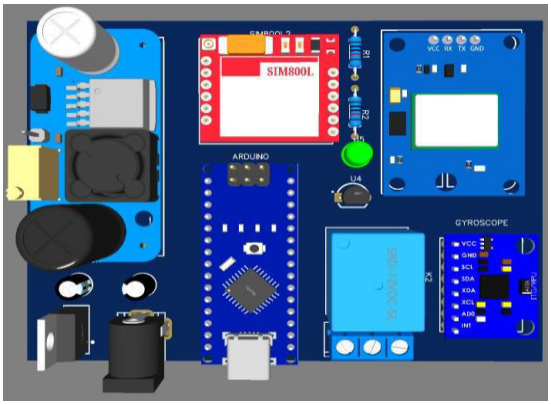
To realize our Prototype we follow several steps. These steps are grouped into five parts:

- In the First part of the process, the proposed circuit is implemented using easyEDA software. This involves creating a schematic circuit, designing the PCB layout, defining the PCB parameters and generating the Gerber files.
- In the second part, the implemented circuit is fabricated. This part of the PCB fabrication process involves several steps.
- The part 3 presents the design of the enclosure (packaging).
- In this fourth part, we utilize the Arduino IDE software for programming our Arduino Nano boards with the others components.
- The last part present the test of our Automatic Multi-Function System. By following these steps, we can effectively realize our Prototype of the Automatic Multi-Function System (AMFS) for car. In the upcoming section, we will delve into the details of these five steps.

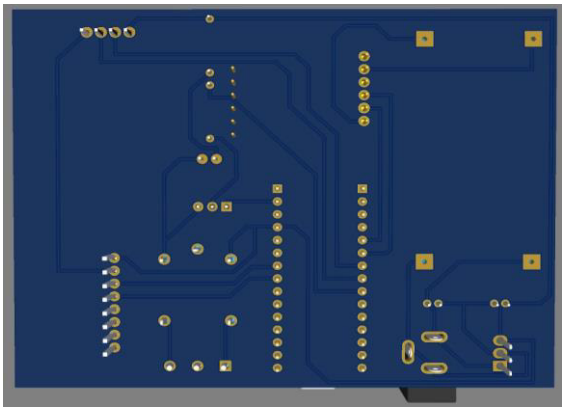
1) Circuit Implementation: Initially, we used the open-source EasyEDA software to design and print the circuit of our prototype. Figure 4 displays the PCB simulation and the 3D visualization for both the top and bottom sides.



(a) PCB Circuit



(b) 3D Model (TOP view)



(c) 3D Model (Bottom view)

Fig.4: Circuit Implementation with EasyEDA

2) PCB fabrication: This part of the PCB fabrication process involves several steps, including Epoxy preparation, application of photoresist, exposure and development, etching, drilling, plating, testing and final inspection. The prototype consisting of all the components integrated is designed as shown in fig.5.

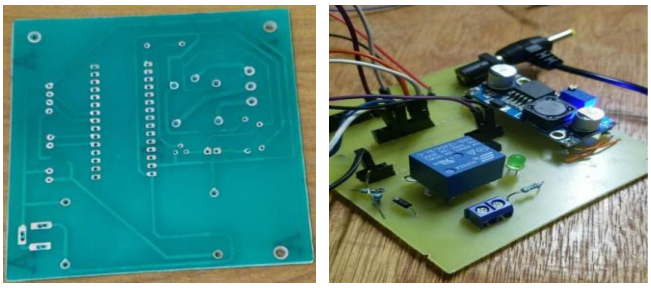


Fig.5: PCB fabrication

3) Designing the Enclosure and Prototyping: The first step in 3D printing a box for the car tracker is to design the enclosure. This involves creating a 3D model using computer-aided design (CAD) software (fig.6). Consider the dimensions and shapes of the components, taking into account their positioning and connections. Ensure that the design allows for proper ventilation and accessibility to buttons, ports, and indicators.

Once the enclosure is printed and post-processed, it's time to assemble the car tracker components inside the box. Carefully position the GPS module, GSM module, Arduino Nano, MPU-6050, and other necessary components within the

enclosure, ensuring proper alignment and securing them in place. Make appropriate openings or cutouts in the enclosure for cables, connectors, and ventilation. A prototype consisting of all the components integrated is designed as shown in Fig.7 below.

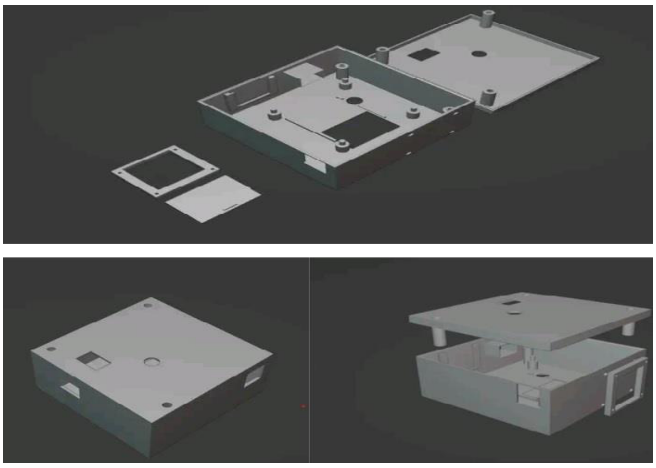


Fig.6: 3D design of the Enclosure using Blender.



Fig.7: Final prototype after assembling the device.

4) Programming Arduino Nano: The following diagram of fig.8 describe the flowcharts that illustrate the algorithms used. The program begins by initializing the critical components, such as the GSM, GPS and the sensor MPU6050. After that, the program checks to see if the GSM module is available, indicating that it is ready to send and receive SMS messages. Next, the system reads values from the sensor MPU6050 to prepare for further action. Moving on to the core of the program, when an SMS is received, the appropriate response is taken:

a) "SMS = ON/OFF"– The Arduino sets the 13th pin of the relay to high or low accordingly.

b) "SMS = location"– The Arduino attempts to obtain the location coordinates and then sends the information to the phone.

c) "SMS = secure"– The system utilises the MPU6050 data to send an alert SMS to the user if motion is detected.

The last function is to read the MPU6050 value by applying the accident detection algorithm. The system then sends an SMS to the user indicating an accident has been detected, along with the relevant location. Throughout, the program continuously tests and waits.

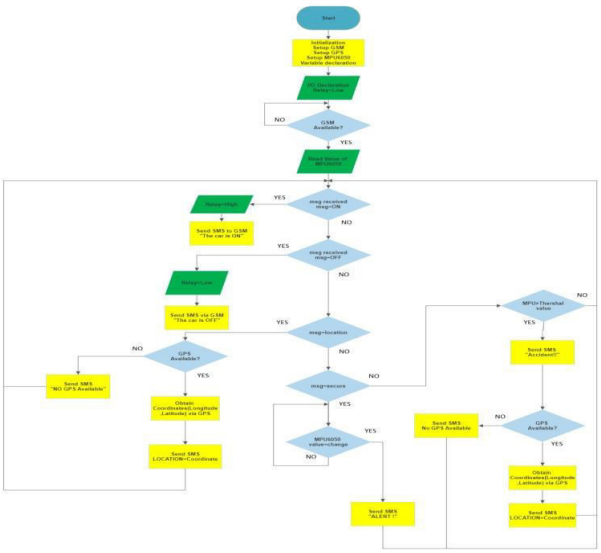
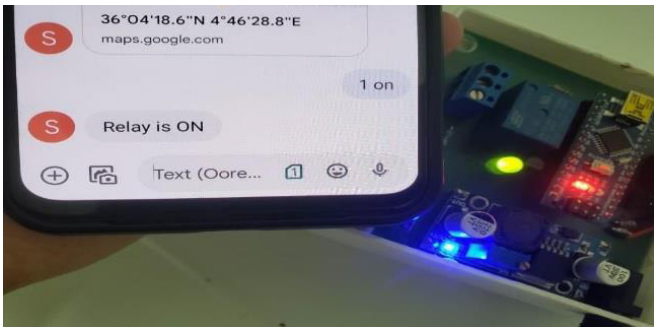


Fig.8: Flowchart explaining the program

5) Functionality result and Testing

a) Test1: In the initial test, our objective is to control the car's power-on function using a relay activated by SMS. By connecting the oil pump to a relay, we can effectively isolate the pump's power circuit from the control circuit. The switch which is a command by the GSM to control the relay, allows us to activate or deactivate the pump with ease. In Fig.9a we observe when we send an SMS "1 on" the relay turns ON(the car starts running) and we receive an SMS confirming that it has turned on. Additionally, the green LED lights up when the relay is on. In contrast, in Fig.9b, when we send an SMS "1 off" the relay turns OFF(the car stops running) and we receive an SMS confirming that it has turned OFF. Moreover, the green Led turns OFF when the Relay is OFF.



(a)Relay turning ON-SMS 1 ON



(b) Relay turning OFF-SMS 1 OFF

Fig.9: Control the car's power-on function using a relay activated by SMS

b) Test2: In the second test, our goal is to determine the car's location by sending an SMS with the keyword "location". Upon sending the message, we receive a link indicating the current location of the car. Fig.10 shows the results of the Test.

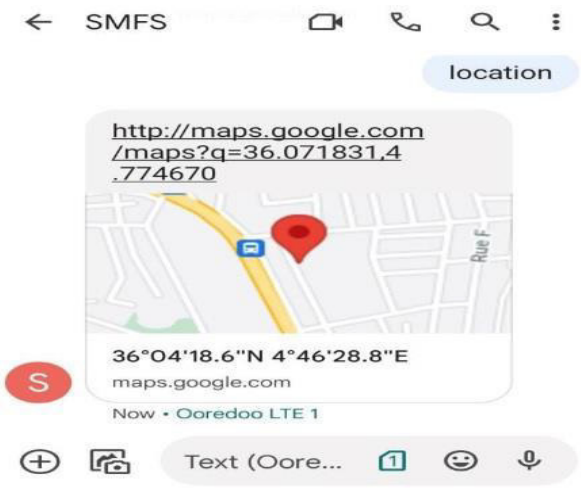
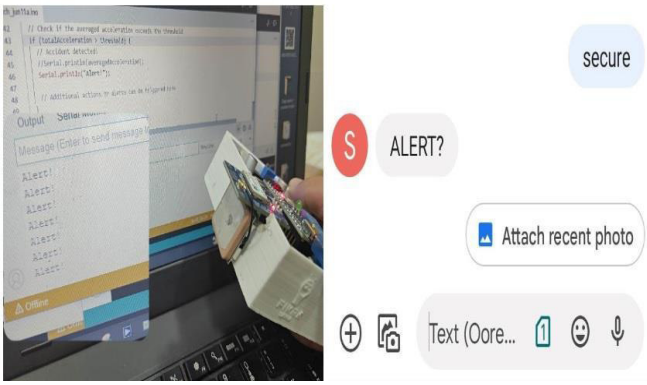


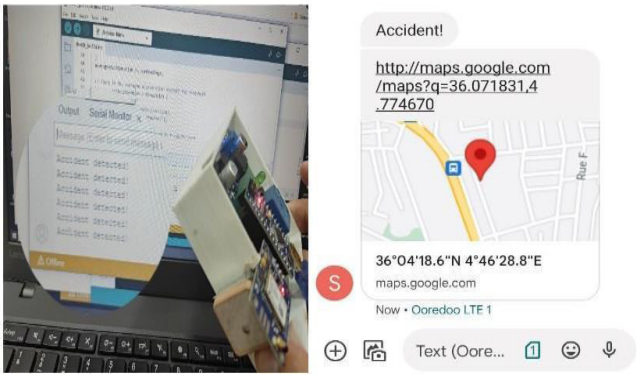
Fig.10: The current location of the car

c) Test3: The last test involves reading the MPU6050 values to apply the accident detection algorithm. Upon detecting an accident, the system sends an SMS to the user (or relevant authorities or emergency contacts) indicating the occurrence and providing the relevant location information. Here the car accident detection system integrates accelerometer-gyroscope (MPU6050) sensors, to monitor vehicle dynamics and detect potential accidents. Accelerometers measure changes in g-forces, offering crucial data on acceleration and deceleration. Common threshold ranges for accidents are categorized as follows: no accident (0-4g), mild accident (4-20g), medium accident (20-40g), and severe accident (>40g).

In our prototype, the system sends an "alert?" SMS to the phone when there is a change in the MPU6050 value . In other scenarios, when the total ac celeration value of the MPU6050 exceeds the threshold, the system sends an "Accident detected" SMS to the phone. To verify the system's functionality, we have lowered the threshold value to determine if the accident detection SMS is triggered. Fig.11 displays the results of the Car Accident Detection Test.



(a) "Alert" SMS



(b) "Accident detected" SMS

Fig.11: The car accident detection system Test

V. Conclusion

In conclusion, we explore the development of an Automatic Multi-Function system for cars that incorporates tracking, alerting, and remote control functionalities. By utilizing components such as Arduino Nano, GPS module, GSM module, and MPU6050, the Automatic system enables precise location tracking, real-time alerting

In case of accidents and thefts, and remote control capabilities. The implementation of the system addresses challenges and presents innovative features that enhance car security and user experience. The findings emphasize the significance of automatic systems in revolutionizing the automotive industry. Through technology and innovation, cars can become intelligent, connected, and responsive, paving the way for a safer and more efficient transportation ecosystem. As research and advancements continue in this field, further refinements and improvements to the automatic system can be expected, enhancing its functionality, usability, and overall appeal to drivers. The future holds immense possibilities for smart systems, bringing us closer to a smarter, more connected, and sustain able transportation landscape. The project also focuses on improved safety by integrating accident detection sensors, which provide instant alerts in case of collisions or changes in vehicle orientation. This allows for swift action, notifying emergency services or concerned individuals for prompt assistance. The continuous monitoring and real time alerts provide peace of mind, allowing you to stay informed about your vehicle's status regardless of your location.

References

[1] So-Eun Lee, Seongcheol Kim: The Influence of Smart Car Technologies on Drivers' Perceived Control and Attachment . 14th Asia-Pacific Regional Conference of the International Telecommunications Society, pp.1-28.(2017)

[2] Pravinraj. R et al:Smart System in Vehicles. International Journal of Engineering Research and Technology (IJERT), pp.2278-0181.(2018)

[3] Mithileysh .S et al:Smart Security System for Vehicles using Internet of Things (IoT). Second International Conference on Green Computing and Internet of Things (ICGCIoT), pp,430-435, 978-1-5386-5657-0/18, IEEE explore.(2018).

[4] Alisher .S.I:SMART CAR SECURITY SYSTEM BASED ARDUINO AND GSM. Science and Education" Scientific Journal 2(2), pp.99-102 (2021).

[5] Vedant.C et al:Smart Car Parking Using Arduino Microcontroller. International Journal of New Technology and Research (IJNTR) 3(6), pp. 2454-4116 (2017).

[6] Hemant.C et al:Advanced CAR Parking System using Arduino. International Con ference on Advanced Computing and

Communication Systems (ICACCS), 978-1 5090-4559-4, IEEE Explore (2017).

[7] Bhavana.P et al:SMART CARMONITORING SYSTEM USING ARDUINO.Inter national Research Journal of Engineering and Technology (IRJET) 5(3), pp. 2395 0072 (2018).

[8] Uzair U.R et al:Implementation of Smart Car Parking System using Arduino. Jour nal of Information and Communication Technology-JICT 13(2), pp.5-9.

[9] Shruthi.M et al:IoT Based Smart Car Parking System. IJSART 5(1), pp. 2395-1052 (2019).

[10] Ye.X: Research on the Development Trend of Intelligent Vehicle. International Con ference on Applied Mathematics, Modelling and Statistics Application (AMMSA 2017) (pp. 392-395). Atlantis Press.

Diagnosis of electrical faults in an asynchronous machine using the sliding mode observer.

1st AMMAOUI YOUCEF
Université de Bejaia, Faculté de Technologie, Laboratoire de Génie Electrique de Bejaia
06000 Bejaia, Algérie
youcef.ammaoui@univ-bejaia.dz

2nd FARID TAFININE
Université de Bejaia, Faculté de Technologie, Laboratoire de Génie Electrique de Bejaia
06000 Bejaia, Algérie
farid.tafinine@univ-bejaia.dz

***Abstract*— In this article, we will diagnose the electricals faults (short-circuit faults in the stator windings and rotor bar breakage fault) of the asynchronous machine. We have used the sliding mode observer to visualize the variation of the machine's characteristics (stator currents, rotor currents, flux, and rotational speed) with and without faults. To validate the results, we have employed spectral analysis (Fast Fourier Transform FFT).**

***Keywords*— Asynchronous machine, diagnosis, bar fault, short-circuit fault, Sliding Mode Observer.**

I. INTRODUCTION

Asynchronous motors (ASM) are widely used in industry due to their simplicity, low cost, reliability, and ability to operate in harsh conditions. They are suitable for a wide range of industrial applications and provide cost-effective solutions for robust and reliable motorization needs.

Significant effort has been made in modeling asynchronous machines in the presence of various faults, such as the breakage of one or more consecutive bars and/or a section of the short-circuit ring, short-circuiting between turns in the windings, and different types of eccentricities [1].

The monitoring and diagnosis of faults in this machine have become essential due to the increasing demands placed by operators. These faults are primarily caused by winding overheating, voltage drops, component wear, surges, and the duration of use, among others.

To detect anomalies in asynchronous machines, several solutions have been proposed by researchers to minimize damage and diagnose the motor in real-time. Among these, there are signal-based diagnostic methods (spectral analysis, artificial intelligence, fuzzy logic, etc.). For instance, the method proposed by [2][3] considered using spectral analysis to solve the set of equations with the cage in the form of special windings. To do this, they used the machine's electrical equations, which include parameters such as stator and rotor resistances, self and mutual inductances. As a result, reliability depends on the accuracy of identifying these parameters. Additionally, model-based diagnostic methods (parity space, parametric identification, state observers) have been explored, such as the method used in [4][5], where they studied flux observers with sliding modes and high gains, which are capable of providing high-quality estimates of flux and current across the entire operating range, as well as effective discretization methods for implementation on test benches.

In this article, we have addressed electrical faults (short circuits in the stator windings and rotor bar breakages), which can occur during the operation of the machine. Initially, we modeled the motor in its healthy state, then in the presence of electrical faults. To monitor the stator currents and rotational speed, we used the Sliding Mode Observer. When comparing the evolution of the machine's characteristics with those of the observer, we found a small error that can be neglected. To localize the faults, we used the Fast Fourier Transform.

II. MODLING OF THE ASYNCHRONOUS MACHINE WITHOUT FAULTS

A. *Modling in the thri-phase reference frame*

The objective is to model an equivalent fictitious machine in which the stator and rotor are always composed of three identical phases carrying three-phase currents, based on several simplifying assumptions (balanced three-phase machine, linearity of the magnetic circuit, sinusoidal distribution of the magnetic field in the airgap) [6].

We define the vectors of stator voltages and currents as U_s and I_s, respectively, as well as the vector of rotor currents I_r for the three-phase windings on the stator and rotor:

$$U_s = \begin{bmatrix} U_a \\ U_b \\ U_c \end{bmatrix}, I_s = \begin{bmatrix} I_{sa} \\ I_{sb} \\ I_{sc} \end{bmatrix}, I_r = \begin{bmatrix} I_{ra} \\ I_{rb} \\ I_{rc} \end{bmatrix} \quad (1)$$

The voltage and flux equations of the asynchronous machine are then written as:

$$U_s = [R_s].[I_s] + \frac{d}{dt}\emptyset_s \quad (2)$$

$$0 = [R_r].[I_r] + \frac{d}{dt}\emptyset_r \quad (3)$$

$$\emptyset_s = [L_s].[I_s] + [M_{sr}].[I_r] \quad (4)$$

$$\emptyset_r = [L_r].[I_r] + [M_{sr}].[I_s] \quad (5)$$

Or :

$$[R_s] = \begin{bmatrix} R_{sa} & 0 & 0 \\ 0 & R_{sb} & 0 \\ 0 & 0 & R_{sc} \end{bmatrix} \qquad [R_r] = \begin{bmatrix} R_r & 0 & 0 \\ 0 & R_r & 0 \\ 0 & 0 & R_r \end{bmatrix}$$

$$[L_s] = \begin{bmatrix} L_{psa} + L_{fsa} & -\frac{L_{sab}}{2} & -\frac{L_{sac}}{2} \\ -\frac{L_{sab}}{2} & L_{psb} + L_{fsb} & -\frac{L_{sbc}}{2} \\ -\frac{L_{sca}}{2} & -\frac{L_{scb}}{2} & L_{psc} + L_{fsc} \end{bmatrix}$$

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

$$[L_r] = \begin{bmatrix} L_{pra} + L_{fra} & -\frac{L_{rab}}{2} & -\frac{L_{rac}}{2} \\ -\frac{L_{rab}}{2} & L_{prb} + L_{frb} & -\frac{L_{rbc}}{2} \\ -\frac{L_{rca}}{2} & -\frac{L_{rcb}}{2} & L_{prc} + L_{frc} \end{bmatrix}$$

$[M_{sr}]$

$$= \begin{bmatrix} M_{s_a r_a}\cos(\theta) & M_{s_a r_b}\cos\left(\theta + \frac{2\pi}{3}\right) & M_{s_a r_c}\cos\left(\theta - \frac{2\pi}{2}\right) \\ M_{s_b r_a}\cos\left(\theta - \frac{2\pi}{2}\right) & M_{s_b r_b}\cos(\theta) & M_{s_b r_c}\cos\left(\theta + \frac{2\pi}{3}\right) \\ M_{s_c r_a}\cos\left(\theta + \frac{2\pi}{3}\right) & M_{s_c r_b}\cos\left(\theta - \frac{2\pi}{2}\right) & M_{s_c r_c}\cos(\theta) \end{bmatrix}$$

$$[M_{rs}] = [M_{sr}]^T$$

Such as:

R_{sx}, R_{rx} : Intrinsic resistance of a stator and rotor phase.

L_{psx} et L_{fsx}: Main inductance and stator leakage inductance.

L_{prx} et L_{frx} Main inductance and rotor leakage inductance.

L_{sxy} : Mutual inductance between two stator phases.

L_{rxy} : Mutual inductance between two rotor phases.

$M_{s_s r_y}$: Mutual inductance between a stator phase x and a rotor phase y.

$\theta = \theta_{mécanique}$: Electrical angle of the rotor position.

P : Number of pole pairs.

The set of assumptions mentioned above allows us to write:

$$R_{sx} = R_s \text{ , } R_{rx} = R_r \text{ , } M_{s_s r_y} = M_{sr} \text{ , } L_{psx} = L_{prx} = L_{sxy} = L_{rxy} = L_p$$

B. Modeling in the two-phase reference frame.

To simplify the study, we will first establish a model in which the quantities are expressed in the reference frame attached to the stator. Thus, the equivalent winding of the three stator phases is formed by two windings with direct axes α_s and quadrature axes β_s. The direct axis α_s is aligned with the axis of the first stator phase α_s. Similarly, in the rotor, two windings α_r and β_r replace the equivalent three-phase windings. We then define the electrical quantities referred to the stator:

Stator quantities: $X_{\alpha\beta s} = T_{23} X_s$

Rotor quantities: $X_{\alpha\beta r} = P(\theta) T_{23} X_r$

$$T_{23} = \sqrt{\frac{2}{3}} \begin{bmatrix} \cos(0) & \cos\left(\frac{2\pi}{3}\right) & \cos\left(\frac{4\pi}{3}\right) \\ \sin(0) & \sin\left(\frac{2\pi}{3}\right) & \sin\left(\frac{4\pi}{3}\right) \end{bmatrix}$$

$$T_{32} = {T_{23}}^T$$

$$P(\theta) = \begin{bmatrix} \cos(\theta) & \cos\left(\theta + \frac{\pi}{2}\right) \\ \sin(\theta) & \sin(\theta + \frac{\pi}{2}) \end{bmatrix}$$ θ is the rotation angle.

Then the voltage and flux equations become:

$$U_{\alpha\beta s} = R_s i_{\alpha\beta s} + \frac{d}{dt}\emptyset_{\alpha\beta s} \quad (6)$$

$$0 = R_r i_{\alpha\beta r} + \frac{d}{dt}\emptyset_{\alpha\beta r} - wP(\theta)\emptyset_{\alpha\beta r} \quad (7)$$

$$\emptyset_{\alpha\beta s} = L_s i_{\alpha\beta s} + M_{sr} i_{\alpha\beta r} \quad (8)$$

$$\emptyset_{\alpha\beta r} = L_r i_{\alpha\beta s} \quad (9)$$

1) Mechanical equation:

$$J\frac{d\Omega_r}{dt} = C_{em} - C_r - f_v\Omega_r \quad (10)$$

With:

$\Omega_r = \frac{d\theta}{dt}$

J : The moment of inertia. f_v : The viscous friction torque.

C_r : The opposing torque applied to the machine. C_{em} : The electromagnetic torque.

The expression for the electromagnetic torque is:

$$C_{em} = \frac{pM_{sr}}{L_r}\left(\phi_{\alpha r} i_{\beta s} - \phi_{\beta s} i_{\alpha r}\right) \quad (11)$$

2) State-space model of the asynchronous machine:

We will write the machine's equations in the following matrix form:

$$\begin{cases} \dot{x} = f(x) + B.u \\ y = Cx + D.u \end{cases} \quad (12)$$

With :

$$x = [i_{\alpha s}\ i_{\beta s}\ \phi_{\alpha s}\ \phi_{\beta s}\ \omega_r]^T, u = [V_{\alpha s}\ V_{\beta s}]^T,$$
$$y = [i_{\alpha s}\ i_{\beta s}\ \omega_r]^T$$

$$f(x) = \begin{bmatrix} -\gamma x_1 + a x_3 + K p x_4 x_5 \\ -\gamma x_2 + a x_4 - K p x_3 x_5 \\ b x_1 - c x_3 - p x_4 x_5 \\ b x_2 - c x_4 - p x_3 x_5 \\ d(x_3 x_2 - x_4 x_1) - \frac{C_r}{J} - \end{bmatrix} \quad ; \quad B = \begin{bmatrix} \frac{1}{\sigma L_s} & 0 \\ 0 & \frac{1}{\sigma L_s} \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad ; \quad D = 0$$

$$\sigma = 1 - \frac{M_{sr}^2}{L_s L_r} \ , \gamma = \frac{R_s}{\sigma L_s} + \frac{R_r M_{sr}}{\sigma L_s} \ , Tr = \frac{L_r}{R_r} \ ,$$
$$K = \frac{M_{sr}}{\sigma L_r L_s} \ , \ a = \frac{K}{T_r} \ , b = \frac{M_{sr}}{T_r} \ ,$$
$$c = \frac{1}{T_r} \ , \ d = \frac{pM_{sr}}{JL_r}$$

III. Model of the Asynchronous Machine with Electrical Fault

A. Stator fault of the short-circuit type:

To account for the existence of short-circuited turns in the stator of the asynchronous machine, an additional short-circuited coil is introduced, with the number of turns η_{cc} equal to the number of defective turns in the machine [7] [8]. This short-circuited winding, which is the source of the stationary field relative to the stator, creates a fault condition.

It is necessary to introduce two parameters to define this fault:

The electrical angle, denoted θ_{cc}, representing the short-circuited winding relative to the reference axis of phase a_s. This parameter allows for the localization of the faulty winding and can only take three values $0 , \frac{2\pi}{3} , -\frac{2\pi}{3}$ Corresponding respectively to a short circuit on the phases a_s , b_s or c_s.

The short-circuit ratio η_{cc} is equal to the ratio of the number of short-circuited turns to the total number of turns in a stator phase without fault:

$$\eta_{cc} = \frac{n_{cc}}{n_s}$$

n_{cc} : is the number of short-circuited turns.

n_s : is the number of turns in a healthy phase.

"Fig. 1" illustrates the stator of the machine when a short-circuit fault appears in the stator phase b_s. It can be observed that the fault may present itself as a new winding B_{cc} in the affected phase. whose number of turns is equal to the number of short-circuited turns and the direction is equal to $\frac{2\pi}{3}$

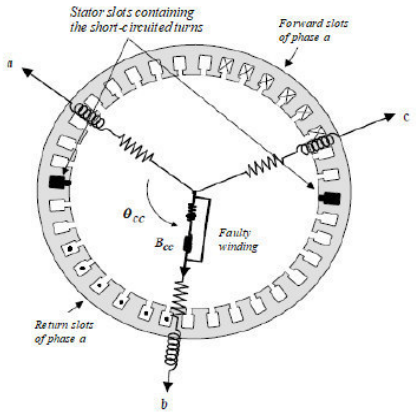


Fig. 1. Short-circuit of turns on phase b of the stator

1) Modeling in the Three-Phase Reference Frame:

We will now develop the voltage and flux equations of the asynchronous machine in the presence of a stator fault of the short-circuit type, taking into account the electrical parameters of the additional coil B_{cc} [6].

$$\begin{aligned} u_s &= [R_s]i_s + \frac{d}{dt}\phi_s \# \\ 0 &= [R_r]i_r + \frac{d}{dt}\phi_r \\ 0 &= [R_{cc}]i_{cc} + \frac{d}{dt}\phi_{cc} \\ \phi_s &= [L_s]i_s + [M_{sr}]i_r + [M_{cc}]i_{cc} \\ \phi_r &= [L_r]i_r + [M_{rs}]i_s + [M_{cc}]i_{cc} \\ \phi_{cc} &= [L_{cc}]i_{cc} + [M_{ccs}]i_s + [M_{ccr}]i_r \end{aligned} \tag{13}$$

We will now define the new parameters that have been introduced compared to the healthy model:

The resistances being directly proportional to the number of turns, the resistance $R_{cc}$of the coil B_{cc} is simply expressed as a function of the resistance $R_s$of the three stator phases without faults [8]:

$$R_{cc} = \eta_{cc} R_s \tag{14}$$

The magnetic circuit traversed by the field lines cutting through the short-circuited coil B_{cc} is identical to the one traversed by the field lines cutting through the faulty stator phase (same reluctance). Thus, we can express the various inductances and mutual inductances of the coil B_{cc} with respect to the stator and rotor phases of the machine [7] [8].

$$L_{cc} = \eta_{cc}^2 (L_p + L_f) \tag{15}$$

$$[M_{scc}] = \eta_{cc} L_p \left[\cos(\theta_{cc}) \ \cos\left(\theta_{cc} - \frac{2\pi}{3}\right) \ \cos\left(\theta_{cc} + \frac{2\pi}{3}\right) \right]$$

$$= \sqrt{\frac{3}{2}} \eta_{cc} L_p [\cos(\theta_{cc}) \ \sin(\theta_{cc})] T_{23} \tag{16}$$

$$[M_{rcc}] = \eta_{cc} L_p \left[\cos(\theta_{cc} - \theta) \cos\left(\theta_{cc} - \theta - \frac{2\pi}{3}\right) \right]$$

$$= \sqrt{\frac{3}{2}} \eta_{cc} L_p [\cos(\theta_{cc}) \ \sin(\theta_{cc})] P(\theta) T_{23} \tag{17}$$

$$[M_{ccs}] = [M_{scc}]^T \quad \text{et} \quad [M_{ccr}] = [M_{rcc}]^T \tag{18}$$

2) Modeling in the Two-Phase Reference Frame:

The short-circuit quantities are localized along a fixed direction in the stator. Their projections onto the α_s and β_s axes allow associating them with stationary vectors relative to the stator [6].

$$i_{\alpha\beta s} = \begin{bmatrix} \cos(\theta_{cc}) \\ \sin(\theta_{cc}) \end{bmatrix} i_{cc} \quad , \quad \phi_{\alpha\beta cc} = \begin{bmatrix} \cos(\theta_{cc}) \\ \sin(\theta_{cc}) \end{bmatrix} \phi_{cc}$$

The voltage equations are the same as those in equations (6) and (7), with the addition of the equation that depends on the short-circuit:

$$0 = R_{cc} i_{\alpha\beta cc} + \frac{d}{dt}\phi_{\alpha\beta cc} \tag{19}$$

The flux equations are written as [7] [8]:

$$\phi_{\alpha\beta s} = L_s i_{\alpha\beta s} + M_{sr} i_{\alpha\beta r} + \sqrt{\frac{2}{3}} M_{sr} \eta_{cc} i_{\alpha\beta cc} \tag{20}$$

$$\phi_{\alpha\beta r} = L_r i_{\alpha\beta r} + M_{sr} i_{\alpha\beta s} + \sqrt{\frac{3}{2}} M_{sr} i_{\alpha\beta cc} \tag{21}$$

$$\phi_{\alpha\beta cc} = \eta_{cc}^2 L_{cc} Q(\theta_{cc}) i_{\alpha\beta cc} + \sqrt{\frac{3}{2}} \eta_{cc} M_{sr} Q(\theta_{cc}) \left(i_{\alpha\beta s} + i_{\alpha\beta r}\right) \tag{22}$$

With:

$$Q(\theta_{cc}) = \begin{bmatrix} \cos(\theta_{cc}) \\ \sin(\theta_{cc}) \end{bmatrix} [\cos(\theta_{cc}) \quad \sin(\theta_{cc})]$$

3) State-Space Model of the Machine with Short-Circuit Fault:

"Fig. 2" represents the electrical circuit of the asynchronous machine in dynamic regime, taking into account a possible short-circuit fault at any phase of the stator [6].

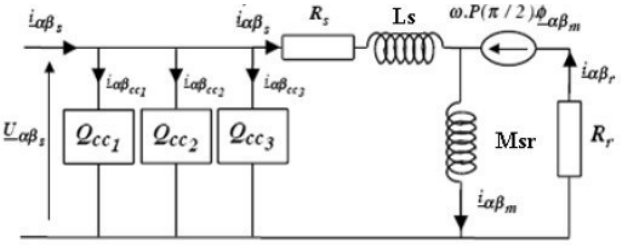


Fig. 2. General Short-Circuit Model in the Park Reference Frame.

The continuous model of the asynchronous machine with a short-circuit fault is then represented in the following form:

$$\begin{cases} \dot{x} = f(x) + Bu \\ y = Cx + Du \end{cases} \tag{23}$$

Where the matrices $f(x)$, B et C are the same as those in equation (12), but only with:

$$D = \sum_{k=1}^{3} \frac{2\eta_{cc_k}}{3R_s} Q\left(\theta_{cc_k}\right) \tag{24}$$

4) Simulation Results

After modeling the asynchronous motor in its normal state and in the presence of a short-circuit fault, the next step is to simulate these models in Matlab. The machine is powered by a balanced 380/220 V, 50 Hz network. We will simulate both cases: the one without a fault, in order to compare it with the one with a fault. A short-circuit of 25 turns at the stator will be introduced at t = 7 seconds. The results of these simulations will then be analyzed and compared.

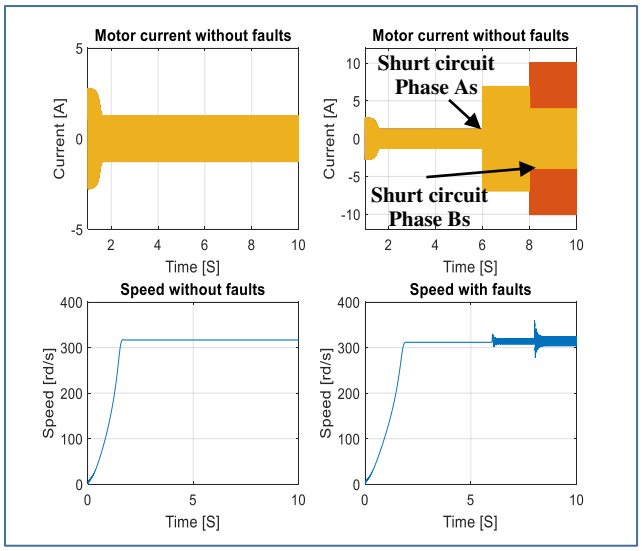


Fig. 3. The evaluation of the stator current and rotational speed in the absence and presence of short-circuit faults in the stator phase.

5) Interpretation of results

Fig. 3 presents the simulation results of an asynchronous motor, both in the absence and presence of a short-circuit fault in the stator windings. It is observed that a fault occurring at t = 6 s leads to an increase in the current of stator phase A as well as vibrations in the rotor speed. Furthermore, the severity of the fault intensifies as the number of short-circuited phases increases.

B. Rotor faults of the bar breakage type:

In this section, we will study the behavior of the asynchronous machine in the presence of faults, such as rotor bar breakage. The squirrel cage of the asynchronous machine is primarily made up of metal conductive bars, connected by short-circuit rings, thus forming the rotor of the machine.

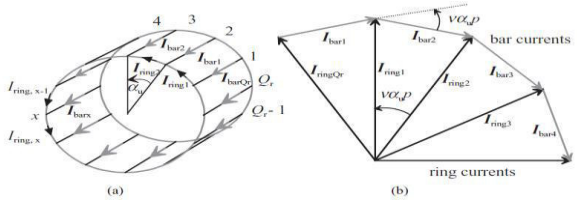


Fig. 4. **(a):** Diagram for a cage winding; **(b)** a sector of a polygon of the phasors of the bar currents and a section of the current phasor diagram [8].

To model the behavior of the machine in the presence of rotor bar breakage, it is sufficient to replace the value of the resistance R_r with [10]:

$$R'_r = R_r(1 - \eta_{bb}) \tag{25}$$

With:

$$\eta_{bb} = \frac{N_b^2 - N_{bs}^2}{N_{bs}^2} \tag{26}$$

$$N_{bs} = N_b - N_{bb}\#$$

N_b : The number of bars, N_{bs} : The number of healthy bars, N_{bb} : The number of broken bars.

C. *State model of the machine with rotor bar faults:*

The continuous model of the asynchronous machine is the same as the model in equation (12), except that the value of the rotor resistance must be changed R_r with R'_r.

1) Résults of the simulation:

We will present the simulation results of the asynchronous machine, both in the absence and in the presence of rotor bar faults (two broken bars), which we obtained using MATLAB software. The machine is powered by a balanced 380/220V, 50Hz power supply.

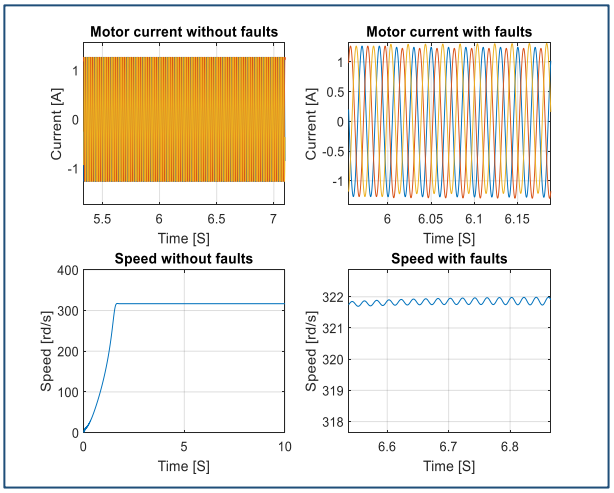


Fig. 5. Stator current variation with and without rotor bar breakage fault.

2) Interpritaion of the results:

According to the simulation results, it is observed that a rotor bar breakage fault causes a slight distortion in the stator currents. Additionally, this fault generates vibrations in the rotational speed. This makes the diagnosis of electrical faults in the asynchronous machine detectable, but relatively difficult to localize.

IV. Use of Observers for Estimating the Characteristics of the Machine

Sliding mode observers are based on the theory of systems with variable structures. This approach is well-suited for uncertain linear dynamic systems. In this section, we will present the synthesis of the sliding mode observer, which is dedicated to motor control and also to the generation of residuals for diagnostics. We will show that the sliding mode observer can accurately reproduce the characteristics of the asynchronous machine.

A. *Synthesis of a Sliding Mode Observer*

The sliding mode observer dedicated to state estimation of the system is given by the following equations [11][12]:

$$\begin{cases} \dot{\hat{x}}_1 = -\gamma x_1 + a\hat{x}_3 + Kp\hat{x}_4 x_5 + V_{s\alpha}/\sigma L_s + \lambda_1^T I_s \\ \dot{\hat{x}}_2 = -\gamma x_2 + a\hat{x}_4 - Kp\hat{x}_3 x_5 + V_{s\beta}/\sigma L_s + \lambda_2^T I_s \\ \dot{\hat{x}}_3 = bx_1 - c\hat{x}_3 - p\hat{x}_4 x_5 + \lambda_3^T I_s \\ \dot{\hat{x}}_4 = bx_2 - c\hat{x}_4 - p\hat{x}_3 x_5 + \lambda_4^T I_s \\ \dot{\hat{x}}_5 = d(\hat{x}_3 x_2 - \hat{x}_4 x_1) - \dfrac{T_r}{J} - \dfrac{f_v}{J} x_5 + q_1(x_5 - \hat{x}_5) + \lambda_5^T I_s \end{cases}$$

With:

$$I_s = [sign(S_1) \;\; sign(S_2)]^T \quad \varepsilon\tau \quad \begin{cases} S_1 = x_1 - \hat{x}_1 \\ S_2 = x_2 - \hat{x}_2 \end{cases}$$

Such that S_1 and S_2 represent the sliding surfaces. The gains q_i and λ_i^T are calculated in a way that ensures the asymptotic convergence of the estimation errors.

They are given by:

$$\begin{bmatrix} \lambda_1^T \\ \lambda_2^T \end{bmatrix} = Q^{-1} \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix} \varepsilon\tau \; Q = \frac{1}{(a^2 + (Kpx_5)^2)} \begin{bmatrix} a & -Kpx_5 \\ Kpx_5 & a \end{bmatrix}$$

$$\begin{bmatrix} \lambda_3^T \\ \lambda_4^T \end{bmatrix} = \left[\begin{pmatrix} -c & -px_5 \\ px_5 & -c \end{pmatrix} + \begin{pmatrix} q_3 & 0 \\ 0 & q_4 \end{pmatrix} \right] \begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix}$$

$$\lambda_5^T = [\lambda_{51} \quad \lambda_{52}]^T \,, \quad \begin{bmatrix} \frac{\lambda_{51}}{\delta_1} & \frac{\lambda_{52}}{\delta_2} \end{bmatrix} = d[x_2 \quad -x_1]$$

Such that the stability of the observer is ensured by the following conditions:

$$\delta_1 \geq \|e_3\|_{max} \,, q_2 \geq \|e_4\|_{max} \,, q_1 > 0 \,, \;\; q_3 > 0 \,, q_4 > 0$$

$$\begin{cases} \dot{e}_3 = -q_3 e_3 \\ \dot{e}_4 = -q_4 e_4 \\ \dot{e}_5 = -q_1 e_5 \end{cases}$$

B. *Results of th simulation:*

After deriving the synthesis of the sliding mode observer, we will simulate it using MATLAB and verify if the estimation results provided by the observer converge toward those of the real asynchronous machine system. The first step will be to simulate the system without faults, then introduce short-circuit faults to observe if the system is sensitive to these faults. Additionally, although the observer can be used for both diagnostic and motor control purposes, we will focus solely on diagnostics in this study.

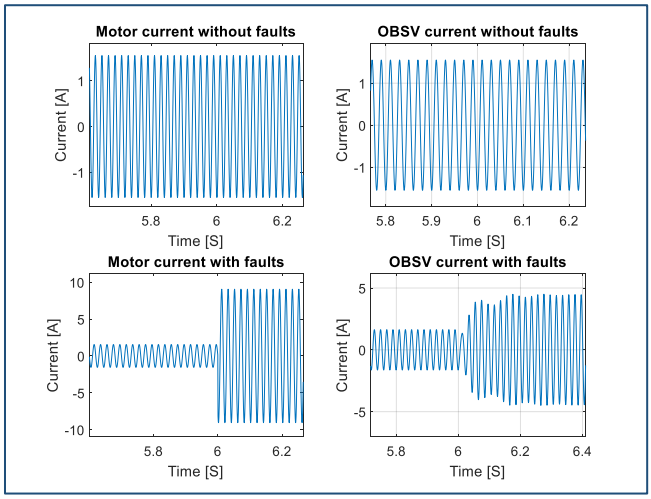


Fig. 6. The comparison between the simulation results of the stator current given by the two-phase model and that of the observer.

1) Interpretation of the results:

It is clearly observed that the current given by the observer follows the system's current, which means that the sliding mode observer model is sensitive to short-circuit faults.

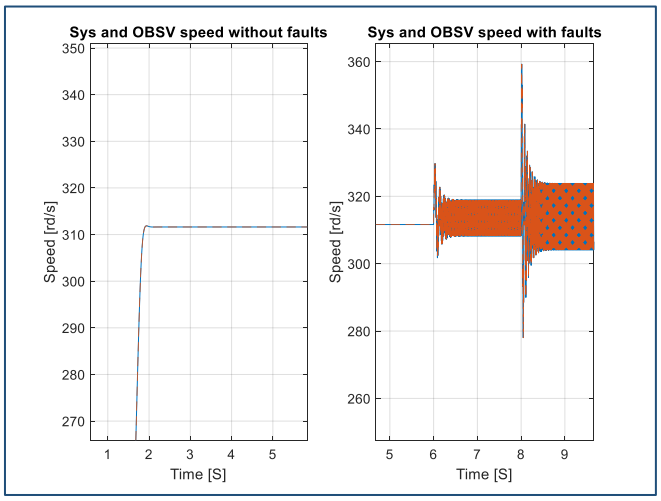


Fig. 7. Comparison between the speed of the model and that of the observer in the presence and absence of faults.

2) Interpretation of the results:

After simulating both the two-phase model and the observer model, it is observed that the speed provided by the observer closely matches the speed of the model.

V. Conclusion

In this article, we studied the phenomenon of electrical short-circuiting and its impact on the asynchronous motor. We analyzed the stator currents and the variation in the motor speed. It was observed that a short-circuit in the stator windings causes a significant distortion in the system's responses, while rotor bar failures do not have a significant impact on the stator current but generate vibrations in the speed.

So far, we can say that the detection of short-circuit faults is relatively easy, while their localization remains a challenge. This has prompted us to explore more effective methods for locating these faults. In this regard, we proposed the use of a sliding mode observer, which is characterized by its robustness, ease of implementation, and fault tolerance. We found that the responses provided by the observer closely follow those of the system, both in the absence and presence of faults. Thus, the generation of residuals could assist us in diagnosing the motor in real-time. Furthermore, the use of spectral analysis methods (FFT) is commonly employed for fault localization.

References

[1] M. E. K. Oumaamar, “Surveillance et diagnostic des défauts rotoriques et mécaniques de la machine asynchrone avec alimentation équilibrée ou déséquilibrée” Thèse de doctorat, Université de Lorraine, Génie Électrique, 2012.

[2] A. Mellakhia, N. Benouzza, A. Bendiabdellah. “Analyse spectrale du courant statorique pour détecter les cassures de barres dans les moteurs asynchrones triphasés à cage en tenant compte des harmoniques d’espace” Université Hassiba Benbouali, Département d’Electrotechnique, BP151, Hay Salem Chlef 02000, Algérie, 2024.

[3] N. Bessous, M. M. Rezaoui, I. Merzouk, A. Borni, L. Zarour. “Analyse détaillée du contenu spectral dans le courant statorique des machines asynchrones sous un défaut de la cassure de barres rotoriques en utilisant la technique de la MCSA-FFT,” Detailed analysis of spectral content in the stator current under broken rotor bar faults in induction motors using MCSA-FFT technique. Research Paper, DOI: 10.5281/zenodo.3738693, Open access, 2020.

[4] S. Alvarez, ‘‘Développement de loi de commande avec observateur pour machine asynchrone,’’ Institut national polytechnique de Grenoble. Laboratoire automatique de Grenoble. 2002.

[5] C. Bonivento, A. Isidori, L. Marconi, A. Paoli. “Implicit fault tolerant control, Application to induction motors” DEIS, University of Bologna, Italy, viale Risorgimento 2, 40136 Bologna, DIS, University of Roma "La Sapienza", Italy, Via Eudossiana 18, 00184 Roma, 2024.

[6] A. Gouichiche. “Diagnostic à base d’observateurs et commande tolérante aux défauts. Expérimentation sur des systèmes électriques” Thèse de Doctorat en Génie Electrique, Option. Automatique, Université de Tiaret, 2015.

[7] E. Schaeffer. “Diagnostic des machines asynchrones, modèles et outils paramétriques dédiés à la simulation et à la détection de défauts, Ph.D. dissertation” École Centrale de Nantes, 1999.

[8] S. Bachir. “Contribution au diagnostic de la machine asynchrone par estimation paramétrique” Ph.D. dissertation, Ecole supérieure d’ingénieurs de Poitiers, 2002

[9] J. Pyrhönen, T. Jokinen, and V. Hrabovcová. “Design of Rotating Electrical Machines” UK, J. Wiley & Sons, Ltd., 2008.

[10] S. Bachir and S. Tnani, “Induction motor modelling of broken rotor bars and fault detection by parameter estimation,” in Proc. IEEE Int. SDEMPED'01, Grado, Italy, pp. 145–149, September 2001.

[11] M. Ghanes, Observation et commande de la machine asynchrone sans capteur mécanique, Ph.D. dissertation, Université de Nantes, 2005

[12] J. Li, L. Xu, and Z. Zhang, “An adaptive sliding-mode observer for induction motor sensorless speed control,” IEEE Trans. Ind. Appl., vol. 41, no. 4, pp. 1039–1046, 2005

Practical Validation of a Control Algorithm for Power Factor Correction and Power Quality Improvement in Distribution Systems.

Messaouda Bedoui
L2GEGI Laboratory, Department of Electrical Engineering, Faculty of Applied Science.
Inb Khaldoun University
Tiaret, Algeria.
messaouda.bedoui@univ-tiaret.dz

Abderrahmane Berkani
L2GEGI Laboratory, Department of Electrical Engineering, Faculty of Applied Science
Inb Khaldoun University
Tiaret, Algeria
abderrahmane.berkani@univ-tiaret.dz

Karim Negadi
L2GEGI Laboratory, Department of Electrical Engineering, Faculty of Applied Science
Inb Khaldoun University
Tiaret, Algeria
karim.negadi@univ-tiaret.dz

***Abstract*— In recent years, the power quality of AC systems has become a significant concern due to the rapid increase in the number of power electronic devices. Most commercial and industrial installations in the country have large electrical loads, which are predominantly inductive in nature, resulting in a lagging power factor. This causes substantial penalties for consumers from the electricity board. To optimize the supply system from both engineering and economic perspectives, it is essential to maintain a power factor as close to unity as possible. This paper focuses on automatic power factor correction techniques designed to improve both power factor and power quality in three-phase network systems. The proposed control system will manage individual capacitors within a capacitor bank, operating in steps based on variations in the power factor. An automatic power factor correction algorithm will monitor the line voltage and current, determining the phase shift between these signals to calculate the power factor. The proposed control algorithm will then assess the compensation requirements and activate the necessary capacitors from the bank, adjusting the power factor to approximately unity.**

Keywords—Power factor, inductive load, Automatic power factor correction, capacitor bank, power quality.

I. INTRODUCTION

The increasing integration of power electronic devices in industrial and commercial systems has significantly impacted the power quality of electrical networks. A major consequence of this trend is the degradation of power factor, which can lead to higher energy costs, penalties from utility companies, and reduced system efficiency. Automatic Power Factor Correction (APFC) systems have emerged as an effective solution to this issue, aiming to adjust the power factor to optimal levels and mitigate the negative effects of reactive power [1-3].

In recent years, there has been a growing interest in the practical validation and simulation of APFC techniques to assess their real-world performance and optimize their application in modern electrical networks. While various simulation models have been developed to predict the behavior of these systems under different operating conditions, it is equally important to validate these models through practical experimentation. Practical validation involves deploying APFC systems in real-world settings and measuring their ability to improve both power factor and overall power quality [2.4-6].

This paper focuses on the simulation and practical validation of an automatic power factor correction system using a combination of theoretical modeling and real-time testing. The validation process includes assessing the accuracy of simulation results and comparing them with actual performance data obtained from distribution network. Recent advancements in digital controllers, sensor technologies, and machine learning algorithms are also explored, as they contribute to the enhanced effectiveness of APFC systems.

In the following section, we will explain how the improvement of power factor operates and discuss the methodology and details of the automatic correction of power factor used in this system. Sections 3 and 4 will present and discuss the results, and finally, in Section 5, we will conclude the findings of the study.

II. ADVANTAGES OF POWER FACTOR IMPROVEMENT

The power factor indicates the actual utilization of real power by the equipment. The power triangle, as shown in Figure 1, illustrates this relationship. The power factor, derived from the power triangle, is the cosine of the angle between the apparent power and the real power, and is denoted as Cos(Ø) [7].

.

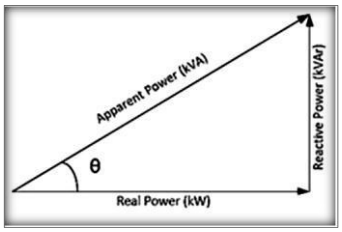


Fig. 1. Diagram power triangle.

$$P = \sqrt{3}.U.I.\cos\varphi \quad (1)$$

$$Q = \sqrt{3}.U.I.\sin\varphi \quad (2)$$

$$S = \sqrt{3}.U.I \quad (3)$$

Where:
P: Active power.
Q: Reactive power.
S: Apparent power.
U: Phase-to-phase voltage.
I: Current.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Power factor is expressed as a numerical value between 0 and 1 (or as a percentage), where a value of 1 (or 100%) signifies maximum efficiency. A power factor less than 1 indicates that current and voltage are not in phase in an AC circuit. The power factor can be denoted mathematically by:

$$\cos \emptyset = \frac{P}{S} \quad (4)$$

A. *Advantages of Power Factor Correction Techniques*

Improving the power factor not only reduces costs but also enhances system efficiency, extends the lifespan of equipment, improves power quality, and ensures compliance with regulations. These benefits make power factor correction an essential practice in both industrial and commercial electrical systems. Additionally, improving the power factor in electrical systems offers several significant advantages [7.8].

1) *Reduced Electricity Bills*: Many utility companies impose penalties for low power factor. By improving the power factor, businesses can avoid these charges, thereby reducing electricity costs. When the power factor is close to unity (1), electrical systems operate more efficiently, resulting in less wasted energy and potentially lower overall energy consumption.
2) *Improved Voltage Regulation:* A low power factor can cause voltage drops in the electrical system, particularly when the load is high. Improving the power factor helps maintain more stable voltage levels, thereby enhancing the performance and longevity of sensitive equipment. Higher power factors reduce the current flow required for the same amount of real power, which can lead to reduced line losses (I^2R losses) in the system.
3) *Improved Power Quality:* Power factor correction often involves the use of capacitors, which can help filter out harmonics and improve the overall quality of the power supply. By reducing reactive power in the system, power factor correction leads to a more stable and higher-quality power supply.
4) *Enhanced System Efficiency:* With a corrected power factor, less reactive power circulates in the system, improving the overall efficiency of the power distribution network. Electric motors, inverters, and other machinery operate more efficiently when running at or near a power factor of 1, leading to better overall system efficiency.
5) *Better Use of Capacitor Banks:* Power factor correction using capacitor banks improves the efficiency of these components, ensuring that the system operates at optimal power factor levels without excessive capacitor bank use.

B. *Compensation Methodology*

To limit line currents and reduce the reactive energy absorbed by the installation, it is essential to install reactive energy sources in parallel with the system. This technique, aims to improve the power factor (cos φ). Consider a device or a group of devices that requires an active power P (watts) and a reactive power Q (volt-amperes reactive, or VAR) [9].

1) *Principle of Compensation*

Compensating an installation involves adding a source of reactive energy compensation, which helps improve the power factor of the system. Figure 2 illustrates the vector representation of this compensation.

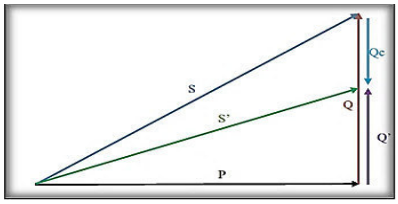


Fig. 2. Diagram of power compensation.

The power triangle method enables us to calculate the reactive power of the compensation capacitor and, consequently, its capacity. The reactive power for compensation, **Q**c, can be expressed as a function of the angles $\boldsymbol{\varphi}$ and $\boldsymbol{\varphi'}$

$$Q_c = P\,(tan\,\varphi' - tan\,\varphi) \quad (5)$$

Where:
P: Active power of the installation.
$\boldsymbol{\varphi}$: Phase angle before compensation.
$\boldsymbol{\varphi'}$: Phase angle after compensation.
Qc: Reactive power produced by a capacitor having at its terminals the voltage U, in a three-phase system is:

$$Q_c = 3.C.\omega.U^2 \quad (6)$$

Where:
U: Phase-to-phase voltage.
C: Capacitance.
ω: Network pulsation rad/s
With: $\omega = 2.\pi.f$ (7)

2) *Calculation of capacitance required*

We place a capacitor in parallel and upstream of the load to compensate for the reactive energy and then we dimension the capacitor to obtain:

$$\cos\varphi = 1 \quad (8)$$

A capacitor of capacity $\boldsymbol{C}$, presenting at its terminals a voltage $\boldsymbol{U}$ of pulsation $\boldsymbol{\omega}$ provides a reactive power can be determined using equation (9):

$$Q_C = -C\omega \times U^2 \quad (9)$$

The reactive power before compensation is given by the equation (10):

$$Q = P \times (\tan \varphi) \quad (10)$$

The reactive power after compensation is given by the equation (11):

$$Q' = P \times (\tan\varphi\,') \quad (11)$$

The compensation reactive power is given by the difference between equations (10) and (11).

$$Qc = Q - Q' \quad (12)$$

So;

$$Qc = P \times (\tan Q - \tan Q') = -C\omega \times U^2 \quad (13)$$

$$C = \frac{Qc}{U^2 \times \omega} \quad (14)$$

III. SIMULATION SYSTEM DESIGN AND DESCRIPTION

Figure 4 presents the practical MATLAB simulation of the model used in this study. The setup consists of a three-phase generator, a three-phase transformer, a transmission line, and an inductive linear load (R-L). The secondary side of the transformer is connected to the transmission line, which provides power to the R-L load. This load operates with a power factor below one. To improve the system's power factor, calculations show that a capacitor bank is required. Prior to compensation, no capacitor bank is installed between the load and the transmission line.

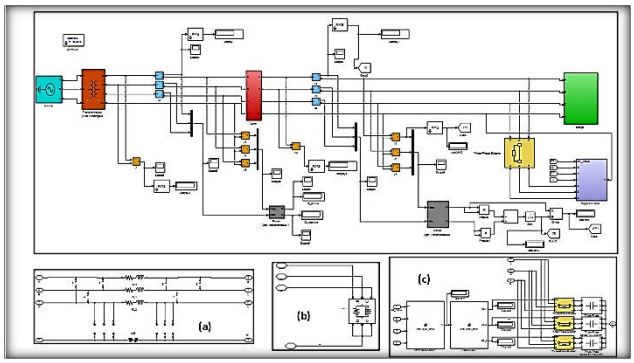


Fig. 3. Schematic Diagram for Automatic Power Factor Correction Using Control Algorithm.: a) Transmission line, b) Three phase loads and c) Capacitor Bank.

For inductive load L=2.4 H, we will show the case before PF correction and after PF correction.

A. *Results Before Power Factor correction*

Table 1 represents the simulation results before the insertion of the compensation capacitor (breaker is open). At the beginning we set the value of the inductive load to 2.4H which is connected in parallel with a resistive load of 1000 Ω then at each step, we measure respectively the following quantities: U_1, I_1, P_1 and Q_1 at the beginning of the line and U_2, I_2 and cos φ at the end of the line.

TABLE I. SIMULATION RESULTS BEFORE COMPENSATION

L(H)	*U_1(V)*	*I_1(A)*	*P_1(W)*	*Q_1(VAR)*	*U_2(V)*	*I_2(A)*	*Cosφ*
2.4	376.9	0.2	114	173.6	331.7	0.31	0.6

From table 1 and figure 4, it can be observed:

- An increase in the line current.
- Voltage drops at the line's arrival.
- The load can consume a significant amount of reactive power, leading to substantial losses in the system.
- A power factor lower than 0.96 (cos φ < 0.96) is considered poor.

Correcting the power factor in this case is highly beneficial, as a low power factor can overload cables due to the high reactive component, resulting in increased current. This overload can influence cable design and lead to significant additional costs in billing. Therefore, correcting the power factor can not only improve energy efficiency but also reduce costs associated with overload.

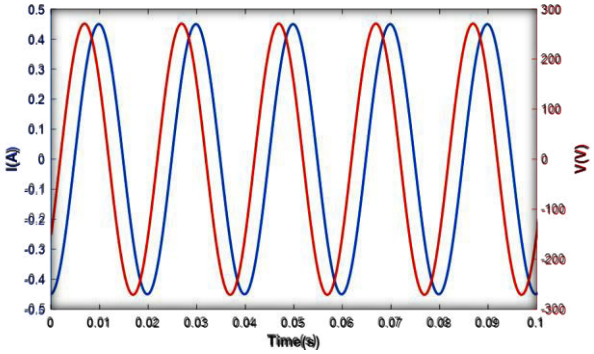


Fig. 4. Phase shift between voltage and current before correction.

B. *Results After Power Factor correction*

In this part, we have inserted a capacitor bank (breaker is closed) in parallel with the load.

Figures 3-c) displays a designed for the automatic calculation of the capacitance value required to improve the power factor of the electrical system.

From the measurements in Table I and by using the equation 14, we can be determined the theory value of the compensation capacitor.

$$C = \frac{Qc}{U^2 \times \omega} = \frac{P \times (\tan Q - \tan Q')}{U^2 \times \omega} = 4.37\ \mu F \qquad (15)$$

TABLE II. SIMULATION RESULTS AFTER COMPENSATION

L(H)	*U_1(V)*	*I_1(A)*	*P_1(W)*	*Q_1(VAR)*	*U_2(V)*	*I_2(A)*	*Cosφ*
2.4	382.1	0.217	143.1	12.06	375.9	0.217	1

Table II shows that if we compare the measurements of **U_1** and **U_2** voltages after compensation, we observe a slight decrease in **U_2** voltages. This may be due to different factors, such as the decrease in reactive power and the power factor correction is perfect to see that the power factor (cos φ) is still greater than 0.96 after compensation.

From figure 5 we mean that the compensation has successfully corrected the phase shift between voltage and current in the circuit, which is essential to maximize efficiency and minimize energy losses and improve overall system performance.

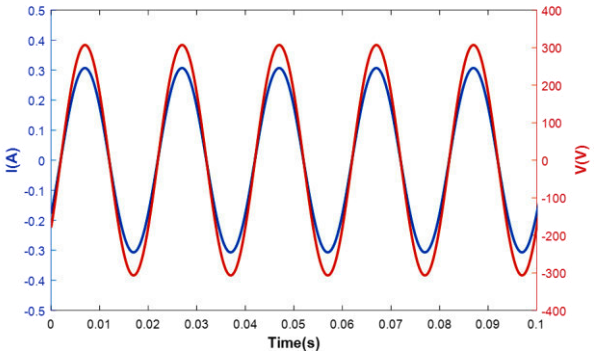


Fig. 5. Phase shift between voltage and current after correction.

IV. HARDWARE SYSTEM OVERVIEW

This section presents an experimental study focused on improving the power factor. An experimental setup was created in the Electrical Network laboratory of the

Electrical Engineering Department at IBN KHALDOUN University in Tiaret. The study specifically investigates the parallel compensation of reactive energy and the techniques employed to correct the power factor.

We use in this part the following materials:

- Three-phase power supply.
- Transmission line.
- Measuring devices (ammeter, voltmeter).
- Arduino UNO card with an ATmega328 microcontroller.
- Sensors (current and voltage).
- Relays.
- Loads (resistances, inductances, capacities).
- Three-phase electrical network analyzer CA 8331.

Figure 6 demonstrates the principle behind the automatic power factor detection and correction circuit. This process involves monitoring the power factor through a microcontroller embedded in the Arduino UNO board, which is programmed with specific code. Voltage and current values are obtained from the main AC line using current and voltage sensors. The microcontroller processes this data and uses an algorithm to calculate the required capacitor size for power factor correction, while also continuously monitoring the load behavior based on the load current.

When the power factor is low, the microcontroller sends a signal to the relay switching unit to activate the necessary capacitor. The microcontroller's actions to correct the low power factor, such as selecting the appropriate capacitor and continuously monitoring the load, are displayed on an LCD screen.

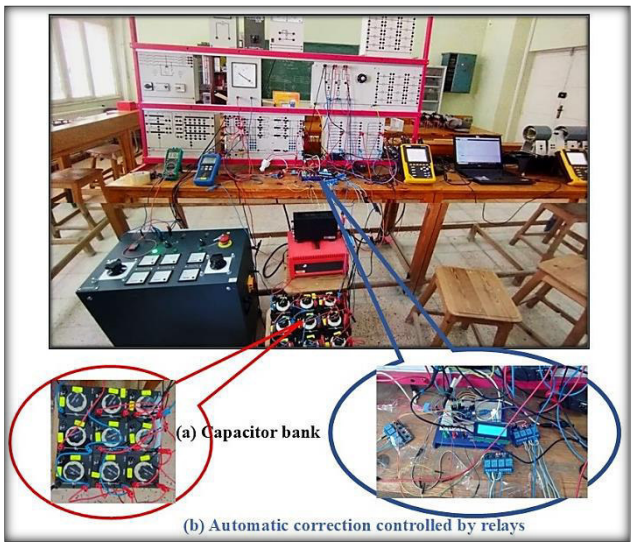


Fig. 6. Implementation of parallel compensation with automatic Arduino control

A. *Results and Discussion Before Compensation*

In this case, the relays are completely deactivated, preventing any capacitors from being connected to the electrical installation. This setup allows us to observe the system prior to power factor (PF) correction. We used a mixed resistive-inductive load with the following specifications: inductance (L) of 2.4 H and resistance (R) of 1000 ohms.

Table III and Figure 7 present the electrical characteristics of this load. The phase shift is observed to be 44.76°, resulting in a power factor of 0.7. This low power factor highlights inefficiencies in the system, which could be mitigated through corrective measures.

TABLE III. PRACTICAL MEASUREMENTS BEFORE CORRECTION

L(H)	*U_1(V)*	*I_1(A)*	*P_1(W)*	*Q_1(VAR)*	*U_2(V)*	*I_2(A)*	*Cosφ*
2.4	379.8	0.27	126.1	124.7	318	0.37	0.7



Fig. 7. Electrical characteristics of load before correction

To verify the results displayed on the LCD, we used a three-phase electrical network analyzer, the CA 833, to record the voltage and current measurement signals, as shown in Figures 8 and 9.

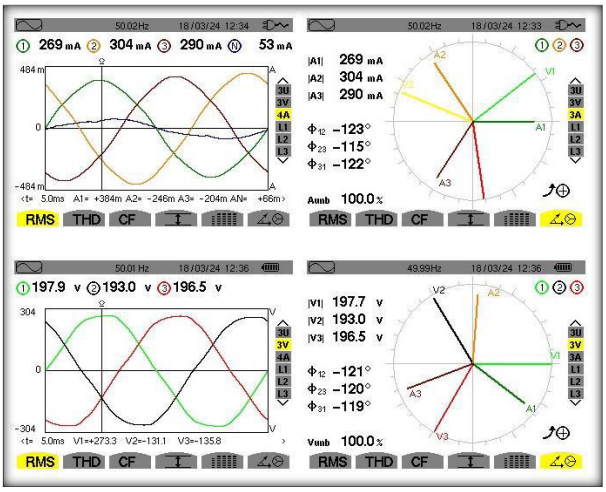
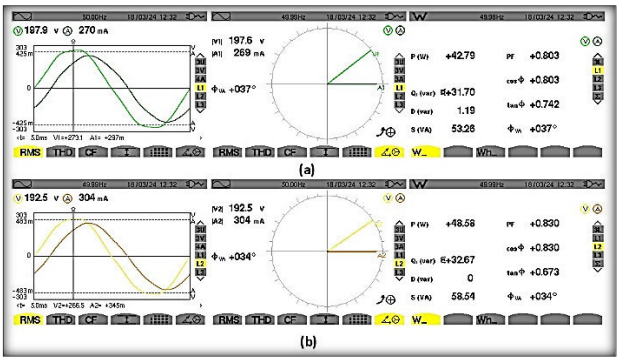


Fig. 8. Experimental results: Current and voltage of three-phase load



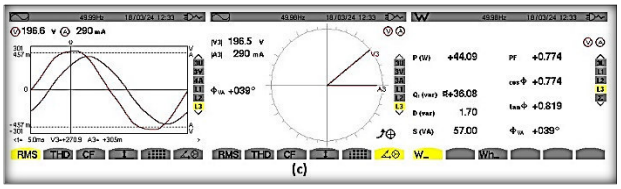


Fig. 9. Experimental results before correction: Phase shift Vector diagram, and powers. a): phase1, b): phase 2 and c): phase 3

Figure 9 shows the current and voltage signals, along with vector diagrams, illustrating the phase shift between the current and voltage. The load induces phase shifts of 36.58°, 33.9°, and 39.28° for each phase, resulting in power factors of 0.802, 0.830, and 0.774, respectively. These low power factors indicate the consumption of reactive power in each phase. Since the power factor is considerably less than unity, it is essential to add compensation capacitors to correct this issue.

B. Results and Discussion After Compensation

In the test conducted after compensation and automatic power factor correction, the relays are controlled automatically by the Arduino, which calculates the required correction value. The Arduino then activates the appropriate relay to close the circuit, connecting the capacitor bank to the electrical network. These capacitors are specifically designed to improve the system's power factor and reduce the amount of non-productive reactive energy circulating within the network.

From the measurements in table 3 and by using equation (14), we can determine the theoretical value of the compensation capacitor.

$$C = \frac{Qc}{U^2 \times \omega} = \frac{P \times (\tan Q - \tan Q')}{U^2 \times \omega} = 4.05\ \mu F \qquad (16)$$

The addition of capacitors has resulted in significant improvements to the electrical system, as shown in Table IV and Figure 10. A notable reduction in current indicates a decreased demand for reactive energy and enhanced efficiency. Active power has also increased, reflecting more effective energy use. The reduction in reactive power highlights the capacitors' effectiveness in improving the power factor. Furthermore, the power factor has improved significantly. This improvement leads to lower energy losses and greater system efficiency. Overall, these changes demonstrate the effectiveness of using capacitors for reactive energy compensation, making the system more cost-effective and reliable.

TABLE IV. PRACTICAL MEASUREMENTS AFTER CORRECTION

L(H)	*U_1(V)*	*I_1(A)*	*P_1(W)*	*Q_1(VAR)*	*U_2(V)*	*I_2(A)*	*Cosφ*
2.4	379.1	0.251	164.7	5.7	379.3	0.25	1



Fig. 10. Electrical characteristics of load after correction

To verify the results displayed on the LCD, we used a three-phase electrical network analyzer, the CA 833, to capture the voltage and current measurement signals, as shown in Figures 11 and 12.

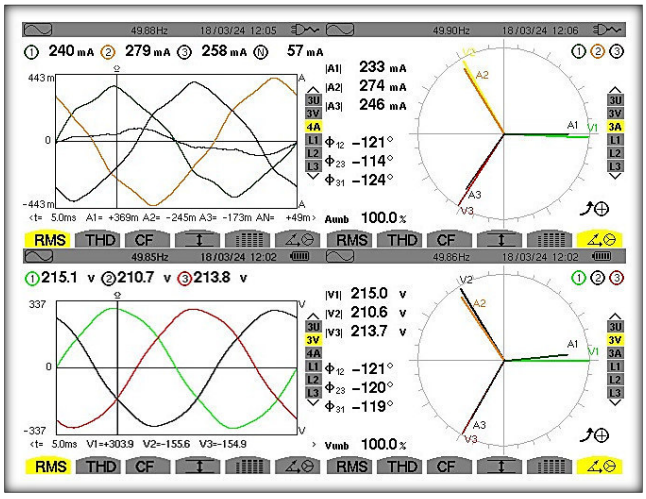


Fig. 11. Experimental results after correction: Current and voltage of three-phase load

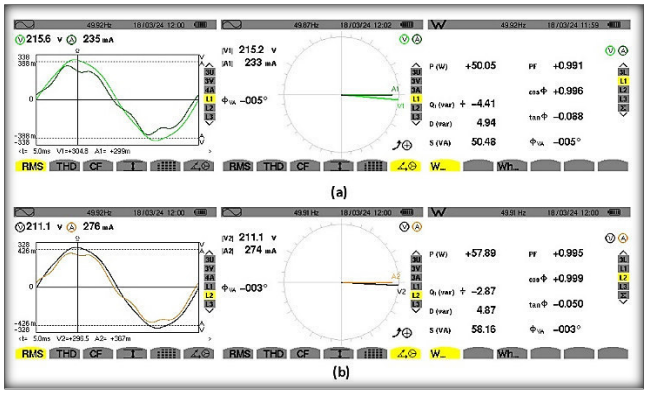


Fig. 12. Experimental results after correction: Phase shift Vector diagram, and powers. a): phase1 and b): phase 2

C. Comparative study

Table 4 summarizes the results before and after power factor correction. Initially, the measured power factor caused cable overloading due to the high current from the reactive component, leading to significant additional costs. However, after implementing the correction device, a noticeable improvement was observed: the current decreased, and the power factor improved, effectively eliminating these extra expenses. This improvement occurs automatically and adaptively, highlighting the effectiveness of the power factor correction system in

optimizing electrical performance and reducing operational costs.

TABLE V. RECAPS THE RESULTS OBTAINED BEFORE AND AFTER CORRECTION

Load	Voltage U(V)	Current I (A)	Active power P(W)	Reactive power Q(Var)	Cosφ
Before compensation					
Simulation	376.9	0.2	114	173.6	0.602
practical	379.8	0.27	126.1	124.7	0.7
After compensation					
simulation	382.1	0.21	164.1	12.06	1
practical	379.1	0.25	164.7	5.7	1

V. CONCLUSION

In this study, both simulation and practical implementation of automatic power factor correction using Arduino have been successfully demonstrated. The simulation provided a solid theoretical foundation, highlighting the significant impact of power factor correction on system efficiency and energy consumption. Through the practical setup, the Arduino-based control system effectively monitored and adjusted the power factor by automatically activating the appropriate capacitors. The results showed a notable improvement in the power factor, reduced reactive energy consumption, and enhanced overall system efficiency. This approach not only demonstrated the feasibility of using Arduino for real-time power factor correction but also emphasized its cost-effectiveness and adaptability for various electrical systems. The integration of automation through the Arduino platform offers a reliable solution for optimizing energy use, reducing operational costs, and improving the sustainability of electrical networks.

ACKNOWLEDGMENT

We would like to express our heartfelt gratitude to everyone who contributed their valuable time and efforts, both directly and indirectly. Your support has been invaluable.

REFERENCES

[1] J. Gupta, R. Kushwaha, B. Singh, and V. Khadkikar, "Improved power quality charging system based on high step-down gain bridgeless SEPIC APFC for light electric vehicles," IEEE Transactions on Industry Applications, vol. 58, pp. 423-434, 2021

[2] H. kumar Naidu and P. Ramtekkar, "Power Factor Improvement Using Artificial Intelligence Technique," in Renewable Resources and Energy Management, ed: CRC Press, 2023, pp. 192-196.

[3] S. M. Lambe and K. J. Karande, "A Long-Term Effect of Power Factor Correction Capacitors on Academic Institute," JOURNAL OF TECHNICAL EDUCATION, p. 166, 2023.

[4] M. Khan, "Automatic power factor correction," Project report, University of Mumbai, 2021.

[5] P. I. Udenze, T. K. Genger, and M. O. Ekoja, "An Arduino-Based Automatic Power Factor Correction Device," vol. 7, pp. 14-22, 2021.

[6] M. L. Katche, P. Tsafack, and E. Tanyi, "Proposed Automatic Load-Side Power Factor Correction System Algorithm and Its Economic Benefits in Battery Based Renewable Energy Systems," International Journal of Advances in Scientific Research and Engineering (ijasre) vol. 7, pp. 14-22, 2021.

[7] B. N. Alhasnawi and B. H. Jasim, "Automated Power Factor Correction for Smart Home," Iraqi Journal of Electrical and Electronic Engineering, vol. 14, 2018.

[8] M. Madhiarasan, "Design of an Internet of Things Powered Automated Power Factor Correction System and Monitoring of Consumption of Energy," Wireless Personal Communications, vol. 133, pp. 31-48, 2023.

[9] X. Zhou, K. Wei, Y. Ma, and Z. Gao, "A review of reactive power compensation devices," in 2018 IEEE International Conference on Mechatronics and Automation (ICMA), 2018, pp. 2020-2024.

FRACTIONAL ADAPTATIVE CONTROL

1st Belgacem Mecheri
dept. Automation and electrificationof systems
(Applied automatic laboratory)
M'hamed Bougara University of Boumerdes
Boumerdes, Algeria
mecheri.belgacem@uni-boumerdes.dz

2nd Aicha Name Djalab
dept. Elecrical Systems Engineering
Faculty of technology
M'hamed Bougara university of Boumerdes
Boumerdes ,Algeria
a.djaleb@univ-boumerdes.dz

3rd Nora Kcimi
dept. Elecrical Systems Engineering
Faculty of technology
M'hamed Bougara university of Boumerdes
Boumerdes ,Algeria
Kacimi.nora@yahoo.dz

4th Abdelkader Nadjem
dept. Elecrical Engineering
University Abbes Laghrour *Khenchela*
Khenchela, Algeria
nadjem.abdelkader@uni.khenchela.dz

Abstract*— Fractional adaptive control is a novel approach that combines the advantages of classical adaptive control and fractional order concepts to improve the performance and robustness of dynamic systems. This paper provides an in-depth exploration of this method, highlighting its theoretical foundations and practical advantages. A review of recent works in this area is presented, highlighting the applications of fractional adaptive controllers in complex environments. Based on a rigorous methodology, the paper proposes a case study applied to a typical dynamic system, demonstrating the effectiveness of this approach in terms of stability and accuracy. The results obtained show that fractional adaptive control outperforms traditional methods, especially in situations where uncertainties and disturbances are significant. Finally, future research perspectives in this area are discussed, paving the way for new industrial applications.

Keywords— Adaptive command, fractional controller, parameter optimization, genetic algorithm, dynamic systems stability

I. INTRODUCTION

The control of dynamic systems is a central area of research in automation, particularly when it comes to managing complex systems subject to uncertainties. In this context, adaptive control techniques have demonstrated their effectiveness by allowing the system to automatically adjust to parameter variations or unforeseen disturbances. However, classical adaptive control approaches sometimes have limitations, particularly in terms of robustness and performance in environments where dynamic phenomena are poorly modeled.

At the same time, fractional order systems, characterized by non-integer derivatives and integrals, provide a theoretical framework capable of modeling real dynamics with high accuracy. These approaches allow capturing intermediate behaviors between strictly integer models, thus providing increased flexibility in system control. Fractional controllers, such as fractional PIDs (PIλDμ), introduce two additional degrees of freedom that improve robustness and performance accuracy, especially in complex environments.

Many recent works have highlighted the advantages of fractional approaches in the field of control. For example, Monje et al. (2010) demonstrated the effectiveness of fractional PIDs in industrial systems subject to disturbances and uncertainties, obtaining better stability compared to classical PIDs. Furthermore, Chen et al. (2018) explored the use of genetic algorithms to optimize the parameters of fractional controllers, showing their rapid adaptation and robustness in dynamic environments. Ahmed et al. (2020) applied these concepts in the field of robotics, significantly improving trajectory tracking for robots operating in unstructured environments. Finally, Li et al. (2019) demonstrated that the integration of stochastic optimization techniques, such as particle swarm algorithms (PSO), allows the design of efficient adaptive fractional controllers suitable for large-scale systems.

Despite these advances, many challenges remain in the practical application of fractional adaptive control. These challenges include real-time optimization of fractional parameters, management of computational costs, and implementation in complex industrial environments. The aim of this paper is therefore to present an in-depth analysis of fractional adaptive control, highlighting its theoretical basis, advantages, and potential applications. A case study will also be introduced to illustrate the effectiveness of this method compared to traditional approaches.

II. STATE OF THE ART

1. Advances in Adaptive Control

μResearch in adaptive control has overcome the limitations of fixed-parameter controllers by introducing methods that can dynamically adjust parameters based on system conditions. Notable approaches include:

Adaptive Reference Model Control (MRC): Used to maintain the output of a system close to that of an ideal reference model.

Robust Adaptive Control: Designed to handle large uncertainties while ensuring stable performance.

However, these conventional approaches show limitations in complex systems where nonlinear dynamics or large disturbances are present.

2. Introduction of fractional concepts

Fractional order systems, integrating non-integer derivatives$D^{\propto}$, have opened new perspectives in the control of dynamic systems. They allow the modeling of complex behaviors with memory effects, absent in classical

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

approaches. Fractional PID controllers, have demonstrated their effectiveness in the robust control of dynamic systems through the addition of fractional orders λ and μ which allow a more precise and adaptable response.

4. Methodology and Modeling

4.1. DC Motor Modeling

$$V_a = L_a \frac{di_a}{dt} + R_a i_a + K_B \omega \quad (1)$$

$$J \frac{d\omega}{dt} + B\omega = K_t i_a \quad (2)$$

4.2. Internal Model

The internal model is a simplified approximation of the engine dynamics, used to facilitate error estimation and parameter adaptation. Its transfer function is given by

$$G(s) = \frac{K}{\tau s+1} \quad (3)$$

K: static gain of the internal model, estimated in steady state
τ : time constant, calculated from the dynamic parameters
This classic internal model is used to calculate the error between the actual output of the engine (y(t)) and that predicted by the internal model (y^(t)).

4.3. Fractional PID Controller

The motor is controlled using a fractional PID ($PI^{\lambda}D^{\mu}$) , whose transfer function is given by:

$$C(s) = K_p + \frac{K_i}{s^{\lambda}} + K_d s^{\mu} \quad (4)$$

K_p , K_i , K_d: proportional, integral and derivative gains.
λ, μ Fractional orders for integration and differentiation (λ<0, μ<0)
The order is calculated based on the error:

$$e(t) = r(t) - \hat{y}(t) \quad (5)$$

Where r(t) is the instruction and $\hat{y}(t)$ the output of the internal model.

4.4. Parameter Adaptation

To optimize performance, the fractional PID parameters (K_p , K_i , K_d, λ, μ)
are dynamically adjusted using a genetic algorithm. This algorithm follows the following steps:
1-Initialization: An initial population of potential solutions is randomly generated.
2-Evaluation: Each solution is evaluated according to a cost function based on the cumulative squared error (ISE):

$$J = \int_0^T e^2(t)dt \quad (6)$$

3-Selection and Crossover: The best solutions are combined to produce a new generation.

4-Mutation: Small perturbations are introduced to further explore the solution space.

5-Convergence: The process is repeated until the optimal parameters are obtained.

4.5. Integration of Adaptive Control

The overall architecture of the system is organized as follows:

Calculating the error: The difference between the motor output y(t) and that of the internal model $\hat{y}(t)$ is used to adapt the parameters

Control: The fractional PID generates the command u(t) to keep the system stable and precise, even in the face of disturbances or parameter variations.

5. Results and Discussions

5.1. Determination of Controller Parameters

The controller parameters were optimized using a genetic algorithm. The cost function used for this optimization is based on the cumulative squared error (ISA)

$$J = \int_0^T e^2(t)dt \quad (6)$$

Where y(t) is the instruction and $\hat{y}(t)$ the output of the internal model.
Parameters obtained for the classic PID:

$K_p = 1{,}5\ ;\ K_i = 0.8;\ K_d = 0.1$

Parameters obtained for fractional PID:

$K_p = 1{,}7\ ;\ K_i = 0.9;\ K_d = 0.7;\ \lambda = 0.8; \mu = 0.7$

These settings ensure optimal performance for each controller under the simulated conditions.

5.2. Step Response

The performance of the optimized controllers was evaluated in response to a step-type input. The results show that:

- Adaptive fractional PID reduces the rise time to 0.3 s, compared to 0.5 s for conventional PID.
- The steady-state error is eliminated with fractional PID, while a slight deviation remains with classical PID.
- Overshoot is better controlled with fractional PID, providing a more stable response.

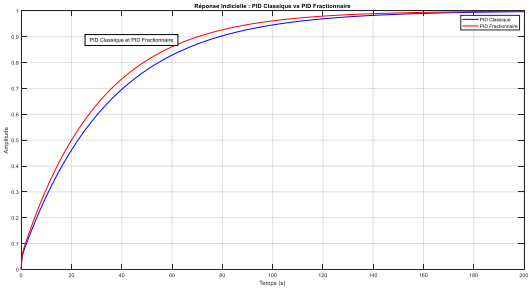
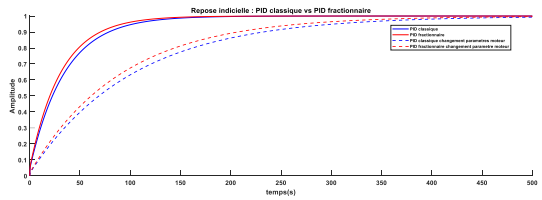


Figure (1) Classic PID vs Fractional PID

The R,B meter parameters are modified at t=5s (increase of R and B)



Classic PID vs Fractional PID with changing motor parameters

Expected results:

Graph 1: Initial parameters

Fractional PID offers a faster response, with controlled overshoot and increased accuracy compared to classical PID.

5.2 Discussion

The results obtained highlight the notable differences between the performances of the classical PID and the fractional PID in the two simulated scenarios:

1- Initial parameters

The fractional PID showed a faster and more accurate response, with minimal overshoot and rapid stabilization. This demonstrates its ability to better manage the dynamics of the initial system, thanks to its more flexible structure and its ability to model complex behaviors.

2-After changing the motor parameters (increasing resistance R and friction B), the fractional PID continued to adapt effectively, maintaining a stable response with minimal disturbance.

The classical PID, however, showed increased oscillations and a prolonged settling time. This illustrates its limitations in the face of large dynamic variations.

6. Conclusion

This work allowed to compare the performances of the classical PID and fractional PID controllers in the context of the control of a direct current motor, both with initial parameters and after a modification of the dynamics of the motor

Main contributions:

Better performance of fractional PID: The fractional controller has demonstrated its superiority in terms of speed, precision and robustness, even in the face of dynamic disturbances or noise.

Limitations of classical PIB: Although effective under stable conditions, classical PID shows weaknesses when faced with parametric variations or uncertainties.

Perspectives:

Extend this study to include a fractional internal model, which could further improve the control accuracy and adaptability.

Apply these methods to more complex systems to validate their effectiveness in real environments.

III. References

[1]. I. Podlubny, “Fractional-Order Systems and Fractional-Order Controllers,” IEEE Transactions on Automatic Control, vol. 44, no. 1, pp. 208–214, Jan. 1999.
[2]. Y. Chen, K. L. Moore, and K. M. Grigoriadis, Fractional-Order Control: Theory and Applications. Springer, 2003.
[3]. M. Ahmed and Y. Wang, “Adaptive Fractional-Order PID Control for Robotic Manipulators Under Dynamic Uncertainty,” International Journal of Control, vol. 93, no. 4, pp. 833–849, Apr. 2020.
[4]. C. Chen and X. Li, “Genetic Algorithm-Based Optimization of Fractional-Order PID Controllers for Industrial Applications,” IEEE Access, vol. 7, pp. 10034–10045, 2019.
[5]. R. C. Dorf and R. H. Bishop, Modern Control Systems, 13th ed. Pearson, 2017.
[6]. Y. Wang and Z. Song, “Integration of Fractional-Order Internal Models in Adaptive Control Systems,” IEEE Transactions on Control Systems Technology, vol. 28, no. 2, pp. 401–412, Mar. 2020.
[7]. Y. Li and Y. Luo, “Comparison of PID and Fractional-Order PID Controllers in DC Motor Control,” Journal of Control Science and Engineering, vol. 2019, Article ID 3045693, 2019.
[8]. B. Mecheri, “Stabilization of chaotic systems by fractional PID controller using Takagi-Sugeno model,” Conference on Control Systems and Applications, 2023.
[9]-Mecheri, B and all. (2021). Fractional order PID controller for the stabilization of chaotic systems using Takagi-Sugeno fuzzy model and all, International journal of systems, control and communications. 12, 1-11

Comparing Two Control Techniques Implemented on the Permanent Magnet Synchronous Machine (PMSM): A Comparative Study

Wassila HATTAB
Laboratory of Biskra LGEB, Electricals Engineering
Biskra University
Biskra, Algeria
wassila.hattab@univ-biskra.dz

Abdelhamid BENAKCHA
Laboratory of Biskra LGEB, Electricals Engineering
Biskra University
Biskra, Algeria
abdelhamid.benakcha@univ-biskra.dz

Seddik TABET
Laboratory of Biskra LGEB, Electricals Engineering
Biskra University
Biskra, Algeria
seddik.thabet@univ-biskra.dz

Amira SLIMANI
Laboratory of Biskra LGEB, Electricals Engineering
Biskra University
Biskra, Algeria
amira.slimani@univ-biskra.dz

***Abstract*—Permanent Magnet Synchronous Motors (PMSM) are popular in motion control applications due to their long lifespan, efficient energy use, and low maintenance needs. However, the performance of PMSM can be affected by changes in parameters and load conditions. To address this, vector control techniques are used, allowing PMSM to operate similarly to separately excited DC motors in demanding applications. This paper examines speed control challenges in PMSM by using both PI control and backstepping control based on the Lyapunov theorem. The PI controller achieves good results for speed regulation. To improve response time and resilience to uncertainties and disturbances, backstepping control is introduced. This method ensures precise speed control, completely eliminating speed errors, and offers robust management of the PMSM with fast response and strong resistance to disturbances across various conditions.**

***Index Terms*—PMSM, Vector control, PI control, Backstepping control, Lyapunov theorem.**

I. Introduction

Given the high efficiency, elevated power density, and broad speed range of Permanent Magnet Synchronous Motors (PMSMs), they have become highly attractive for various applications, including renewable energy systems and electric vehicles [1]. PMSMs offer a unique advantage of reduced power losses due to the absence of rotor windings, making them suitable for high-performance applications where efficiency and precise control are paramount. However, PMSMs exhibit nonlinear and highly coupled dynamics, posing significant challenges in designing an effective control strategy [2]. To address these challenges, advanced control strategies such as Field-Oriented Control (FOC) and Direct Torque Control (DTC) have been implemented, allowing precise control over the torque and flux components by transforming the stator currents into a rotating reference frame using Park's transformation [3].

In particular, FOC enables a decoupled control of torque and flux through the use of d and q axis current components. The torque-producing current, aligned with the q-axis, and the flux-producing current, aligned with the d-axis, allow PMSMs to emulate the control structure of a DC motor. This setup, often realized through Proportional-Integral (PI) controllers, is widely adopted due to its simplicity. Nonetheless, PI controllers may not be sufficient to address dynamic disturbances and parametric uncertainties in PMSM applications with varying load conditions .

As a result, researchers have extensively explored nonlinear control techniques to enhance the robustness and performance of PMSM systems. Techniques such as Model Predictive Control (MPC) [4], Sliding Mode Control (SMC) [5], and Backstepping Control [6] have been proposed to improve the dynamic response and fault tolerance of PMSMs. Among these, the backstepping control strategy has gained attention due to its recursive design methodology, which is particularly suited for nonlinear systems like PMSMs [7]. Backstepping control enables precise tracking of speed and position by systematically addressing each state in the PMSM's dynamic model [8]. However, for backstepping control to perform effectively, an accurate understanding of machine parameters and operating conditions is essential.

To further enhance system robustness under model uncertainties and external disturbances, adaptive backstepping control methods have been developed. These methods allow real-time adaptation to variations in machine parameters, thus improving stability and performance across diverse operational conditions [9]. Recent studies have shown that adaptive backstepping control strategies can achieve faster response times and increased robustness, especially in scenarios where load changes are frequent [10]. Furthermore, intelligent control

methods, such as neural networks and fuzzy logic, have been integrated with adaptive control to provide additional robustness and adaptability to unforeseen variations in PMSM dynamics [11], [12].

The remainder of this paper is structured as follows: Section 2 discusses an advanced speed control strategy for PMSMs using vector-based algorithms and backstepping control to achieve high-performance speed regulation. Section 3 presents the simulation parameters, conditions, and results. Finally, Section 4 provides a comprehensive summary and conclusions.

II. Control of the speed produced by PMSM

To achieve an optimal speed output from the Permanent Magnet Synchronous Motor (PMSM), it is essential to employ suitable optimization control techniques for regulating the rotor's speed. In this context, we introduce two control techniques:

A. *Vector control based on PI controller*

This approach is among the strategies employed to enhance the control quality of the mechanical power generated by the Permanent Magnet Synchronous Motor (PMSM). In this method, the stator current vector is decomposed into two components one governing flux control and the other influencing torque, aligning its dynamics with that of the separately excited DC machine [13]. The model of the PMSM is represented by the following equations:

$$\begin{cases} V_d = (R_s + L_d p) i_d - w L_q i_q \\ V_q = (R_s + L_q p) i_q + w L_d i_d + w \varphi_f \\ j \frac{d}{dt} w = C_e - C_r - f w \end{cases} \tag{1}$$

With:

$$\begin{aligned} w &= p \cdot w_r \\ C_e &= \frac{3}{2} p \left(L_d - L_q\right) i_d \cdot i_q + \varphi_f \cdot i_q \end{aligned} \tag{2}$$

The essence of vector control for the Permanent Magnet Synchronous Motor (PMSM) involves aligning the rotor flux along the d-axis. This strategy aims to maintain a constant alignment of the d-axis with the magnet's flux vector. The reference for the current, id is kept at zero. Within this control scheme, nonlinear decoupling is employed to close the Park current regulation loop using Proportional-Integral (PI) controllers [14].

$$i_d = 0 \Rightarrow i_q = i_s \tag{3}$$

The system of equations for the Permanent Magnet Synchronous Motor (PMSM) is expressed as:

$$\begin{cases} V_d = -\omega L_q i_q \\ V_q = (R_s + L_q p) i_q + \omega \, \varphi_f \\ C_e = \frac{3}{2} p \, \varphi_f i_q \end{cases} \tag{4}$$

Since the flux (φ_f) is held constant, the torque is directly proportional to the current component i_q.

B. *Backstepping Controller*

In this section, we have presented the application of backstepping control in the PMSM motor. This approach is designed to maintain the same general structure as vector control, while ensuring regulation and limitation of currents.

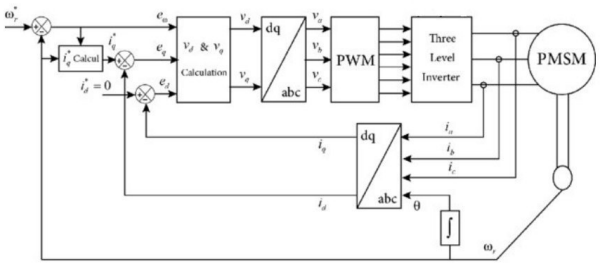


Fig. 1. Block diagram of backstepping control of PMSM

The model is written in the following form:

$$\begin{bmatrix} \frac{di_d}{dt} \\ \frac{di_q}{dt} \\ \frac{dw}{dt} \end{bmatrix} = \begin{bmatrix} \frac{-R_s}{L_d} i_d + \frac{w_r L_q}{L_d} i_q \\ \frac{-R_s}{L_q} i_q - \frac{w_r L_q}{L_d} i_d - \frac{w_r \varphi_f}{L_q} \\ \frac{3}{2} \frac{P \varphi_f}{J} i_q - \frac{P(L_q - L_d)}{J} i_d i_q - \frac{f}{J} w \end{bmatrix} + \begin{bmatrix} \frac{1}{L_d} & 0 & 0 \\ 0 & \frac{1}{L_q} & 0 \\ 0 & 0 & \frac{-1}{J} \end{bmatrix} \begin{bmatrix} v_d \\ v_q \\ c_r \end{bmatrix}$$

Can be rewritten in the following form:

$$\begin{cases} \frac{di_d}{dt} = \frac{-R_s}{L_d} i_d + \frac{L_q}{L_d} w_r i_q + \frac{1}{L_d} v_d \\ \frac{di_q}{dt} = \frac{-R_s}{L_q} i_q - \frac{L_q}{L_d} w_r i_d - \frac{\varphi_f}{L_q} w_r + \frac{1}{L_q} v_q \\ \frac{dw}{dt} = \frac{3}{2} \frac{P \varphi_f}{J} i_q - \frac{P(L_q - L_d)}{J} i_d i_q - \frac{f}{J} w - \frac{c_r}{J} \end{cases} \tag{5}$$

The fundamental idea behind backstepping control is to make the closed-loop system of cascaded first-order subsystems stable in the sense of Lyapunov, providing it with robust qualities and asymptotic global stability. Using these variables i_q, i_d the voltage v_q, v_d commands are then calculated to ensure speed control of the PMSM and the overall stability of the system.

1) First Step - Calculation of the V_{dref} Control Law
Given that the direct current is a controlled quantity, its desired value and regulation error are defined by:

$$\begin{aligned} i_{\text{dref}} &= 0 \\ e_1 &= i_{\text{dref}} - i_d \end{aligned} \tag{6}$$

From equations (5) and (6), the dynamic equations of the error are:

$$\begin{aligned} \dot{e}_1 &= \dot{i}_{dref} - \dot{i}_d \\ e_1 &= i^*_{dref} + \frac{R_s}{L_d} i_d - \frac{L_q}{L_d} w_r i_q - \frac{1}{L_d} v_d \end{aligned} \tag{7}$$

Since the objective requires the error e_1 to converge to zero and also demands that the current be regulated and limited, we use the Lyapunov function V_1, which represents in a sense, the energy of the error

$$V_1 = \frac{1}{2}e_1^2 \tag{8}$$

The derivative of the function, written from equations (6) and (7), is as follows:

$$\begin{aligned} \dot{V_1} &= e_1\dot{e}_1 \\ \dot{V_1} &= e_1\left(\dot{i}_{\text{dref}} + \frac{R_s}{L_d}i_d - \frac{L_q}{L_d}w_r i_q - \frac{1}{L_d}v_d\right) \end{aligned} \tag{9}$$

To ensure that the derivative of the criterion is always negative, the derivative of V_1 must take the form $\dot{V_1} = -K_1e_1^2$ introduced by the backstepping method, where

$$\dot{V_1} = e_1\left(\dot{i}_{\text{dref}} + \frac{R_s}{L_d}i_d - \frac{L_q}{L_d}w_r i_q - \frac{1}{L_d}v_d\right) = -k_1e_1^2 \tag{10}$$

This equation allows us to define the voltage command V_d for the subsystem to ensure Lyapunov stability and force the current i_d to follow its reference $i_{dref} = 0$. We obtain the reference voltage V_{dref}:

$$v_{\text{dref}} = L_d\left(k_1e_1 + \frac{R_s}{L_d}i_d - \frac{L_q}{L_d}w_r i_q\right) \tag{11}$$

2) Second Step - Calculation of the Virtual Current Control Law $i_{qref} = 0$

Since the rotor speed is the primary modulating variable, we define its trajectory specified by a reference value and regulation error:

$$\begin{aligned} e_2 &= w_{\text{ref}} - w \\ \dot{e_2} &= \dot{w}_{\text{ref}} - \dot{w} \\ \dot{e_2} &= \dot{w}_{\text{ref}} - \left(\frac{P\varphi_f}{J} + \frac{P\left(L_d - L_q\right)}{J}i_d\right)i_q - \frac{f}{J}w - \frac{c_r}{J} \end{aligned} \tag{12}$$

Our objective requires the error e_2 to converge to zero, which is achieved by choosing i_q as the virtual control in equation (12). The Lyapunov function

$$\begin{aligned} V_2 &= V_1 + \frac{1}{2}e_2^2 \\ V_2 &= \frac{1}{2}\left(e_1^2 + e_2^2\right) \end{aligned} \tag{13}$$

By choosing $\dot{V}_2$ semi negative definite such that:

$$\dot{V_2} = -k_1e_1^2 - k_2e_2^2 \leq 0$$

We obtain:

$$k_2e_2 + \dot{w}_{ref} - \left(\frac{P\varphi_f}{J} + \frac{P\left(L_d - L_q\right)}{J}i_d\right)i_q + \frac{f}{J}w + \frac{c_r}{J} = 0 \tag{14}$$

Considering that $i_{dref} = 0$, this leads to defining the $i_{qref} = 0$ command necessary to determine the voltage V_{qref}.

$$i_{\text{qref}} = \left(k_2e_2 + \dot{w}'_{\text{ref}} + \frac{f}{J}w + \frac{c_r}{J}\right)\left(\frac{J}{P\varphi_f}\right) \tag{15}$$

3) Third Step - Calculation of the Final Control Law V_{qref}

This step aims to determine the reference voltage, with the new regulation objective being the current considered as the virtual control of this step. We define a new regulation error as follows:

$$\begin{aligned} e_3 &= i_{\text{qref}} - i_q \\ \dot{e}_3 &= \dot{i}_{\text{qref}} - i_q \\ \dot{e}_3 &= \dot{i}_{\text{qref}} - \left(\frac{-R_s}{L_q}i_q - \frac{L_d}{L_q}w_r i_d - \frac{\varphi_f}{L_q}w_r + \frac{1}{L_q}v_q\right) \end{aligned} \tag{16}$$

Since the objective requires this error to also converge to zero and demands that the current be regulated and limited, we use the extension of the Lyapunov function as follows:

$$\begin{aligned} V_3 &= V_1 + V_2 + \frac{1}{2}e_3^2 \\ V_3 &= \frac{1}{2}\left(e_1^2 + e_2^2 + e_3^2\right) \end{aligned} \tag{17}$$

The derivative of the function is written as follows:

$$\dot{V_3} = \dot{V_1} + \dot{V_2} + e_3\dot{e}_3 \tag{18}$$

$$\dot{V_3} = \dot{V_1} + \dot{V_2} + e_3\left(\dot{i}_{qref} - \left(\frac{R_s}{L_q}i_q - \frac{L_d}{L_q}w_r i_d - \frac{\varphi_f}{L_q}w_r + \frac{1}{L_q}v_q\right)\right) \tag{19}$$

By choosing $\dot{V}_3$ semi negative definite such that:

$$\dot{V_3} = -k_1e_1^2 - k_2e_2^2 - k_3e_3^2 \leq 0 \tag{20}$$

We obtain:

$$\begin{aligned} \dot{V_3} &= -k_1e_1^2 - k_2e_2^2 - k_3e_3^2 = -k_1e_1^2 - k_2e_2^2 + e_3\dot{e}_3 \\ e_3\dot{e}_3 &= -k_3e_3e_3 \\ \dot{e}_3 &= -k_3e_3 \end{aligned} \tag{21}$$

We obtain:

$$i_{\text{qref}} - \left(\frac{-R_s}{L_q}i_q - \frac{L_d}{L_q}w_r i_d - \frac{\varphi_f}{L_q}w_r + \frac{1}{L_q}v_q\right) = -k_3e_3 \tag{22}$$

We deduce the final control law $\dot{V}_{qref}$:

V_{aref} :

$$v_{\text{qref}} = L_q\left(\dot{i}_{\text{qref}} + \frac{R_s}{L_q}i_q - \frac{L_d}{L_q}w_r i_d - \frac{\varphi_f}{L_q}w_r + k_3e_3\right) \tag{23}$$

III. Comparative analysis of the proposed controls

To evaluate the two control techniques applied to our system in this document, we will conduct a comparative study between these methods. This comparison involves a series of tests:

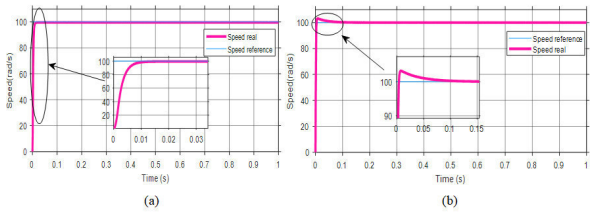


Fig. 2. Speed response, (a) Backstepping Controller, (b) Vector control based on PI controller

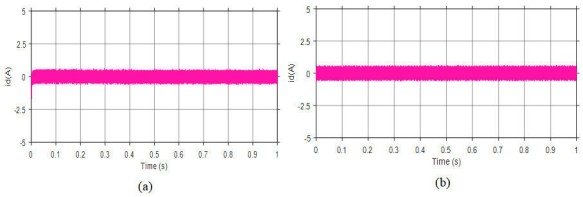


Fig. 3. Current d response, (a) Backstepping Controller, (b) Vector control based on PI controller

A. The initial test is conducted with a constant resisting torque of zero

B. The second test fig.5 , incorporating two criteria, includes monitoring the mechanical speed for the reversal of the direction of rotation and applying a resisting torque equal to 5N.m at t=0.2s

C. The final test fig.6, defining overspeed control, introduces the application of a resisting torque equal to 5N.m at t=0.2s

The simulation results presented the Figures 1, 2, 3, and 4 clearly indicate that both control schemes exhibit commendable performances. The measured speed closely follows its new reference, with low tracking

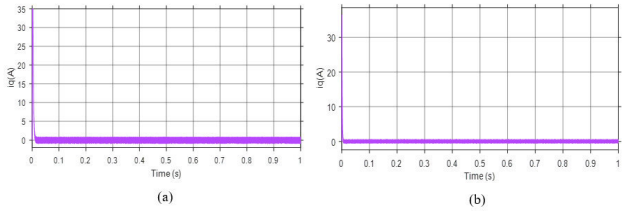


Fig. 4. Current q response, (a) Backstepping Controller, (b) Vector control based on PI controller

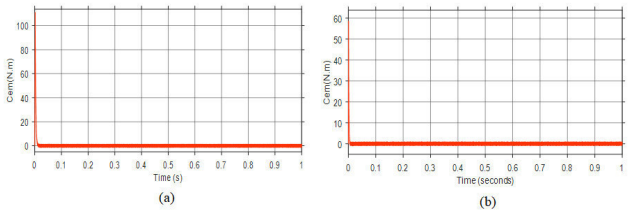


Fig. 5. Electromagnetic torque, (a) Backstepping Controller, (b) Vector control based on PI controller

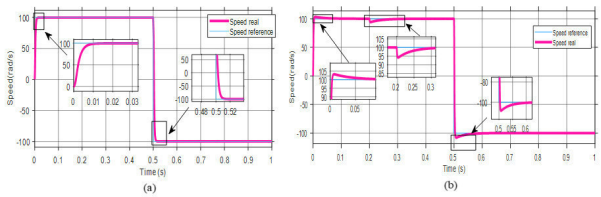


Fig. 6. Mechanical speed for reversal of direction of rotation, (a) Backstepping Controller, (b) Vector control based on PI controller

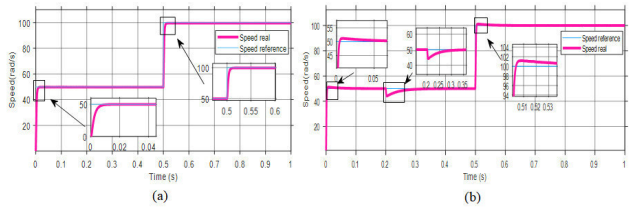


Fig. 7. Overspeed control, (a) Backstepping Controller, (b) Vector control based on PI controller

errors and acceptable exceedances. The response times, which characterize the transient regime and the change of instructions, are notably swift. Furthermore, a significant enhancement is observed in the results achieved by the backstepping controller when compared to vector control:

1) The backstepping controller demonstrates zero overshoot in the transient regime, contrasting with vector control where an overshoot is noticeable.

2) Regarding the responses of dq-axes current and electromagnetic torque, there is no significant difference between the two control methods.

The responses obtained in Fig.5 and 6 for both control types reveal the superior efficiency of the backstepping controller system in handling variations in resistive torque compared to vector control. It is noteworthy that:

1) Upon the application of the load (Cr=5Nm at t=0.2 s), a decrease in speed is evident in the vector control.

2) The backstepping controller exhibits no overshoot during the transient phase, in contrast to vector control where overshooting is evident.

IV. Conclusion

In this study, we conducted a comparative analysis of two control methods applied to the Permanent Magnet Synchronous Machine to evaluate their efficiency and performance under varying parametric and external conditions. The investigation focused on three key criteria: without resisting torque (Cr=0N.m), with resisting torque (Cr=5N.m), and with a change in speed. The results indicate that the backstepping controller system proves to be the most efficient and effective

in controlling our system compared to vector control, demonstrating superior performance in the presence and absence of parametric and external variations.

REFERENCES

[1] B. Adhavan, A. Kuppuswamy, G. Jayabaskaran, and V. Jagannathan, "Field oriented control of permanent magnet synchronous motor (pmsm) using fuzzy logic controller," in *2011 IEEE Recent Advances in Intelligent Computational Systems*. IEEE, 2011, pp. 587–592.

[2] A. A. Hagras, "Nonlinear adaptive extended state space predictive control of permanent magnet synchronous motor," *International Transactions on Electrical Energy Systems*, vol. 29, no. 1, p. e2677, 2019.

[3] W. Song, Y. Gang, Q. Zhi-Jian, S. Shuang-shuang, and C. Chao, "Identification of pmsm based on ekf and elman neural network," in *2009 IEEE International Conference on Automation and Logistics*. IEEE, 2009, pp. 1459–1463.

[4] T. Li, X. Sun, G. Lei, Y. Guo, Z. Yang, and J. Zhu, "Finite-control-set model predictive control of permanent magnet synchronous motor drive systems—an overview," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 12, pp. 2087–2105, 2022.

[5] M. Huang, Y. Deng, H. Li, M. Shao, and J. Liu, "Integrated uncertainty/disturbance suppression based on improved adaptive sliding mode controller for pmsm drives," *Energies*, vol. 14, no. 20, p. 6538, 2021.

[6] L. B. Freidovich and H. K. Khalil, "Lyapunov-based switching control of nonlinear systems using high-gain observers," *Automatica*, vol. 43, no. 1, pp. 150–157, 2007.

[7] Y. E. Mourabit, H. Salime, B. Bossoufi, S. Motahhir, A. Derouich, S. Mobayen, and A. Zhilenkov, "Enhanced performance in pmsg-based wind turbine systems: Experimental validation of adaptive backstepping control design," *Energies*, vol. 16, no. 22, p. 7481, 2023.

[8] Y. Bensalem, A. Abbassi, R. Abbassi, H. Jerbi, M. Alturki, A. Albaker, A. Kouzou, and M. Abdelkrim, "Speed tracking control design of a five-phase pmsm-based electric vehicle: A backstepping active fault-tolerant approach," *Electrical Engineering*, vol. 104, no. 4, pp. 2155–2171, 2022.

[9] F. Wang, J.-m. Wang, K. Wang, Q. Zong, and C. Hua, "Adaptive backstepping sliding mode control of uncertain semi-strict nonlinear systems and application to permanent magnet synchronous motor," *Journal of Systems Science and Complexity*, vol. 34, no. 2, pp. 552–571, 2021.

[10] G. Bramerdorfer, G. Lei, A. Cavagnino, Y. Zhang, J. Sykulski, and D. A. Lowther, "More robust and reliable optimized energy conversion facilitated through electric machines, power electronics and drives, and their control: State-of-the-art and trends," *IEEE Transactions on Energy Conversion*, vol. 35, no. 4, pp. 1997–2012, 2020.

[11] Y. Yi, D. M. Vilathgamuwa, and M. A. Rahman, "Implementation of an artificial-neural-network-based real-time adaptive controller for an interior permanent-magnet motor drive," *IEEE transactions on industry applications*, vol. 39, no. 1, pp. 96–104, 2003.

[12] H.-K. Hoai, S.-C. Chen, and C.-F. Chang, "Realization of the neural fuzzy controller for the sensorless pmsm drive control system," *Electronics*, vol. 9, no. 9, p. 1371, 2020.

[13] H. Celik and T. Yigit, "Field-oriented control of the pmsm with 2-dof pi controller tuned by using pso," in *2018 international conference on artificial intelligence and data processing (IDAP)*. IEEE, 2018, pp. 1–4.

[14] J. Sabrine and N. Khaled, "A comparative study on the pmsm control system using the nonlinear dynamic inversion method," in *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*. IEEE, 2018, pp. 490–494.

Improved sensorless backstepping controller using model reference adaptive system of a Permanent Magnet Synchronous Motors

1st SLIMANI Amira
Electrical Engineering
Laboratory of Biskra LGEB
University of Biskra
Biskra,Algeria
amira.slimani@univ-biskra.dz

2nd *BOUREK Amor*
Electrical Engineering
Laboratory of Biskra LGEB
University of Biskra
Biskra,Algeria
a.bourek@univ-biskra.dz

3rd *AMMAR Abdelkarim*
Signals and Systems Laboratory
of Boumerdes LSS
University of Boumerdes
Boumerdes,Algeria
a.ammar@univ-boumerdes.dz

4th *DAHNOUN Ilyes*
Electrical Engineering
Laboratory of Biskra LGEB
University of Biskra
Biskra,Algeria
ilyasdahnoun@gmail.dz

5th *HATTAB Wassila*
Electrical Engineering
Laboratory of Biskra LGEB
University of Biskra
Biskra,Algeria
wassila.hattab@univ-biskra.dz

***Abstract*— To address the issue of applying permanent magnet synchronous motors (PMSM) without sensors in certain environments, and accounting for the uncertainty in load torque, we propose a sensorless model reference adaptive system control for PMSM based on the backstepping method. firstly, we shed the light on non-linear backstepping control, offering a comprehensive perspective on its impact on the stability and robust performance of PMSMs. employing Lyapunov functions, Secondly, the rotor speed and position are extracted by the MRAS, enhances system reliability, improves control performance, and reduces costs. Finally, the simulation is executed using MATLAB/Simulink. The results show the sensorless control algorithm demonstrates proficient capability in accurately estimating both rotor speed. the system has promising for applications demanding reliable and adaptive sensorless control in the face of varying loads.**

Keywords— backstepping control, Lyapunov functions, MRAS, permanent magnet synchronous motor.

I. INTRODUCTION

Permanent magnet synchronous motors (PMSMs) have established widespread attention and research because of their compact structure, high torque-to-current ratio, high power density, good fault tolerance, and high work efficiency [1]-[2]. Currently, PMSMs find extensive applications in various sectors, including new energy power generation systems, electric vehicles, aerospace engines, and intelligent robots.

The common control methods of PMSM drive systems are field-oriented control (FOC) [3], and direct torque control (DTC) [4]-[5], and direct flux control (DFC) technique [6]. Compared to the FOC and the DFC, the DTC offers better torque response and is considered a more appropriate control algorithm for PMSM [7]. But the traditional Direct Torque Control (DTC) faces challenges, notably variable switching frequencies, fluctuations in torque, flux, and current, particularly at low speeds. These issues could be addressed by employing modern control technology. Therefore, scholars persist in exploring advanced and effective control strategies to enhance the performance of (PMSM) drive systems, such as sliding mode control (SMC) [8], backstepping control, adaptive control [9], linearization control [10]. predictive control [11]. and intelligent control [12], which improve the performance of motor control manifests across various facets. Among these control methods, backstepping presents promisingly alternative methods for controlling nonlinear systems by combining the choice of the Lyapunov function with the control laws which allows the overall stability of the system [13]. Various speed observers are employed in Permanent Magnet Synchronous Motors (PMSM) for accurate and efficient speed estimation. One commonly utilized method is the Extended Kalman Filter (EKF), which leverages a state-space representation to estimate the rotor speed and position. Another approach involves the Model Reference Adaptive System (MRAS), which adapts the system parameters to closely match the actual motor behavior for accurate speed estimation. Additionally, Sliding Mode Observers (SMOs) are applied to mitigate the effects of uncertainties and disturbances in PMSM systems. Each of these speed observers has its advantages and is chosen based on specific application requirements, system constraints, and the desired balance between accuracy and computational complexity. Among those model reference adaptive observer is

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

widely used because ofits simple mathematical model, strong robustness.

In this paper, a sensorless control approach for Permanent Magnet Synchronous Motors (PMSM) is proposed, combining the Model Reference Adaptive System (MRAS) with the backstepping method. The integration of MRAS allows the system to adapt and estimate the rotor position and speed without the need for direct sensors, addressing the challenges associated with sensor costs and complexities. The backstepping method is employed to design a controller that enhances the overall stability and performance of the sensorless control system. Through a series of simulations and experiments, the effectiveness of the sensorless MRAS control strategy is demonstrated, showcasing improved dynamic performance and accurate observation of rotor parameters. This approach holds promise for applications where sensor installation is impractical or cost-prohibitive, offering a viable solution for achieving precise control in Permanent Magnet Synchronous Motors.

The structure of this paper is outlined as follows. In the second section, we present the mathematical model of the (PMSM). The third section focuses on the construction of the backstepping controller. In fourth section, is devoted to the design of sensor-less control utilizing the Model Reference Adaptive System (MRAS) method. The results of the simulations demonstrate that the control scheme exhibits superior dynamic performance and achieves higher observation accuracy in fifth section, which is followed by conclusion in sixth section.

II. DRIVE MODEL:

The backstepping strategy is applied to a (PMSM) powered by a two-level voltage source inverter (2L-VSI). This section provides a mathematical model for both the power converter and the electrical machine.

A. *Power Converter:*

In applications involving medium and low power rate drives, the 2L-VSI power converter is the most widely used. The converter produces the voltage that feeds the machine's stator. Eight voltage vectors, six active vectors, and two zero vectors are produced by the 2L-VSI

The power converters voltage vector in a stationary αβ-frame is:

$$v_{s\,\alpha,\beta} = v_{dc}.\frac{2}{3}\left(S_a + aS_b + a^2 S_c\right) \quad (1)$$

where v_{dc} is the dc-link voltage, $\mathbf{S} = [S_a\ S_b\ S_c]$ are the switching state of the converter, and $a = e^{j\frac{2\pi}{3}}$

Next, in a synchronous *d-q* frame oriented with the rotor angle of the PMSM θ_r is, the power converter voltage is

$$v_s = v_{s\,\alpha,\beta}.e^{-j\theta_r} \quad (2)$$

B. *PMSM Model:*

The mathematical model of the (PMSM) in the rotating (d-q) reference frame can be expressed as:

$$\begin{bmatrix} U_{sd} \\ U_{sq} \end{bmatrix} = \begin{bmatrix} L_d & 0 \\ 0 & L_q \end{bmatrix}\begin{bmatrix} \dot{I_{sd}} \\ \dot{I_{sq}} \end{bmatrix} + \begin{bmatrix} R_s & -L_q\omega_e \\ L_d\omega & R_s \end{bmatrix}\begin{bmatrix} I_{sd} \\ I_{sq} \end{bmatrix} + \begin{bmatrix} 0 \\ \varphi_f\ \omega_e \end{bmatrix} \quad (3)$$

Where U_{sd} and U_{sq} are the stator input voltage, R_s is stator resistance, $I_{sd}\ and\ I_{sq}$ are the stator currents, φ_f is flux linkages, L_d and L_q are the stator inductances, respectively. Additionally, ω_e is the rotor electrical angular speed, calculated as $(\omega_e = \mathrm{p} * \omega\ )$, where p is pole pairs, and ω_m is the rotor mechanical angular speed.

The PMSM's dynamical model can be written as follows:

$$\frac{d\omega}{dt} = \frac{1}{J}(Te - B\omega\ \ - T_L) \quad (4)$$

Where B is the friction factor, J is the inertia of PMSM T_L is the load torque, T_e is the electromagnetic torque, which can be derived as.

$$Te = \frac{3}{2}I_{sq}p\left[\varphi_f + \left(L_d - L_q\right)i_d\right] \quad (5)$$

In particular, for a surface-mounted PMSM where L_d equals L_q, the electromagnetic torque can be rewritten as:

$$Te = \frac{3\,p\varphi_f}{2}I_{sq} \quad (6)$$

III. DESIGN OF BACKSTEPPING CONTROLLER

The fundamental idea of the backstepping approach is the use of the so-called "virtual control" quantities to break down a complex non-linear control design problem into simpler and smaller ones [13]. The Backstepping Controller for a Permanent Magnet Synchronous Motor (PMSM) involves a systematic process aimed at achieving stable and efficient control of the motor's dynamics.

A. *Step1: Stator current reference determination I_{sq}:*

According to the reference rotor angular speed ω^*, define the error signal $e_1 = \omega^* - \omega$, The dynamic response of the speed error is [14] .

$$\dot{\boldsymbol{e}}_{\mathbf{1}} = \dot{\boldsymbol{\omega}}^* + \dot{\boldsymbol{\omega}} = \ \dot{\boldsymbol{\omega}}^* - \frac{\mathbf{3}\boldsymbol{p}\boldsymbol{\varphi}_{\boldsymbol{f}}}{\mathbf{2}\boldsymbol{J}} + \frac{\boldsymbol{B}}{\boldsymbol{J}}\boldsymbol{\omega} + \frac{\boldsymbol{T}_{\boldsymbol{L}}}{\boldsymbol{J}} \quad (7)$$

In order to satisfy the tracking performance criteria, the initial Lyapunov function linked to the speed errors is expressed as (8). This is accomplished by employing equation (7).

$$\begin{cases} V_1 = \dfrac{1}{2}{e_1}^2 \\ \dot{V_1} = e_1\dot{e_1} \end{cases} \quad (8)$$

The dynamics of the Lyapunov function are expressed as follows:

$$\dot{V}_1 = e_1\left(\dot{\omega}^* - \frac{3p\varphi_f}{2J} + \frac{B}{J}\omega + \frac{T_L}{J}\right) \quad (9)$$

Equation (8) can be expressed such as:

$$\dot{V}_1 = -k_1 {e_1}^2 \quad (10)$$

Where k_1 must be positive parameter, in array to guarantee a stable tracking, so

$$\dot{e}_1 = \dot{\omega^*} + \dot{\omega} = -k_1 e_1 \quad (11)$$

Equation (11) allow the generation of current reference to ensure the satisfaction of the Lyapunov stability condition. The current reference is represented as follows:

$$i_{sq}{}^* = \left(\dot{\omega}^* + \frac{B}{J}\omega + \frac{T_L}{J} + k_1 e_1\right)\frac{2J}{3p\varphi_f} \quad (12)$$

The virtual control in equation (12) is selected to fulfil the objective of stator current regulation and is regarded as a reference for the next step.

B. Step2: Stator voltages reference determination :

the current errors as follows:

$$\begin{cases} e_2 = I_{sq}{}^* - I_{sq} \\ e_3 = I_{sd}{}^* - I_{sd} \end{cases} \quad (13)$$

Then, accounting to (11), (12) will take the following form:

$$\begin{cases} e_2 = \left(\dot{\omega}^* + \dfrac{B}{J}\omega + \dfrac{T_L}{J} + k_1 e_1\right)\dfrac{2J}{3p\varphi_f} - \mathrm{I}_{sq} \\ e_3 = \mathrm{I}_{sd}{}^* - \mathrm{I}_{sd} \end{cases} \quad (14)$$

Then:

$$\begin{cases} \dot{e}_2 = \dot{I_{sq}}{}^* - \dot{I_{sq}} \\ \dot{e}_3 = \dot{I_{sd}}{}^* - \dot{I_{sd}} \end{cases} \quad (15)$$

Following the substitutes of (3) in (15), The dynamics of current errors can be expressed as follow:

$$\begin{cases} \dot{e}_2 = \dot{I_{sq}}{}^* + \left(\frac{R_s}{L_q} I_{sq} + p\omega\, I_{sd} + \frac{p\varphi_f}{L_q}\omega - \frac{U_{sq}}{L_q}\right) \\ \dot{e}_3 = \frac{R_s}{L_d} I_{sd} - p\omega\, I_{sq} - \frac{U_{sd}}{L_d} \end{cases} \quad (16)$$

The Lyapunov function for the second subsystem is chosen as:

$$V_2 = \frac{{e_1}^2 + {e_2}^2 + {e_3}^2}{2} \quad (17)$$

$$\begin{cases} \dot{V}_2 = e_1\dot{e}_1 + e_2\dot{e}_2 + e_3\dot{e}_3 \\ \dot{V}_2 = -k_1{e_1}^2 - k_2{e_2}^2 - k_3{e_3}^2 + f_1 e_2 + f_2 e_3 \end{cases} \quad (18)$$

where:

$$\begin{cases} f_1 = k_2 e_2 + \dot{I_{sq}}{}^* + \left(\frac{R_s}{L_q} I_{sq} + p\omega\, I_{sd} + \frac{p\varphi_f}{L_q}\omega - \frac{U_{sq}}{L_q}\right) \\ f_1 = k_3 e_3 + \frac{R_s}{L_d} I_{sd} - p\omega\, I_{sq} - \frac{U_{sd}}{L_d} \end{cases} \quad (19)$$

The dynamic Lyapunov function V2 could exhibit negative definiteness if f_1 and f_2 are chosen to be equal to zero.

$$\begin{cases} \mathrm{k_2 e_2} + \dot{\mathrm{I}}_{\mathrm{sq}}{}^* + \left(\dfrac{\mathrm{R_s}}{\mathrm{L_q}}\mathrm{I_{sq}} + \mathrm{p}\omega\,\mathrm{I_{sd}} + \dfrac{\mathrm{p}\varphi_\mathrm{f}}{\mathrm{L_q}}\omega - \dfrac{U_{sq}}{\mathrm{L_q}}\right) = 0 \\ k_3 e_3 + \dfrac{R_s}{L_d} I_{sd} - p\omega\,\mathrm{I}_{sq} - \dfrac{U_{sd}}{L_d} = 0 \end{cases} \quad (20)$$

The stator voltages control can be deduced as follows:

$$\begin{cases} U_{s_q} = L_q k_2 e_2 + L_q \dot{I_{sq}}{}^* + R_s L_q + L_q p\omega\, I_{sd} + p\varphi_f\omega \\ U_{sd} = R_s I_{sd} - p\omega L_d + k_3 e_3 L_d \end{cases} \quad (21)$$

The structural block diagram representing the sensorless MRAS of PMSM based on the backstepping method is depicted in Figure 1. The control mode with $I_{sq}{}^* = 0$ is implemented on the permanent magnet synchronous motor and the control parameter is specifically chosen as follows: k_1= 1400, k_2= 1450 and k_3= 120.

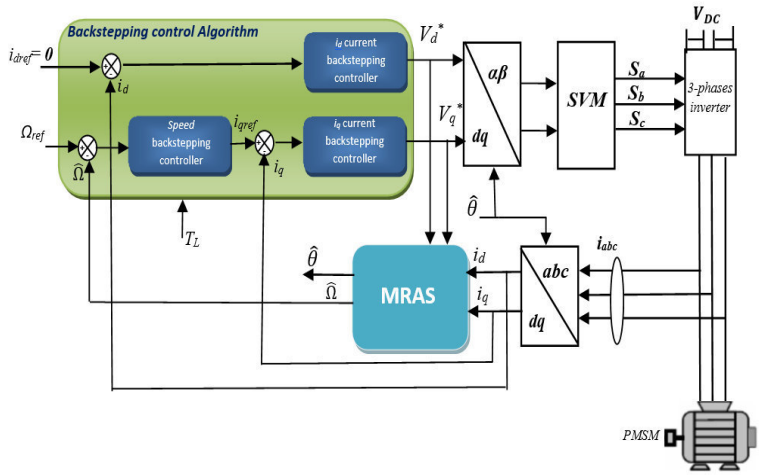


Fig. 1. Block diagram of PMSM control system.

IV. MODEL REFERENCE ADAPTIVE SYSTEM (MRAS):

Model Reference Adaptive System (MRAS) is a widely utilized methodology in control system design, particularly in the field of adaptive control. The primary goal of MRAS is to make a dynamic system behave like a specified reference model, even in the presence of uncertainties or varying operating conditions. It achieves this by continuously adjusting the parameters of the system based on the error between the actual system output and the output of the reference model.

The equation for estimating speed is formulated as follows:

$$\hat{\omega} = \int_0^t k_p[(I_{sd}\widehat{I_{sq}} - I_{sq}\widehat{I_{sd}} - \frac{\varphi_f}{L_q}(I_{sq} - \widehat{I_{sq}}))]dt + k_i[(I_{sd}\widehat{I_{sq}} - I_{sq}I_{sd}\frac{\varphi_f}{L_q}(I_{sq} - \widehat{I_{sq}}))] + \hat{\omega}(0) \quad (22)$$

And the equation for estimating position is expressed as:

$$\hat{\theta} = \int_0^t \hat{\omega}\, dt \quad (23)$$

Where k_p and k_i are the proportional and integral coefficients.

Figure 2 illustrates a comprehensive Simulink model of the MRAS.

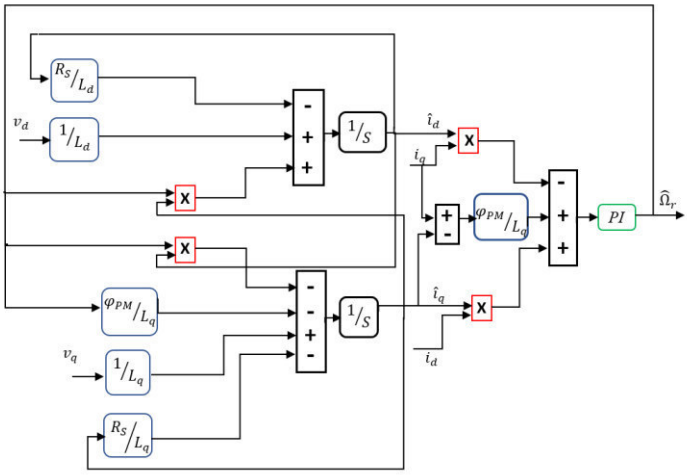


Fig. 2. Block diagram of MRAS

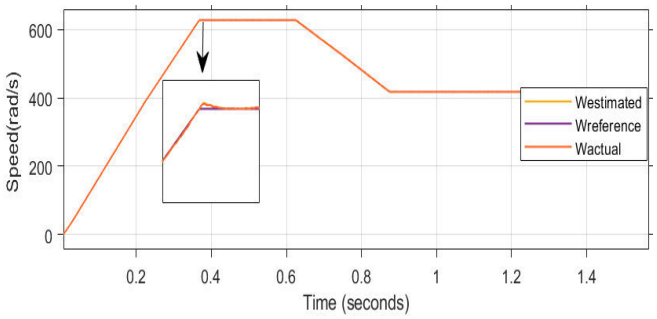
V. SIMULATION RESULTS AND DISCUSSION:

Simulation is carried out on a PMSM in the Matlab/Simulink environment. This simulation evaluates the sensorless of the backstepping control using MRAS. The nominal parameters of the PMSM used are presented in table I

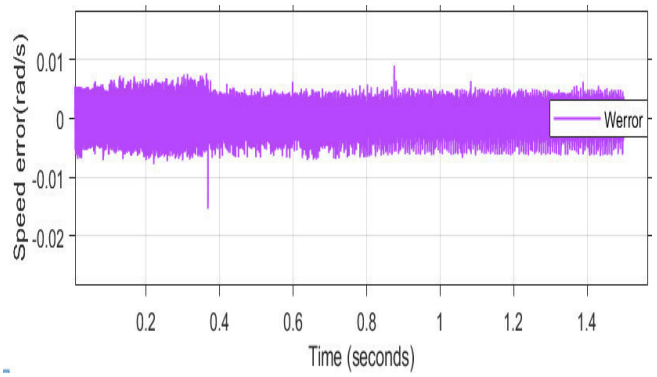
TABLE I. PMSM NOMINAL PARAMETERS USED IN NUMERICAL SIMULATION.

Parameters	Values
Stator inductance L_d, L_q (H)	0.0076
Stator resistance Rs (Ω)	2.3
Friction coefficient B (N.m.s)	0.000169
Moment of inertia J ($kg.\ m^2$)	0.0032
Flux linkages φ_f (wb)	0.4
Number of pole pairs p	4

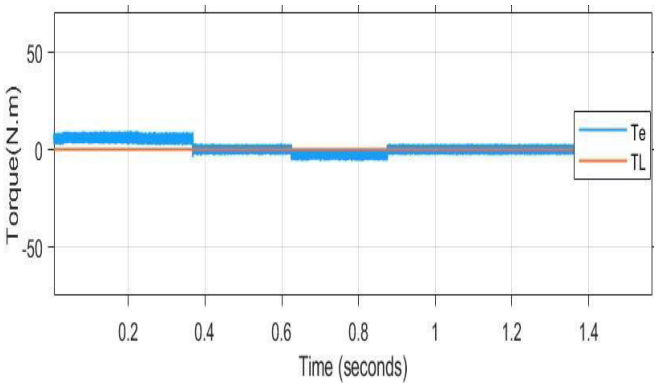
In the simulation, the reference speed signals are set to 628rad/s and 418 rad/s. The results of the simulation under no load conditions are presented in figures 3 (a, b, c, d and e). It is obvious that the Model Reference Adaptive System exhibits global convergence, ensuring minimal overshoot in speed changes. Moreover, the speed error between the estimated and actual values remains within an acceptable range. The torque T_e, and I_{sq} current can quickly converge to the stable values and the d-axis current is maintained at zero with small oscillations. In addition, Three-phases I_a, I_b and I_c stator currents can be small with no-load.



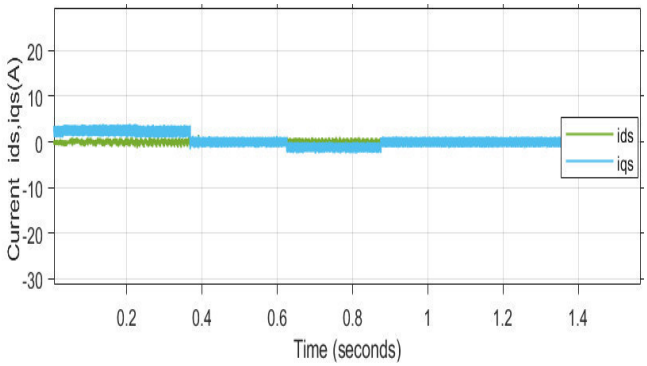
(a). Reference, actual and estimated rotor speed



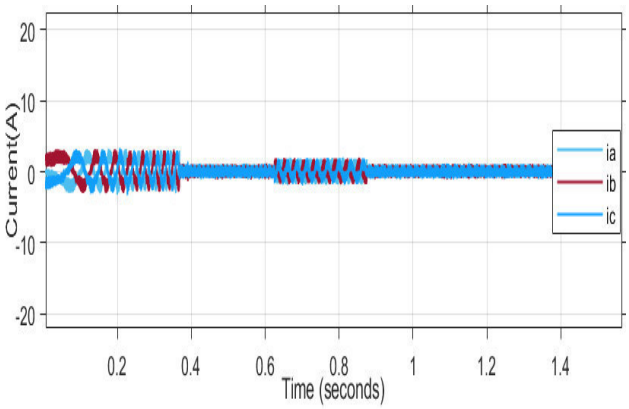
(b). The error between actual and estimated value of speed.



(c). electromagnetic and load torque.



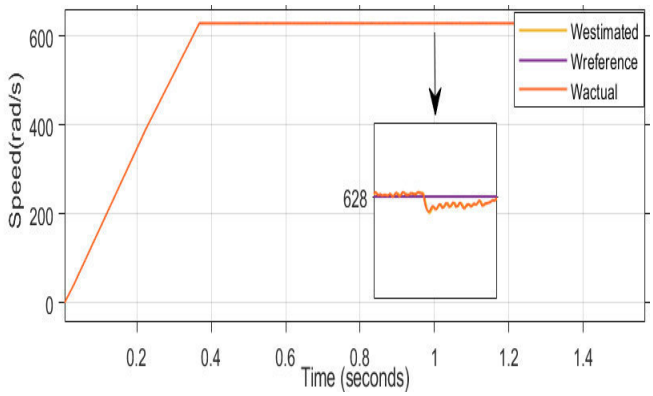
(d). (d-q) axes current response



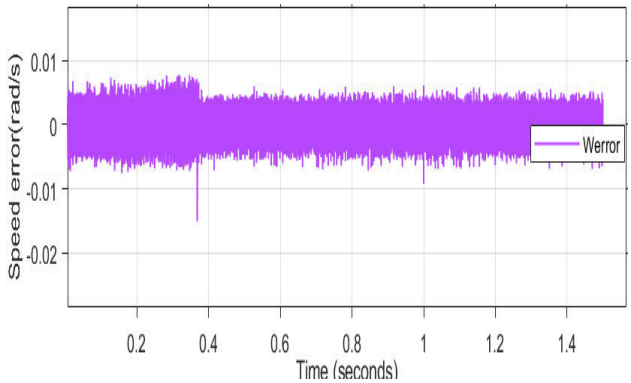
(e). Three-phases ia, ib and ic stator currents.

Fig. 3. System performance with Speed change from 628rad/s to 418 rad/s without load torque

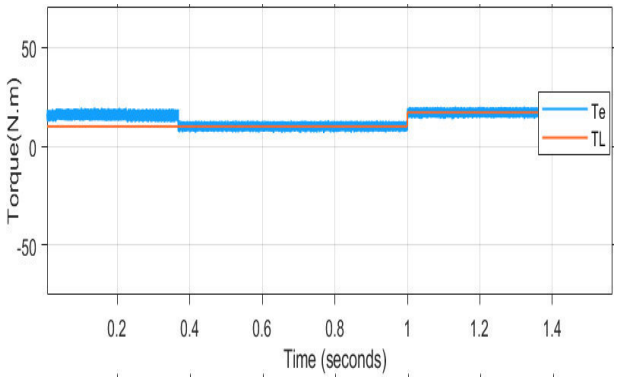
To assess the robustness of the proposed sensorless control algorithm, During the operation of the motor 628 (rad/s) for a duration of 1 seconds, the load torque is given to 17N.m Before was 10 (N.m). The simulation results are presented in Figures 4 (a, b, c, d and e). It is evident that after addition of the the load. The motor speed exhibits some oscillation after the load is applied. However, the error between the estimated value of the MRAS and the value actual of remains within an acceptable range. The proposed control effectively manages the motor speed, swiftly bringing it back to the designated reference value. Figure 4(b) depicts the speed error between the estimated and actual values confirming the algorithm's fast response and robust performance. Therefore, the MRAS control algorithm proposed based on the backstepping method demonstrates effective estimation of the rotor speed in a permanent magnet synchronous motor.



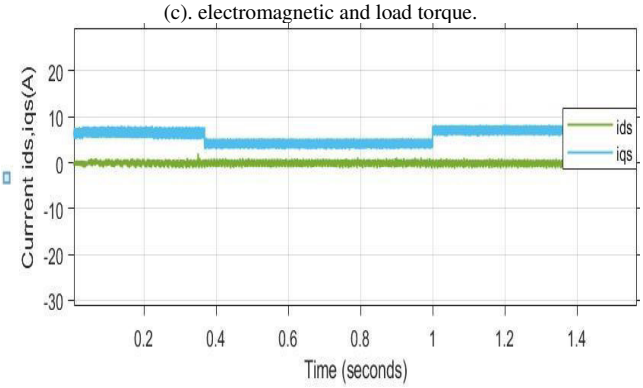
(a). Reference, actual and estimated rotor speed



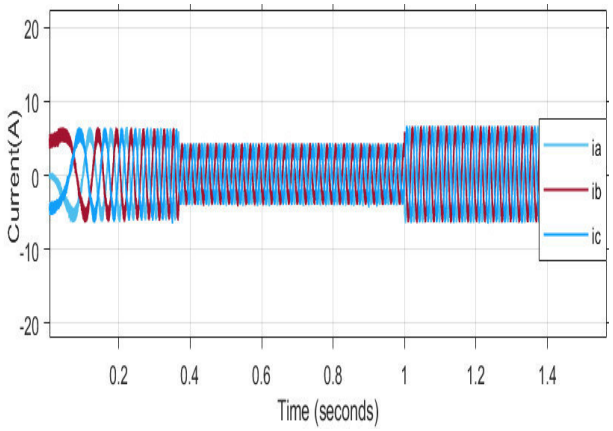
(b). The error between actual and estimated value of speed.



(c). electromagnetic and load torque.



(d). (d-q) axes current response



(e). Three-phases ia, ib and ic stator currents.

Fig. 4. System performance with load torque charge from 10(N.m) to 17(N.m)

VI. CONCLUSION:

In this paper, a MRAS control of PMSM based on backstepping is proposed. Firstly, the comprehensive evaluation of the control is performance reveals its efficacy in accurately estimating the rotor speed of the permanent magnet synchronous motor. The combination of MRAS and backstepping proves to be a potent strategy, ensuring precise control and estimation in sensorless operation. The algorithm's ability to maintain rotor speed within acceptable ranges underscores its practical utility. Then the proposed control demonstrates exceptional robustness, particularly in the presence of varying loads and uncertainties. The control strategy effectively mitigates the impact of load torque uncertainties, contributing to stable and reliable motor operation. Finally, its notable features include a rapid speed response and enhanced robustness, making it a promising solution for applications demanding reliable and adaptive sensorless control in the face of varying loads.

REFERENCES

[1] Chen, H.; He, J.; Demerdash, N.A.O.; Guan, X.; Lee, C.H.T. Diagnosis of Open-Phase Faults for a Five-Phase PMSM Fed by a Closed-Loop Vector-Controlled Drive Based on Magnetic Field Pendulous Oscillation Technique. IEEE Trans. Ind. Electron. 2021, 68, 5582–5593.

[2] Rassõlkin, A.; Kallaste, A.; Orlova, S.; Gevorkov, L.; Vaimann, T.; Belahcen, A. Re-Use and Recycling of Different Electrical Machines. Latv. J. Phys. Tech. Sci. 2018, 55, 13–23

[3] Pedrosa, D.; Carvalho, J.; Gonçalves, H.; Monteiro, V.; Fernandes, A.; Afonso, J.L. Field oriented control of an axial flux permanent magnet synchronous motor for traction solutions. In Proceedings of the IECON 2014—40th Annual Conference of the IEEE Industrial Electronics Society, Dallas, Texas, USA, 29 October 2014; pp. 1466–1472.

[4] S. Vaez-Zadeh, “Analysis of a DTC with back emf oriented voltage for PMS motor drives,” IEEE Trans. Energy Convers., vol.33, no. 3, pp.1594–1596, Sept. 2018.

[5] X. D. Sun, K. K. Diao, G. Lei, Y. G. Guo, and J. G. Zhu, “Direct torque control based on a fast modeling method for a segmented-rotor switched reluctance motor in HEV application,” IEEE J. Emerg. Sel. Topics Power Electron., vol.9, no.1, pp.232–241, Feb. 2021.

[6] Emanuele, G.; Marco, P.; Fabio, C.; Matthias, N.; Francesco, C.; Francesco, G. Detection of stator turns short-circuit during sensorless operation by means of the Direct Flux Control technique. In Proceedings of the 2020 AEIT International Annual Conference (AEIT), Catania, Italy, 23–25 September 2020; pp. 1–6

[7] K. Kakouche, A. Oubelaid, S. Mezani, D. Rekioua, and T. Rekioua, "Different Control Techniques of Permanent Magnet Synchronous Motor with Fuzzy Logic for Electric Vehicles: Analysis, Modelling, and Comparison," Energies, vol. 16, no. 7, p. 3116, 2023.

[8] L. Yuan, Y. Jiang, L. Xiong, and P. Wang, “Sliding Mode Control Approach with Integrated Disturbance Observer for PMSM Speed System,” CES Trans. Electr. Mach. Syst., vol. 7, no. 1, pp. 118–127, 2023, doi: 10.30941/CESTEMS.2023.00009.

[9] H. N. Tran and J. W. Jeon, “Robust Speed Controller Using Dual Adaptive Sliding Mode Control (DA-SMC) Method for PMSM Drives,” IEEE Access, vol. 11, no. June, pp. 63261–63270, 2023, doi: 10.1109/ACCESS.2023.3288124.

[10] Y. Feng, H. Zhao, M. Zhao, and H. Chen, “A Feedback Linearization Control Scheme Based on Direct Torque Control for Permanent Magnet Synchronous Motor,” Chinese Control Conf. CCC, vol. 2018-July, pp. 87–92, 2018, doi: 10.23919/ChiCC.2018.8483803

[11] X. F. Wang, X. C. Fang, S. Lin, F. Lin, and Z. Yang, “Predictive common-mode voltage suppression method based on current ripple for permanent magnet synchronous motors,” IEEE J. Emerg. Sel. Top. Power Electron., vol.7, no.2, pp.946–955, Jun. 2019.

[12] Lin CH, Lin CP (2013) The hybrid RFNN control for a PMSM drive electric scooter using rotor flux estimator. Appl Mech Mater 145(3):542–546

[13] L. Dongliang, Z. Xiehui, and C. Lili, (2011). Backstepping control of speed sensorless permanent magnet synchronous motor,Transactions of China Electrotechnical Society, vol. 26, no. 9, pp. 67-72.

[14] A. Bennassar, A. Abbou, M. Akherraz, M. Barara, (2013).Sensorless Backstepping Control Using an Adaptive Luenberger Observer with Three Levels NPC Inverter, International Journal of Electrical, Robotics, Electronics and Communications Engineering, Vol.7, No.8, pp. 612-618.

Design of a Super Twisting Sliding Mode Controller for a *LC* filtered Three-Phase Inverter

Houssam Eddine Mansouri
ETA Laboratory, Departement of Electronics.
University Mohamed El Bachir El Ibrahimi of Bordj Bou Arreridj
Bordj Bou Arreridj, Algeria
houssameddine.mansouri@univ-bba.dz

Billel Talbi
ETA Laboratory, Departement of Electronics.
University Mohamed El Bachir El Ibrahimi of Bordj Bou Arreridj
Bordj Bou Arreridj, Algeria
billel.talbi@univ-bba.dz

Idris Messaoudene
ETA Laboratory, Departement of Electronics.
University Mohamed El Bachir El Ibrahimi of Bordj Bou Arreridj
Bordj Bou Arreridj, Algeria
idris.messaoudene@univ-bba.dz

Moussa Abderrahim Mehiris
ETA Laboratory, Departement of Electronics.
University Mohamed El Bachir El Ibrahimi of Bordj Bou Arreridj
Bordj Bou Arreridj, Algeria
moussaabderrahim.mehiris@univ-bba.dz

***Abstract*— This paper introduces a novel nonlinear super twisting sliding mode control (ST-SMC) strategy for grid-forming converters operating within a droop control framework. As a second-order system, the three-phase voltage source inverter with an *LC* filter presents challenges for traditional sliding mode control due to potential chattering. The ST-SMC technique mitigates these challenges by incorporating robust feedback linearization principles. A graphical method based on Lyapunov's direct method is employed to design the switching gain, ensuring effective disturbance rejection. The study demonstrates that the implementation of ST-SMC significantly enhances the frequency response of the power-frequency (P-f) loop, leading to faster dynamics. Additionally, the proposed control method improves both the transient and steady-state performance of the grid-forming converter.**

Keywords—Grid-forming, voltage source inverter, supper twisting sliding mode control (ST-SMC), LC filter.

I. INTRODUCTION

With the increasing integration of inverter-based generation, ensuring converter stability is crucial for maintaining overall grid stability [1]. Grid-forming (GFM) converters have gained significant traction in industry applications due to their ability to regulate voltage and frequency, mimicking the behavior of traditional synchronous generators in power networks [2]. The primary approaches to GFM converter control can be categorized into four main types [3]: (a) virtual synchronous machines (VSM), (b) Type A droop control, (c) Type B droop control, and (d) virtual oscillator control. Among these methods, droop control is well-established, with Type A droop control being particularly prevalent in microgrid applications [4].

Droop control strategies are further divided into single-loop and multi-loop configurations, based on their internal control mechanisms [5]. This study utilizes Type A droop control with conventional droop characteristics [6], where active and reactive power inputs determine the system's angular frequency and voltage magnitude outputs, respectively, as depicted in Figure 1. In contrast to single-loop systems, multi-loop droop control incorporates a dedicated current control element, typically implemented with a PI regulator. While this addition enhances current regulation, it can reduce the stability margin of the power-frequency (P-f) droop loop [5,7], particularly in weak grids experiencing significant frequency fluctuations.

To address these challenges, advanced techniques such as nonlinear model predictive control have been introduced to enhance the frequency response of droop-based GFM converters [8]. Furthermore, various sophisticated control strategies [9-12] aim to improve dynamic performance and stability across a wide range of operating conditions. However, these methods often face limitations due to constraints in selecting controller gains while preserving system stability.

Sliding mode control (SMC) provides greater adaptability compared to traditional methods. Motivated by this advantage, the study focuses on developing and analyzing a nonlinear super twisting sliding mode control (ST-SMC) approach to enhance the dynamic and frequency response of droop-based grid-forming converters (GFMs) equipped with *LC* filters. Higher-order designs are favored over lower-order variants due to their ability to minimize switching frequency effects and facilitate state feedback control.

This research addresses several the design of the proposed ST-SMC methodology which is tailored grid-forming converter (GFM) equipped with *LC* filter.

The study begins by establishing a mathematical model for grid-connected converters with *LC* filter. It then delves into ST-SMC theory, detailing the design of voltage and current controllers using this technique. A thorough control system model is subsequently derived, followed by Lyapunov stability analysis to define appropriate control law. Finally, various contingency scenarios are simulated to evaluate the performance advantages of the ST-SMC controller over its conventional counterpart. The paper is organized as follows: Section 2 presents the mathematical model of the three-phase Grid forming inverter system. Section 3 details the development of the proposed ST-SMC control strategy in the natural reference frame. Section 4 provides simulation results obtained using MATLAB® and Simulink®, demonstrating the effectiveness of the proposed control method. Finally, Section 5 concludes the paper with a summary of the key findings and contributions.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

II. Mathematical Model of Grid-Forming Inverter

A. Inverter model

A three-phase voltage source inverter (VSI) is employed, comprising six bidirectional switches, each consisting of an Insulated Gate Bipolar Transistor (IGBT) and an anti-parallel diode. This inverter is interfaced with the three-phase grid through an *LC* filter, as illustrated in Fig. 1. The voltage vector produced by the inverter can be expressed as a function of the DC-link voltage and the switching states, as detailed in equation [13]:

$$v_i = V_{dc}\left[\frac{2}{3}\left(S_a + e^{j2\pi/3}S_b + e^{j4\pi/3}S_c\right)\right] \quad (1)$$

where V_{dc} is the DC supply voltage, S_a , S_b and S_c are the switching states of the inverter.

B. Three phase grid-forming voltage source inverter

The dynamic behavior of the grid-forming inverter (Fig.1) is described by a mathematical model expressed in the natural frame. This model, essential for controller design, is given by the following equations:

$$\begin{cases} L_1 \dfrac{di_{ip}}{dt} = -r_1 i_{ip} + v_{ip} - v_{op} \\ C \dfrac{dv_{op}}{dt} = i_{ip} - i_{op} \qquad p = a,b,c \end{cases} \quad (2)$$

Where:

- v_{op} is the *abc* components of the filter output voltage.
- i_{ip} is the *abc* components of the inverter and filter output currents, respectively.
- C is the filter capacitance.
- $L1$ is the inverter-side filter inductances, with *r1 as* its equivalent resistance.

III. Controller Design

This section delves into the design and implementation of the ST-SMC strategy. The goal is to compel the output voltage to track a reference signal in the presence of measured and unknown disturbances. The desired reference signal is expressed as:

$$v_{op}^{*} = E\sin(\omega t + \theta_p) \quad (3)$$

Where θ_a = 0°, θ_b = 120° and θ_c = 240°, E is the peak voltage and ω is the nominal angular frequency.

Let the sliding function and its derivative be defined as

$$\begin{cases} \sigma_p = v_{op}^{*} - v_{op} \\ \dot{\sigma}_p = \dot{v}_{op}^{*} - \dot{v}_{op} \end{cases} \quad (4)$$

The core component of Sliding Mode Control (SMC) is the sliding manifold, which dictates the system's dynamic behavior.

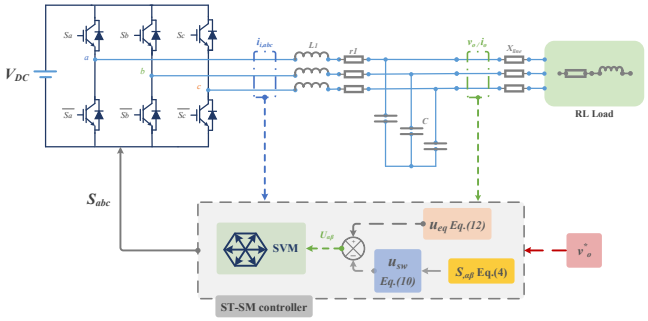


Fig. 1. Three-phase grid-forming inverter with the suggested ST-SMC controller.

The proposed sliding manifold and its derivative are defined as follows:

$$\begin{cases} S_p = k_d\sigma_p + \dot{\sigma}_p \\ \dot{S}_p = k_d\dot{\sigma}_p + \ddot{\sigma}_p \end{cases} \quad (5)$$

The derivative from equation (5) can be further expanded to:

$$\dot{S}_p = \ddot{v}_{op}^{*} + k_d\dot{v}_{op}^{*} + \frac{1}{C}\left(\frac{di_{op}}{dt} + k_d i_{op}\right) + \frac{1}{L_1 C}v_{cp} - \frac{1}{L_1 C}v_{ip} - \frac{k_d}{C}i_{ip} \quad (6)$$

The global control law U_p is defined as the sum of a continuous control known as the equivalent control u_{eq} and the switching control term u_{sw} as follows:

$$U_p = u_{eqp} + u_{swp} \quad (7)$$

The switching term guarantees the convergence of the system dynamics to the sliding surface, while the equivalent term is responsible for sustaining the system's motion on this surface, using the condition $\dot{S}_p = 0$ The latter is derived as:

$$u_{eqp} = CL_1\ddot{v}_{op}^{*} + CL_1 k_d\dot{v}_{op}^{*} + L_1\left(\frac{di_{op}}{dt} + k_d i_{op}\right) + v_{cp} - L_1 k_d i_{ip} \quad (8)$$

In accordance with ST-SMC, the switching control term (u_{sw}) is designed to implement the super-twisting algorithm. This term comprises two distinct components:

$$u_{swp} = u_{1p} + u_{2p} \quad (9)$$

where :

$$\begin{cases} u_{1p} = \alpha\left|S_p\right|^{\lambda} sign(S_p) \\ \dot{u}_{2p} = \beta sign(S_p) \end{cases} \quad (10)$$

In this context, the parameter λ is a positive constant typically ranging from 0 to 1. The signum function, denoted by *sign*, is defined as follows:

$$sign(S) = \begin{cases} -1 \ if \ \ S < 0 \\ 0 \ \ if \ \ S = 0 \\ 1 \ if \ \ S > 0 \end{cases} \quad (11)$$

Additionally, the convergence to the sliding surface can be ensured by selecting appropriate values for the parameters k_1 and k_2, as suggested in reference [16].

$$\alpha > \frac{\delta}{\eta}\ ,\ \beta^2 \geq \frac{4\delta\eta_{Max}(k_1+\delta)}{\eta^2\eta_{\min}(k_1-\delta)} \tag{12}$$

For linearly dependent systems, the super-twisting control equation can be simplified further by introducing the constants η and δ. In such cases, the switching control term, u_{sw}, can be expressed as:

$$u_{swp} = \alpha\left|S_p\right|^{\lambda} sign(S_p) + \beta\int sign(S_p) \tag{13}$$

From Equation (10) and, the final control law can finally be derived as follows:

$$U_p = v_{cp}\left(\frac{L_1}{L_2}+1\right) - \frac{L_1}{L_2}v_{op} + CL_1\ddot{v}^*_{cp} + CL_1k_d\dot{\sigma}_p - u_{swp} \tag{14}$$

The control law derived in Equation (13) is subsequently applied to the inverter using Space Vector Modulation (SVM). This strategy minimizes the error between the actual and reference grid currents, ensuring stable system operation.

The parameters of the controller were determined through a trial-and-error optimization process, and their values are listed in Table I.

For the effective design of the proposed ST-SMC, the following Lyapunov function is considered:

$$\begin{cases} V = \frac{1}{2}S^2 \\ \dot{V} = \dot{S}.S \end{cases} \tag{15}$$

Next, to meet the Lyapunov stability criteria, which requires $\dot{V} \leq 0$, we consider:

$$\dot{V} = S[k_d\dot{\sigma}_p + \ddot{\sigma}_p]$$

$$= \ddot{v}^*_{op} + k_d\dot{v}^*_{op} + \frac{1}{C}\left(\frac{di_{op}}{dt} + k_d i_{op}\right) + \frac{1}{L_1C}v_{cp} - \frac{1}{L_1C}U_p - \frac{k_d}{C}i_{ip} \tag{16}$$

Substituting Equation (14) into Equation (16) gives:

$$\dot{V} = \alpha\sqrt{\left|S_p\right|}S_p * sign(S_p) + \beta\int S_p * sign(S_p) \tag{17}$$

Since S_p*$sign\ (S_p) > 0$, $\alpha > 0$ and $\beta > 0$, then selecting α and β large enough would dictate $\dot{V} < 0$. Equation (12) can be used to select the minimum values of α and β. Thereafter, these values can easily be tuned for the optimum performance which makes the closed-loop system stable.

IV. Simulation Results and Discussion

Numerical simulations using MATLAB®/Simulink® environment are carried out to confirm the performance of the suggested controller. System parameters are listed in Table I.

The simulation results of active power P and reactive power Q are presented in Fig. 2. the active and reactive power requested by the load amounts to 1500W and 300 VAR respectively. At t= 0.06 s a higher load is connected resulting in a total of 5.7 kW of active power and a 1.8 kVAR. It can be seen that the ST-SMC and droop controllers are successful at delivering the requested power.

TABLE I. System Parametars

Parameter	*Value*
DC link voltage, *Vdc*	700 V
Filter inductance L_1	2.5 mH
Filter capacitance C	50 µF
Nominal angular frequency ω	2*π*50 rad/s
Peak nominal voltage E	$230\sqrt{2}$
SVM switching frequency f_{SVM}	15 kHz
Sampling time (Fixed-step)	30 µs
Gains k_d, α, β	3e3, 13, 5e9

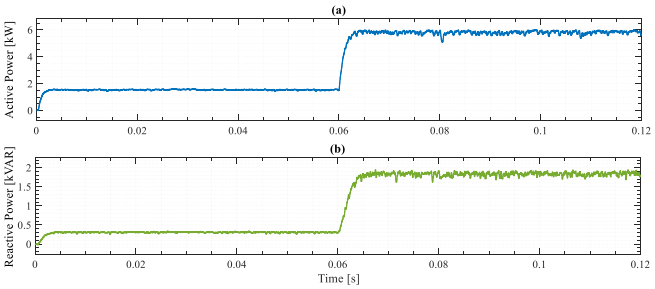


Fig. 2. Simulation results of (a) Active power and (b) Reactive power delivered to the load.

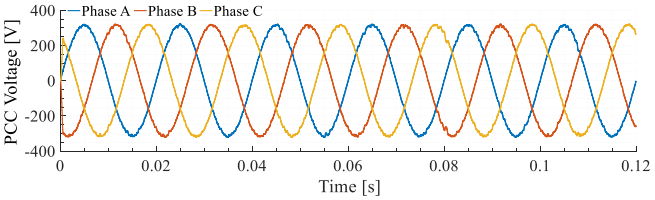


Fig. 3. Three phase voltage at point of common coupling.

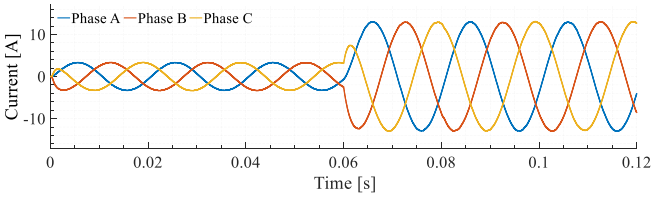


Fig. 4. Three phase output current.

The simulated responses of the output voltage and current are plotted in Fig. 3 and Fig 4 respectively. The phase-a output THD% is found to be 0.84% while the voltage is 1.04%. Fig. 7 displays the obtained simulation results when the filter inductance and capacitance L_1 and C are varied within the

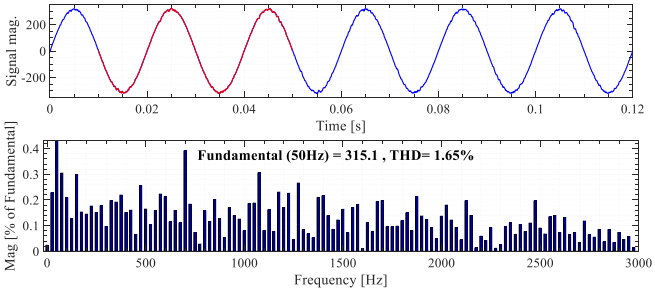


Fig. 5. FFT window and THDv% of phase a voltage at PCC.

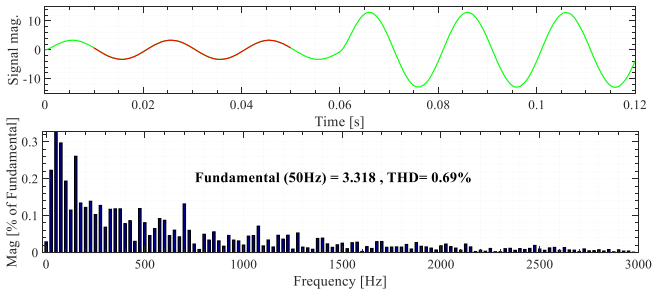


Fig. 6. FFT window and THDi% of phase a load current.

specified range.the tests were conducted during high load operation. It can be noticed that these variations have minimal impact on the output voltage THDv%. These results are very promising and demonstrate that the suggested ST-SMC controller exhibits low sensitivity to changes in these parameters.

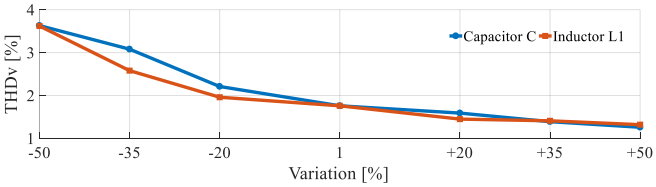


Fig. 7. Phase a output voltage THDv% results with grid filter parameters varaitions (C and L_1).

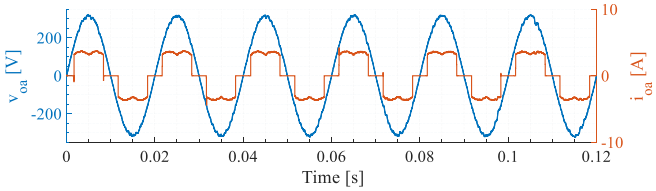


Fig. 8. Simulation results for the phase-a output voltage and current during non-linear load scinario.

A non-linear load scenario was investigated, the output voltage and current are displayed in Fig. 8, The phase-a output voltage v_{oa} THDv% during non-linear load test scenario is found 1.58%.

V. Conclusion

This paper presents a novel super twisting sliding mode control ST-SMC strategy for three-phase grid-forming inverters. The primary objective of this control approach is to mitigate harmonic distortion and simplify the controller design while eliminating the chattering phenomena. A detailed mathematical model of the inverter system is developed, followed by the formulation of the ST-SMC controller. The proposed controller ensures asymptotic stability and enables the injection of sinusoidal currents into the load. The effectiveness of the proposed control strategy is validated through MATLAB®/Simulink® simulations. Simulation results demonstrate the robustness, dynamic performance, and steady-state characteristics of the ST-SMC controller, leading to a significant reduction in high-frequency harmonics and the achievement of active and reactive power injection operation.

References

[1] Ndreko, M., Rüberg, S., Winter, W.: Grid forming control scheme for power systems with up to 100% power electronic interfaced generation: a case study on Great Britain test system. In: IET Renewable Power Generation,vol. 14, pp. 1268–1281. Institution of Engineering and Technology,London (2020).

[2] I. Phurailatpam, C., Flynn, D.: Requirements and grid standards for grid-forming inverters. In: Grid-Forming Power Inverters: Control and Applications, pp. 15–38. CRC Press, Boca Raton, FL (2023).

[3] R Ramasubramanian, D., Wang, W., Farantatos, E., Huque, M.: Toward performance-based requirements and generic models for grid forming inverters. In:Grid-Forming Power Inverters: Control and Applications, pp. 55–77. CRC Press, Boca Raton, FL (2023)

[4] Du, W., Liu, Y., Tuffner, F.K., Huang, R., Huang, Z.: Model specification of droop-controlled, grid-forming inverters (GFMDRP_A) (2021).

[5] Du,W., Chen, Z., Schneider, K.P., Lasseter, R.H., Pushpak Nandanoori, S.,Tuffner, F.K., et al.: A comparative study of two widely used grid-forming droop controls on microgrid small-signal stability. IEEE J. Emerging Sel. Top. Power Electron. 8(2), 963–975 (2020).

[6] Olivares, D.E., Mehrizi Sani, A., Etemadi, A.H., Cañizares, C.A., Iravani,R., Kazerani, M., et al.: Trends in microgrid control. IEEE Trans.Smart Grid 5(4), 1905–1919 (2014).

[7] Pogaku, N., Prodanovi´c, M., Green, T.C.: Modeling, analysis and testing of autonomous operation of an inverter-based microgrid. IEEE Trans. Power Electron. 22(2), 613–625 (2007).

[8] Samanta, S., Lagoa, C., Chaudhuri, N.R.: Nonlinear model predictive control for droop-based grid forming converters providing fast frequency support. IEEE Trans. Power Delivery 39(2), 790–800 (2023).

[9] Xiao, H., He, H., Zhang, L., Liu, T.: Adaptive grid-synchronization basedgrid-forming control for voltage source converters. IEEE Trans. PowerSyst. 39(2), 4763–4766 (2023).

[10] Samanes, J., Rosado, L., Gubia, E., Lopez, J., Perez, M.A.: Deadbeat voltage control for a grid-forming power converter with LCL filter. In: IEEE Transactions on Industry Applications, vol. 59, Institute of Electricaland Electronics Engineers Inc., pp. 2473–2482. IEEE, Piscataway, NJ (2023).

[11] Zhao, Z., Han, Z., Liu, X., Yao, J., Ji, B., Wang, S., et al.: Optimal tuning of the current loop for dual-loop controlled grid-forming converters based on active damping optimization. IEEE Access 9, 35801–35813 (2021).

[12] Wu, H., Wang, X.: Passivity-based dual-loop vector voltage and current control for grid-forming VSCs. IEEE Trans. Power Electron. 36(8), 8647–8652 (2021).

[13] Hou, Q.; Ding, S.; Yu, X. Composite super-twisting sliding mode control design for pmsm speed regulation problem based on a novel disturbance observer. IEEE Trans. Energy Convers. 2020, 36, 2591–2599.

[14] S. Bayhan and H. Abu-Rub, “A simple control technique for distributed generations in grid-connected and islanded modes,” IEEE 27th International Symposium on Industrial Electronics (ISIE). IEEE, 2018, pp. 1237–1242.

Examining the Stability of a Coupled Axial-Torsional Drilling System Across Diverse Operating Scenarios

Samira Hadibi
Laboratory of Applied Automation, Department of Automation Faculty of Hydrocarbons and Chemistry, University M'hamed Bougara
Boumerdes, Algeria
s.hadibi@univ-boumerdes.dz

Madjid Kidouch
Laboratory of Applied Automation, Department of Automation Faculty of Hydrocarbons and Chemistry, University M'hamed Bougara
Boumerdes, Algeria
kidouche_m@hotmail.com

Mohamed Zinelabidine Doghmane line
Bob L. Herd Department of Petroleum Engineering
Texas Tech University
Lubbock 79401, Texas, USA, Country
doghmane_m@yahoo.com

***Abstract*—This In this study, we examine the stability of stick-slip and bit-bounce phenomena encountered in rotary drilling operations using a fully coupled two-degrees-of-freedom model. The effects of varying operating conditions on the stability of the coupled axial-torsional system are assessed by analyzing the decay rate of the system's response, with the Hilbert transform employed to extract the signal envelope. Numerical simulations further illustrate how these operating conditions influence drilling stability.**

Keywords—Torsional vibration, axial vibration, coupled models, vibration stability

I. INTRODUCTION

The phenomenon of self-excited vibrations in drill strings has been a longstanding research subject, primarily aimed at mitigating adverse effects such as early fatigue of drill pipes and premature failure of drill bits. Drill strings predominantly exhibit axial, torsional, and lateral vibrations. In extreme cases, these vibrations manifest as axial bit-bounce, with the drill bit intermittently losing contact with the bottom; lateral whirl motions, causing the drill bit to deviate from its central axis; and torsional stick-slip motions, resulting in the abrupt halting of the drill string's rotational motion. This research seeks to understand and control these vibration modes to enhance the operational efficiency and longevity of drilling equipment.

Understanding the stability boundaries of drilling systems is crucial for designing effective control strategies to mitigate equipment damage and reduce well costs. Several studies have explored the complex dynamics of axial-torsional vibrations in rotary drilling. Richard et al [1] used the RGD model to analyze self-excited vibrations caused by time delay and feedback, linking instability to phenomena such as bit-bounce and stick-slip. Subsequent research refined this model, incorporating axial dissipation, compliance, and damping to improve stability predictions. Germay et al [2] highlighted the influence of axial dynamics on mean torque, while [3] demonstrated that axial vibrations exacerbate torsional instability. Advanced models, such as those by [4], incorporated high degrees of freedom to study chaotic behavior and stability variations with system parameters. Other works, including those by [5-7], expanded upon these findings by integrating bit-rock interaction nonlinearities, state-dependent delays, and bifurcation analyses, revealing key factors affecting stability. Studies Zheng et al [8-10], emphasized the roles of drill string length, penetration rate, and rock properties in shaping stability boundaries. Despite these advancements, challenges remain in predicting stability ranges for coupled systems, as highlighted by Jakob et al., due to the inherent complexity of distributed axial-torsional dynamics.

Here we will use a Hilbert transform to study the stability of the system without need to linearize it. Based directly on the system response including the nonlinear interactions, we will establish a law to detect the damping or the increasing oscillation of both the axial and torsional responses over a range of operating parameters. The Hilbert transform is a popular approach in signal analysis for determining instantaneous properties of a time series, particularly the envelope amplitude which is the complex Hilbert transform's amplitude and instantaneous frequency, referring to the rate at which the instantaneous phase angle changes over time. These qualities can be used to detect the dynamic characteristics of both linear and nonlinear systems [11, 12].

This study focuses on the stability analysis of coupled axial-torsional vibrations in a drilling system, employing a novel approach based on the decay rate percentage of the system's dynamic response. By utilizing the Hilbert Transform, we effectively extract the envelope of oscillatory signals, enabling a precise evaluation of stability under various operating conditions. This method captures the influence of system nonlinearities, providing a comprehensive understanding of how these nonlinear interactions impact stability across a wide range of parameters. The findings offer valuable insights for optimizing drilling performance and mitigating instability-related issues.

II. DRILL-STRING SYSTEM MODELING

A. The Model

Here, a two-degree-of-freedom system is presented, where U(t) and Φ(t) denote the axial and angular displacements of the drill string, respectively:

$$\begin{aligned} M\ddot{U} + C_a\dot{U} + K_aU &= -Wob \\ I\ddot{\varphi} + C_t\dot{\varphi} + K_t\varphi &= -Tob \end{aligned} \quad (1)$$

Where I is the moment of inertia, M is the drill string mass, Ca and Ct is the axial and torsional drill pipe damping

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

respectively, Ka and Kt is the axial and torsional drill pipe stiffness respectively.

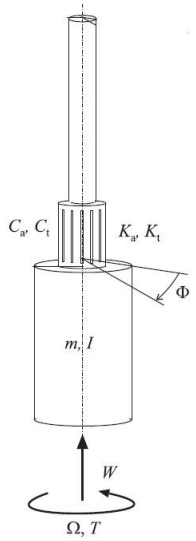


Fig. 1. A diagram of the drill-string model.

B. Bit Rock Interaction Model

A variety of torque and weight-on-bit functions have been proposed in the literature to investigate stick-slip vibrations in models that couple torsional and axial dynamics [1, 13-14]. In this study, the model presented in [1] is employed for further analysis.

$$\begin{cases} Wob = \zeta a \varepsilon d(t) H(\dot{\varphi}) H(d(t)) - \sigma a l \\ Tob = \dfrac{1}{2}(a^2 \varepsilon d(t) H(\dot{\varphi}) H(d(t)) + a^2 \mu \gamma \sigma l \operatorname{sgn}(\dot{\varphi}) H(\dot{U})) \end{cases} \quad (2)$$

In which, ξ is cutter inclination coefficient, σ is the contact pressure, γ is the geometry parameter of drill bit, a is the radius of bit, μ is the friction coefficient and l is the wear flat length.

Here, the depth of cut d(t) is defined as in [15]:

$$d(t) = d_0 \sin(n_b \varphi) \quad (3)$$

Where d0 is the initial depth of cut and nb is a constant.

III. Stability Analysis

A. Hilbert Transforms

The Hilbert transform (HT) differs from other integral transforms like Fourier and Laplace in that it does not include domain transformation. The algorithm assigns a corresponding imaginary portion to a real part or vice versa by shifting each component signal by one-quarter of a period. The HT pair can determine a signal's amplitude and frequency in real-time [16].

The HT of the function x(t) is defined by an integral transform:

$$H[x(t)] = \tilde{x}(t) = \pi^{-1} \int \frac{x(\tau)}{t - \tau} d\tau \quad (4)$$

The aforementioned mathematical integral expression provides little information about the HT's knowledge and application. Nevertheless, the physical significance of the HT enables us to obtain a far deeper understanding of the transition. On a physical level the HT is comparable to a specific form of linear filter, wherein every single amplitudes of the spectral components are maintained intact, only their phases are altered by π/2.

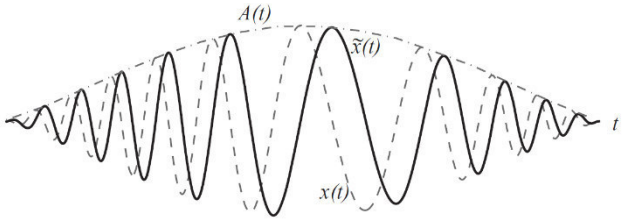


Fig. 2. x(t) the initial signal, $\tilde{x}(t)$ the HT pair projection, A(t)the envelope.

B. Decay Rate

Consider the signal :

$$y(t) = A_0 e^{-\gamma t} \sin(\omega_0 t) \quad (5)$$

where A_0 is the starting amplitude, ω_0 is the sinusoid frequency, and γ is the damping coefficient that quantifies energy dissipation. The signal's envelope follows an exponential decay determined by the damping rate. This free decay, y(t), is a well-known oscillatory function with an amplitude that gradually diminishes to zero over time. The time-dependent amplitude, or envelope, is expressed as

$$A(t) = A_0 e^{-\gamma t} \quad (6)$$

This envelope illustrates how the amplitude of the oscillations varies with time. Whenever the signal is oscillating with diminishing amplitude, the envelope will gradually decrease over time.

To detect the decreasing amplitude, the natural logarithm (log) is applied to linearize the expression. If the envelope decreases significantly, its logarithm will display a linear trend over time, described by:

$$\log(A(t)) = \log(A_0) - \gamma t \quad (7)$$

where γ represents the decay rate:

- Negative slope indicates envelope decay (damped oscillation).
- A zero or positive slope denotes the signal's amplitude is not decreasing (the signal may be constant or increasing).

IV. Results And Discussion

In this study, a MATLAB code was developed to compute the analytic signal of the axial and angular responses of the system using the Hilbert Transform (HT). Stability maps were constructed by analyzing the decay rate percentage of the responses. For the axial stability map, the decay rate was calculated over a range of axial damping parameters (C_a) while maintaining a fixed torsional damping value. Conversely, the torsional stability map was generated by varying the torsional damping (C_t) parameter while keeping the axial damping constant. As illustrated in “fig. 3” and “fig. 4”.

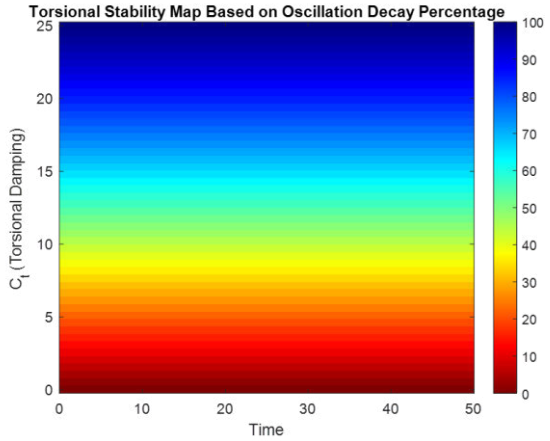


Fig. 3. Stability map of the torsional response with C_a=10^6

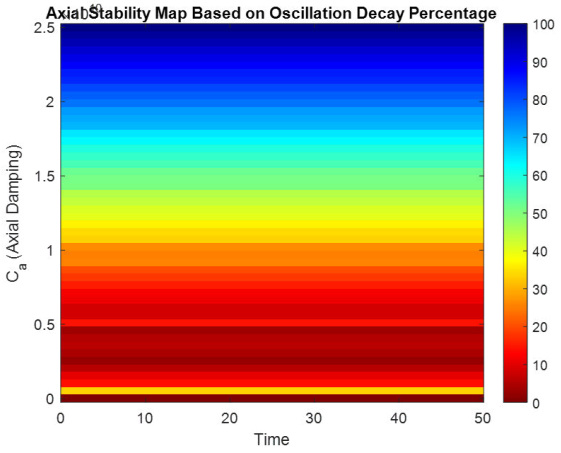


Fig. 4. Stability map of the axial response with C_t=1.

The axial stability map reveals pronounced oscillatory behavior for axial damping values below 10^8. In contrast, for axial damping values exceeding 1.8×10^8 , the system stabilizes more rapidly, as illustrated in "Fig. 3". Conversely, the torsional stability map, "Fig. 4" indicates that stabilization occurs only for torsional damping values greater than 10.

V. Conclusion

Note The application of the Hilbert Transform provided a straightforward and efficient method to extract critical information about the oscillatory behavior of a coupled axial-torsional drilling system. By analyzing the variations in the decay rate of the system's response, we identified specific ranges of axial and torsional damping parameters where the oscillations dissipate rapidly, leading to a stable system. This approach not only facilitated a clear understanding of the stabilization dynamics but also enabled the precise characterization of damping conditions required for optimal performance in drilling operations.

References

[1] T. Richard, C. Germay and E. Detournay, "Self-excited stick–slip oscillations of drill bits,"Journal of Petroleum Science and Engineering 59 (2007) 289–299

[2] C. Germay, N. V. de Wouw, H. Nijmeijer, and R. Sepulchr, "Nonlinear Drill-string Dynamics Analysis," SIAM J. APPLIED DYNAMICAL SYSTEMS c_ Vol. 8, No. 2, pp. 527–553

[3] B. Besselink, N. van de Wouw and H. Nijmeijer, "A semi-analytical Study of Stick-Slip Oscillations in Drilling Systems," Journal of Journal of Computational and Nonlinear Dynamics, April 2011, DOI: 10.1115/1.4002386

[4] X. Liu, N. Vlajic, X. Long, G. Meng and B. Balachandran, "State-Dependent Delay Influenced Drill-String Oscillations and Stability Analysis," Journal of Vibration and Acoustics, October 2014, Vol.136/051008-1

[5] K. Nandakumar, M. Wiercigroch and C. Pearson, "Bit-Bounce and Stick-Slip in Drill-String Dynamics," DOI: 10.1007/978-94-007-5742-4_25, 2013

[6] F. Bakhtiari-Nejad and A. Hosseinzadeh, "Nonlinear dynamic stability analysis of the coupled axial-torsional motion of the rotary drilling considering the effect of axial rigid-body dynamics," International Journal of Non-Linear Mechanics, S0020-7462(16)30255-4, October 2016

[7] S. K.Gupta and P. Wahi, "Global axial–torsional dynamics during rotary drilling," Journal of Sound and Vibration, May 2016, DOI:10.1016/j.jsv.2016.04.021

[8] X. Zheng and B. Balachandrana, "State-Dependent Delay and Drill-String Dynamics, " Procedia IUTAM 22 (2017) 31 – 38

[9] A. Hosseinzadeh and F. Bakhtiari-Nejad, "A New Dynamic Model of Coupled Axial–Torsional Vibration of a Drill String for Investigation on the Length Increment Effect on Stick–Slip Instability," Journal of Vibration and Acoustics, December 2017, Vol. 139 / 061016-1

[10] P. Deng, A. Zhang, K. Fu and H. Li, "a Nonlinear Vibration of a Time-Space Coupled Drill String System Based on the Surface Morphology of Rock," Journal of Sound and Vibration 506 (2021) 116153

[11] Luo, Huageng, X. Fang, and B. Ertas. "Hilbert transform and its engineering applications." AIAA journal 47.4 (2009): 923-932.

[12] Abdulmunem, Matheel E., and A. A. Badr. "Hilbert Transform and its Applications: A survey." Int. J. Sci. Eng. Res 8.2 (2017).

[13] Navarro-López, E. M., & Suárez, R, "Practical approach to modelling and controlling stick-slip oscillations in oilwell drillstrings," In Proceedings of the 2004 IEEE International Conference on Control Applications, September 2004, (Vol. 2, pp. 1454-1460). IEEE

[14] Yigit, A. S., & Christoforou, A. P, "Coupled torsional and bending vibrations of actively controlled drillstrings," In Journal of sound and vibration, 2000, (234(1), 67-83).

[15] Li, Xinye, et al, "Theoretical and Numerical Analysis of Coupled Axial-Torsional Nonlinear Vibration of Drill Strings," In Shock and Vibration (1-15), 2021,

[16] "Hilbert Transform Applications in Mechanical Vibration, First Edition. Michael Feldman," 2011 John Wiley & Sons, Ltd. Published 2011 by John Wiley & Sons, Ltd. ISBN: 978-0-470-97827-6K.

Trajectory Tracking of Mobile Manipulator Based on Second Order Sliding Mode Control

Samah Riache
Department of Electrical Systems Engineering,
Faculty of Technology
M'Hamed Bougara University,
Boumerdes, Algeria.
samahriache@univ-boumerdes.dz

Madjid Kidouche
Applied Automation Laboratory, Faculty of Hydrocarbons and Chemestry;
M'Hamed Bougara University,
Boumerdes, Algeria
mkidouche@univ-boumerdes.dz

Issam, Younes Hmadache
Department of Electrical Systems Engineering,
Faculty of Technology
M'Hamed Bougara University,
Boumerdes, Algeria.
youneshamadache513@gmail.com

***Abstract*— This paper focuses on trajectory tracking control of a wheeled robot integrated with a two-degree-of-freedom robotic manipulator, based on second order Sliding Mode Control (SMC). The dynamic model was formulated by treating the mobile manipulator as a coupled system. The controller included a Super Twisting Algorithm to maintain a high tracking performance and reduce chattering effects. Simulation results were carried out to validate the performance and advantages of the proposed controller.**

Keywords—Mobile Manipulator, Wheels, Trajectory, Super Twisting,

I. INTRODUCTION

Nowadays, mobile manipulators have attained significant relevance across a multitude of applications, encompassing fields such as agriculture, industry, construction, and others [1-3]. These systems have been designed utilizing a variety of locomotion mechanisms, which include wheeled configurations, legged structures, underwater robotic platforms, and aerial robotics.

A wheeled mobile manipulator (WMM) is characterized by a wheeled mobile base, upon which a manipulator affixed. This robotic system is recognized as a non-holonomic uncertain system, thereby rendering its modeling and control processes significantly challenging.

Various methods have been employes to solve the aforementioned problems, including nonlinear feedback control [4], input-output decoupling control [5], task space and null-space decoupling control [6], and computed torque control [7]. However, these control schemes depend on the knowledge of the complex dynamics of the system and ignore external disturbances.

An adaptive robust output-feedback force/motion control strategy was presented in [8], under non-holonomic constraints in the presence of uncertainties and disturbances. An adaptive Neural Network (NN)-based control for the arm and the base was proposed in [9], for joint-space position control of a mobile manipulator. In [10], an adaptive model predictive control (AMPC) was implemented for a two-wheeled mobile robot. In [11], an optimized differential evolution algorithm based on kinematic limitations and structural complexity constraints was proposed to solve the trajectory tracking problem for a mobile manipulator robot. Adaptive control strategies were presented in [12], for a WMM operating in task space under external torques and disturbances.

Sliding mode control (SMC) is a robust control strategy widely applied to nonlinear systems, enhancing their stability and trajectory tracking capabilities under various conditions. This approach is particularly effective in managing uncertainties and external disturbances, making it suitable for diverse operational environments it can be applied to various robotic systems [13-15]. In [16], finite time SMC method was applied to WMM under uncertain tire/road structure TRs. A comparative study between conventional and dual closed loop finite time SMC was conducted in [17]. Later, in [18], the authors proposed an adaptive trajectory tracking control of the manipulator based on extreme learning machine and sliding mode control ELM-SMC.

The main drawback of SMC is the chattering effect, for which various solutions have been proposed, such as adaptive laws in the discontinuous control part [19],high order SMC[20] , and so on. In this paper, we propose a dynamic modeling and design of a SMC scheme for trajectory tracking of a WMM, using Super Twisting algorithm (STW) to attenute chattering and improve performaces.

The remainder of this paper is organized as follows: Section II derives the dynamic model of the system, while Section III details the proposed controller design based on SMC. Section IV presents simulation results, and finally Section V concludes this work.

II. MODELLING

The considered system consists of a two-link manipulator mounted on 3 degrees of freedom wheeled omnidirectional robot, and located at its gravity center, as shown in Fig1.

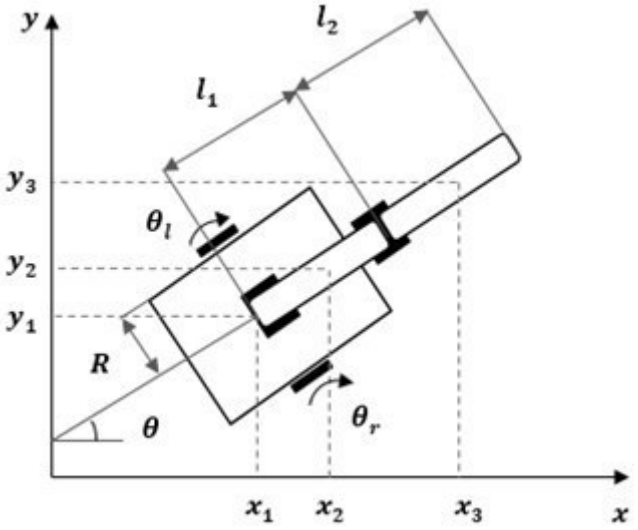


Fig. 1. Wheeled mobile manipulator presented in x-y plane

The vector of generalized coordinate variables for the system is given by

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

$$q = \begin{bmatrix} q_m{}^T & q_p{}^T \end{bmatrix}, \quad (1)$$

Where:

$q_p{}^T$, $q_m{}^T$ present the generalized coordinate variables for the wheeled robot and the manipulator respectively. $q_m = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}$, with θ_1 and θ_2 represent the joint angles of the 2-DOF manipulator.

$q_p = \begin{bmatrix} x_p & y_p & \theta \end{bmatrix}$, with x_p , y_p , θ are the Cartesian coordinates of the mobile platform.

In most research on dynamic models of omnidirectional mobile manipulators, the redundant actuation of the platform is often overlooked. Additionally, the integrated dynamic model of the omnidirectional mobile manipulator, derived from the driving wheels, is not explicitly addressed [21]. Using Lagrange's theory, the dynamic model of the combined system with external disturbances can be given as follows:

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) + \tau_d = J^T(q)\lambda + D(q)\tau \quad (2)$$

Where

- $M(q) \in \Re^{5\times5}$ is the inertia matrix;
- $C(q,\dot{q}) \in \Re^{5\times5}$ represents the vector of centripetal and Coriolis torques;
- $G(q) \in \Re^{5\times1}$ is the gravitational torque vector;
- $\tau \in \Re^{4\times1}$ is the vector of generalized input torques;
- $\tau_d \in \Re^{5\times1}$ denotes the external disturbances;
- $D(q) \in \Re^{5\times4}$ is a full rank input transformation matrix and is assumed to be known because it is the function of fixed geometry of the system;
- $J^T(q) \in \Re^{5\times4}$ is the kinematic constraints matrix;
- $\lambda \in \Re^{4\times1}$ is a constraint force vector.

The expressions of those matrixes are detailed in [22].

The non-holonomic constraints are considered independent of time and can be expressed as [21]:

$$J(q)\dot{q} = 0 \quad (3)$$

By using the following equation:

$$S(q)J(q) = S^T(q)J^T(q) = 0 \quad (4)$$

With :

$$S(q) = \begin{bmatrix} \frac{r}{2}\cos\theta & \frac{r}{2}\cos\theta & 0 & 0 \\ \frac{r}{2}\sin\theta & \frac{r}{2}\sin\theta & 0 & 0 \\ \frac{r}{2} & -\frac{r}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Where:

r is the radius of the wheel.

The dynamic model becomes:

$$\bar{M}(q)\ddot{q}_a + \bar{C}(q,\dot{q})\dot{q}_a + \bar{G}(q) = \bar{D}(q)\tau \quad (5)$$

with:

- $\bar{M}(q) = S(q)^T M(q) S(q)$
- $\bar{C}(q,\dot{q}) = S(q)^T (C(q,\dot{q})S(q) + M(q)\dot{S}(q))$
- $\bar{G}(q) = S(q)^T G(q)$
- $\bar{D}(q) = S(q)^T D(q)$
- $q_a = \begin{bmatrix} \theta_r & \theta_l & \theta_1 & \theta_2 \end{bmatrix}^T$ $\dot{q}_a = \begin{bmatrix} \dot{\theta}_r & \dot{\theta}_l & \dot{\theta}_1 & \dot{\theta}_2 \end{bmatrix}^T$

III. CONTROL DESIGN

The suggested control strategy is based on Sliding Mode Control, with the sliding variable is defined as:

$$S = \dot{e} + We \quad (6)$$

With:

$S = \begin{bmatrix} s_1,...,s_4 \end{bmatrix}^T$, $W = diag(w_i)$, $w_i > 0$ for $(i = 1\ to\ 4)$

The tracking error is given by $e = q_a - q_d$ and its derivative is $\dot{e} = \dot{q}_a - \dot{q}_d$, where the vectors q_d and $\dot{q}_d$ represents the desired trajectory and its time derivative respectively.

We define the control vector as: $U = \tau = [U_1\ U_2\ \tau_1\ \tau_2]$ with:

U_1 U_2 are the forces given by the mobile base motors,

τ_1 τ_2 are the input of manipulator actuation torques.

By assuming that $\dot{S} = 0$, we get the equivalent control as:

$$U_{eq} = \bar{C}\dot{q}_a + \bar{G} - \bar{M}(\lambda\dot{e} - \ddot{q}_d) \quad (7)$$

The global control law is:

$$U = U_{eq} + U_{dis} = \bar{C}\dot{q}_a + \bar{G} - \bar{M}(\lambda\dot{e} - \ddot{q}_d) - KSign(S) \quad (8)$$

In order to attenuate the chattering effects, We proceed to the second order SMC by using the 'Super Twisting Algorithm' [23] the control law becomes:

$$U = \bar{C}\dot{q}_a + \bar{G} - \bar{M}(\lambda\dot{e} - \ddot{q}_d) - \alpha|S|^{\frac{1}{2}}Sign(S) - \beta\int_{t_0}^{t_f} Sign(S) \quad (9)$$

IV. SIMULATION RESULTS

Simulation of the proposed controller has been carried out using an omnidirectional wheeled robot with a 2-DOF robotic manipulator given in [6]. Simulations have been performed in trajectory tracking mode using the MATLAB environment.

We consider the position error of the mobile platform with the following coordinates:

$$\begin{pmatrix} e_x \\ e_y \\ e_z \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_d - x \\ y_d - y \\ z_d - z \end{pmatrix} \quad (11)$$

For small variation of θ $(\theta \simeq 0)$ the position error become:

$$\begin{pmatrix} e_x \\ e_y \\ e_z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_d - x \\ y_d - y \\ z_d - z \end{pmatrix} \quad (12)$$

and for $\theta_r = \theta_l$ the kinematic matrix of the mobile platform is as follow:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} \frac{r}{2}\cos\theta & \frac{r}{2}\cos\theta \\ 0 & 0 \\ \frac{r}{2R} & \frac{r}{2R} \end{pmatrix} \begin{pmatrix} \dot{\theta}_r \\ \dot{\theta}_l \end{pmatrix} \quad (13)$$

R is the space between the mass center and the common drive wheel axis;

The initial values of all the system's generalized joints are $q_a = [\theta_r \ \theta_l \ \theta_1 \ \theta_2] = [0 \ 0 \ 0.1 \ 1]$.

The desired trajectory is defined by: $[\theta_{rd} \ \theta_{ld} \ \theta_{1d} \ \theta_{2d}]$; it consists of a displacement from 0 to 0.2 m along the x-axis without any change in y and θ. Hence, the trajectory must be tracked by the wheels is $\theta_r = \theta_l = t$,

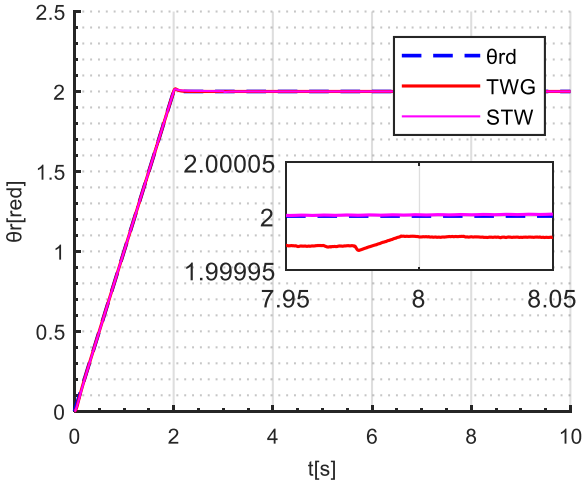
When the WMM reach the position $(x, y) = (0.2, 0)$, the 1st joint of the robotic arm θ_1 should track the path given by: $\theta_{1d}=t$. When θ_1 reaches 1rad, the 2nd joint θ_2 should change its angular position from 0 to 1rad through the path $\theta_{2d}=t$.

The parameters of the Super-Twisting sliding mode controller (STW) for all joints are:

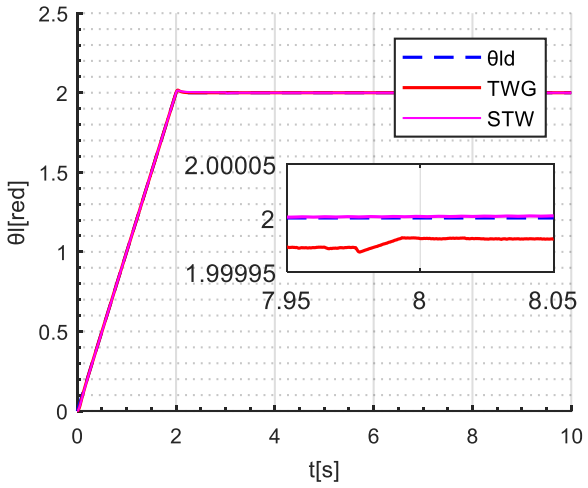
$w = diag(15,15,10,10)$, $\alpha = diag(1,1,2,1)$

$\beta = diag(0.5,0.5,0.5,0.5)$

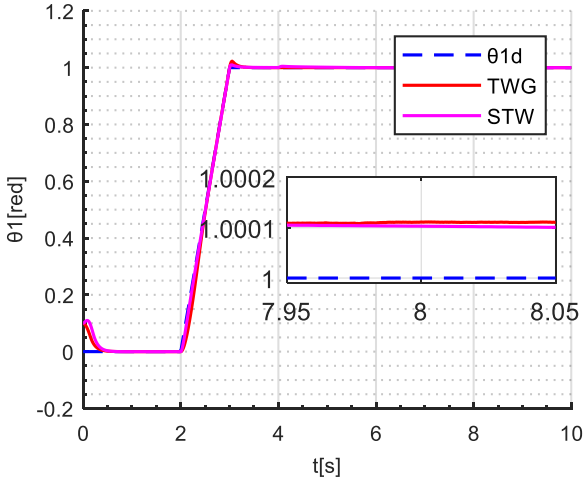
The proposed control law STW (equation 9), with the desired trajectory, the chosen parameters, and gains as described previously is implemented and compared with Twisting Algorithm (TWG). The comparison results are presented in Fig. 2-4, which illustrate the linear and angular positions, position errors, and control signals for all controllers.



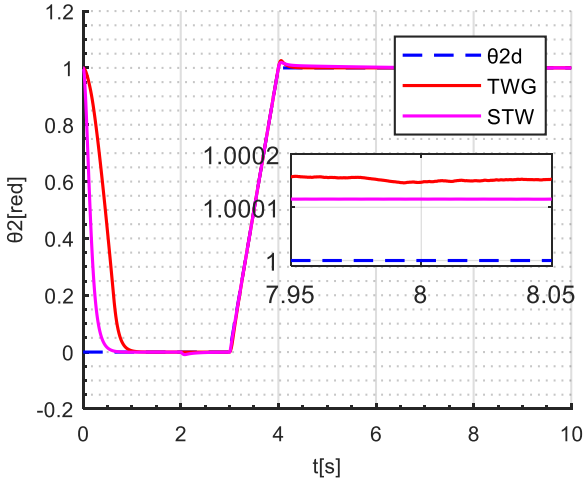
(a)



(b)

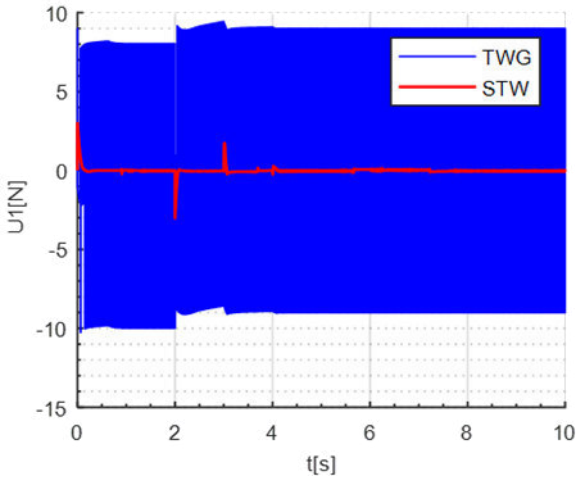


(c)

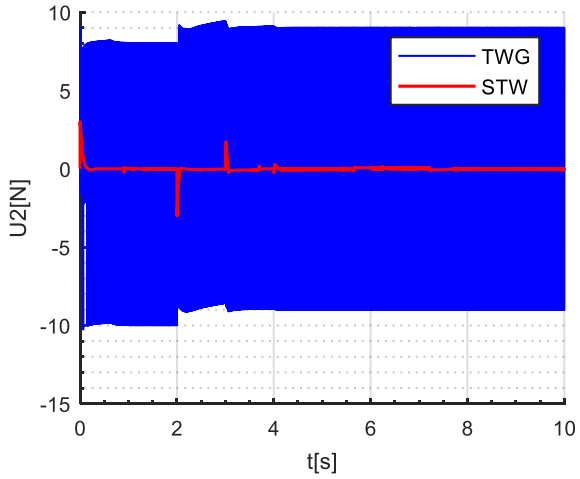


(d)

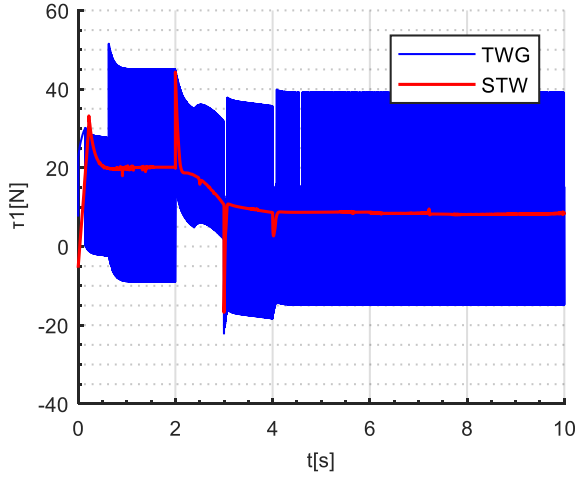
Fig. 2. Application of SMC in tracking trajectory mode.



(a)



(b)



(c)

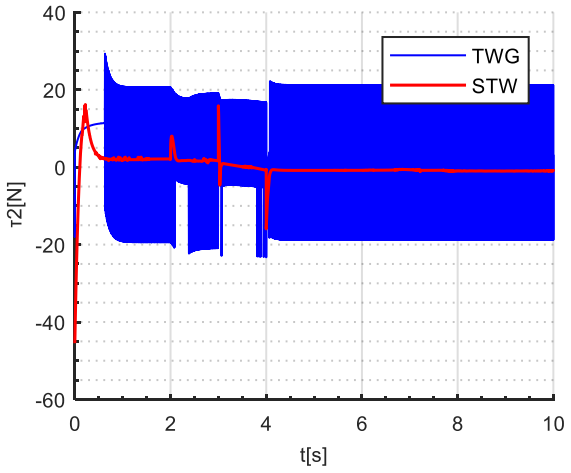


Fig. 3. Control signals

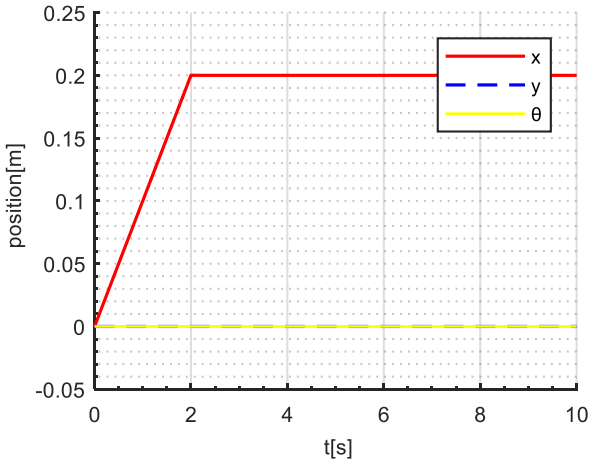


Figure. 4. Platform tracking trajectory

Figure 2 shows that the WMM is able to successfully track the desired path with best precision using STW control, particularly at the moment corresponding to the mobile base deflection at t=2s (as shown in Fig. 2 (a,b)). In term of rapidity, the STW control present fast response compared to TWG control (as shown in Fig. 2 (b,c))

Additionally, the use of STW algorithm leads to effectively reduce the chattering effects in the control signals as shown in Figure 3. The pics appeared in control signals are due to trajectories changements.

V. CONCLUSION

This paper aims to design robust control for omnidirectional wheeled mobile manipulator. Super Twisting Sliding Mode Control (STW) has been applied to this mobile manipulator for trajectory tracking. The proposed control strategy has significantly improved WMM performances and successfully attenuated chattering effects . The future work will focus on testing this controller under disturbances in complex trajectory tracking in order to further improve their performance and robustness.

REFERENCES

[1] Y. Liu and G. Liu, ‘‘Interaction analysis and online tip-over avoidance for a reconfigurable tracked mobile modular manipulator negotiating slopes,’’ IEEE/ASME Trans. Mechatronics, vol. 15, no. 4, pp. 623–635, Aug. 2010.

[2] J. Mattila, J. Koivumaki, D. G. Caldwell, and C. Semini, ‘‘A survey on control of hydraulic robotic manipulators with projection to future trends,’’ IEEE/ASME Trans. Mechatronics, vol. 22, no. 2, pp. 669–680, Apr. 2017,

[3] S. Aguilera-Marinovic, M. Torres-Torriti, and F. Auat-Cheein, ‘‘General dynamic model for skid-steer mobile manipulators with wheel–ground interactions,’’ IEEE/ASME Trans. Mechatronics, vol. 22, no. 1, pp. 433–444, Feb. 2017,

[4] Y. Yamamoto, X. Yun, “Effect of the dynamic interaction on coordinated control of mobile manipulators”, IEEE Transactions on Automatic Control 42, pp 816–824, 1996.

[5] A. Mazur, New approach to designing input–output decoupling controllers for mobile manipulators, Bulletin of the Polish Academy of Sciences, Technical Sciences 53 (1), 2005.

[6] G. White, R. Bhatt, V. Krovi, Dynamic redundancy resolution in a nonholonomic wheeled mobile manipulator, Robotica 25 (2), pp 147–156, 2007.

[7] E. Papadopoulos, J. Poulakakis, “Planning and model based control for mobile manipulators”, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems,, pp. 1810–1815., 2000.

[8] Z. Li, W. Chen, “Adaptive neural fuzzy of uncertain constrained multiple coordinated nonholonomic mobile manipulators”, Engineering Applications of Artificial Intelligence vol 21 (7), pp 985–1000, 2008.

[9] Yukun Zheng, Yixiang Liu, Rui Song, Xin Ma, Yibin Li, “Adaptive neural control for mobile manipulator systems based on adaptive state observer”, Neurocomputing, Vol 489, pp 504-520, 2022.

[10] M. Önkol and C. Kasnakoğlu, “Adaptive Model Predictive Control of a Two-wheeled Robot Manipulator with Varying Mass”, Measurement and Control, pp 1–19, 2018.

[11] Lijun Qiao, Xiao Luo, and Qingsheng Luo, “Control of Trajectory Tracking for Mobile Manipulator Robot with Kinematic Limitations and Self-Collision Avoidance”, Machines, 10, 1232, 2022.

[12] M. Boukattaya., M. Jallouli? & T.Damak, “On trajectory tracking control for non-holonomic mobile manipulators with dynamic uncertainties and external torque disturbances”. Robotics and autonomous systems, vol 60(12), pp 1640-1647, 2012.

[13] S. Riache, M. Kidouche and A.Rezoug, “Adaptive robust nonsingular terminal sliding mode design controller for quadrotor aerial manipulator”, Telkomnika, vol. 17(3),pp. 1501-1512, 2019.

[14] M.Z,Doghmane, M. Kidocuhe, S.Riache, and A. Aibeche, “Hybrid Adaptive Backstepping-Sliding Mode Control Design for Non-linear Under-Actuated Systems., Hatti, M., Ed.; Lecture Notes in Networks and Systems; Springer: Cham, Switzerland; Vol 361, 2022.

[15] M. Kidouche, S. Riache, A. Rezoug, “Adaptive Sliding Mode Control for Quadrotor Helicopter,” International Conference on Technological Advances in Electrical Engineering Skikda, Algeria, 2016.

[16] Z. Cao, Yugang. Niu, “Finite-time sliding mode control of Markovian jump systems subject to actuator nonlinearities and its application to wheeled mobile manipulator”, Journal of the Franklin Institute,Vol 355(16), pp 7865-7894, 2018.

[17] In Seok Seo, Seong Ik Han, “Dual closed-loop sliding mode control for a decoupled three-link wheeled mobile manipulator”, ISA Transactions, Vol 80, pp 322-335, 2018.

[18] X. Wang, D. Wang, M. Du, K. Song, Y. Ni and Y. Li, "A Two-Layer Trajectory Tracking Control Scheme of Manipulator Based on ELM-SMC for Autonomous Robotic Vehicle," IEEE Transactions on Automation Science and Engineering, pp1-12, 2023.

[19] Riache, S., Kidouche, M., Doghmane, M.Z. and Tee, K.F. ‘Optimal control design for uncertain aerial manipulator system based on an adaptive approach’, Int. J. Automation and Control, Vol. 18, No. 4, pp.425–443, 2024.

[20] Alizadeh M, Samaei MH, Vahid Estakhri M, Momeni H, Beheshti MT. “Robust trajectory tracking of a 3-DOF robotic arm using a Super-Twisting Fast finite time Non-singular Terminal Sliding Mode Control in the presence of perturbations”. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering.; 238(7), pp.1285-1295, 2024.

[21] M. Boukattaya, M. Jallouli, T. Damak, On trajectory tracking control for nonholonomic mobile manipulators with dynamic uncertainties and external torque disturbances, Robotics and Autonomous Systems, Vol 60 (12) pp 1640-1647, 2012.

[22] S. Lin and A. A. Goldenberg, Fellow, IEEE, Neural-Network Control of Mobile Manipulators, IEEE TRANSACTIONS ON NEURAL NETWORKS, Vol 12 (5), pp 1121-1133, 2001.

[23] Arie LEVANT, « Sliding ordre and sliding acceracy in sliding mode control », INT.J.COTROL, VOL 58 NO.6, pp 1247-1263, 1993.

Combining MLP and Gradient Boosting for Efficient Intrusion Detection

Nabila BOUDIA
Université Abderrahmane Mira, Laboratoire d'Informatiques MEDicale (LIMED), Bejaia, Algeria,
nabila.boudia@univ-bejaia.dz.

Malika BELKADI
Université Mouloud Mammeri, Laboratoire de Recherche en Informatique (LARI), Tizi ouzou, Algeria,
malika.belkadi@ummto.dz.

Hocine CHEBI
Faculty of Electrical Engineering, Djillali Liabes University of Sidi Bel Abbes, Laboratory Intelligent Control and Electrical Power System (ICEPS), Algeria.
chebi.hocine@yahoo.fr, hocine.chebi@univ-sba.dz,
chebi.hocine@univ-boumerdes.dz.

***Abstract*—The complexity of data currently exchanged on networks has significantly increased the number of attacks in different domains. Intrusion Detection Systems (IDS) face great difficulties in identifying new attacks using traditional detection methods. In this paper, we propose an approach combining two techniques: Boosting and MLP neural networks. We have integrated a neural network as a basis for the GBM model, in order to improve the performance of the Boosting technique. The method has been evaluated on the NSL-KDD dataset using two classification modes: binary and multi-class. The results obtained are promising, surpassing those of the individual GBM and MLP approaches. Our method has achieved an accuracy of 99.26% for binary classification and 93.66% for multi-class classification.**

Keywords—Intrusion detection systems (IDS), deep learning (DL), machine learning (ML), boosting, GBM, NSL-KDD.

I. INTRODUCTION

In recent years, the Internet has become an essential part of everyday life, connecting billions of devices around the world. It plays a crucial role in various sectors such as healthcare, industry, and business, facilitating information exchange and connectivity [1]. However, the amount of data exchanged over networks is becoming increasingly large, which significantly increases the risk of attack. In order to improve security and protect the network from vulnerabilities, it is essential to implement intrusion detection and prevention (IDS) mechanisms [2].

Intrusion detection systems (IDS) monitor network traffic to detect malicious behavior. They are divided into signature-based IDS (S-IDS), which detect known attacks, and anomaly-based IDS (A-IDS), which identify deviations from normal network behavior, including new attacks [3]. On the other hand, faced with the explosion in the volume of network traffic, traditional intrusion detection systems (IDS) are unable to process the data efficiently. Their ability to differentiate legitimate traffic from malicious activity is reduced, compromising their ability to reliably detect intrusions.

In this work we propose an approach combining a Deep Learning method which is Multi-Layer Perceptron (MLP), with a Boosting technique which is the Gradient Boosting Classifier. The objective of this integration allows to exploit the strengths of the two methods in a complementary way: the ability of MLP to model complex and non-linear relationships and the effectiveness of Boosting to progressively improve the accuracy of the models. This method allows to create a more efficient classification model. We also explore in this work the application of the combined method for both binary and multi-class classification. By treating these two types of classification problems, the study demonstrates the flexibility and efficiency of the MLP-GBM approach in various scenarios. The performance analysis on the dataset with multiple classes and binary cases allows to better understand the advantages and limitations of this technique. The contributions of this work are as follows:

- Integration of a Deep Learning MLP method with a Boosting technique
- Improvement of classification performance using an MLP as the basis of the Boosting technique
- Exploration of the method on binary and multi-class classification.

The rest of the article is organized as follows: in section 2 presents an overview of the state of the art on Boosting methods, section 3 describes the methodology used, section 4 experiments and discussion of the results and finally we end with a conclusion.

II. STATE OF THE ART

Ensemble and Boosting methods are widely used in the field of intrusion detection. Bukhariet al. [15] developed a model exploiting ensemble techniques to analyze traffic in fog nodes surrounding IoT infrastructures. Their experiments, conducted on the UNSW-NB15 and CICIDS201 datasets, mobilized several machine learning algorithms, such as RF,

SVM, KNN and ANN. A cross-validation dividing the ensembles into 10 subgroups allowed to evaluate their approach, which proved to be more efficient than individual models in identifying attacks. In another study [16], a detection model based on natural language processing (NLP) and ensemble techniques was proposed to spot abnormal network traffic flows. NLP extracts relevant information from HTTP requests, while ensemble methods provide classification. This model, tested on the HTTP CSIC 2010 dataset, achieved remarkable performances: an accuracy of 99.95%, an F1-score of 99.96%, and a low false alarm rate of 0.07%. However, its evaluation has not been extended to other datasets or to large data, which limits its analysis of generalizability and overfitting risk. Look and Tama [17] introduced a dual ensemble model combining bagging and gradient boosting techniques (GBDT).

III. Gradient Boosting Model (GBM) and Proposed Method

Boosting is a widely used ensemble method in the field of machine learning. This technique was introduced by Schapire in 1990 to improve the performance of traditional Machine Learning algorithms [15] such as decision trees, Support Vector Machines, NaiveBays, KNN, etc. The principle of Boosting is based on the idea of combining several weak learners (usually shallow decision trees) to create a stronger and more efficient model [16]. The principle of Boosting is to progressively correct the errors of previous models by giving a higher weight to misclassified examples, so that the following models focus on these observations. Unlike other ensemble techniques such as Bagging [17], where models are built independently and in parallel, Boosting is a sequential process. Each model is trained by taking into account the errors of previous models. There are several Boosting methods namely AdaBoost (Adaptive Boosting) [7], Gradient Boosting [8], XGBoost [9] and LightGBM [18]. Boosting methods have been widely used in face recognition tasks [19], computer vision applications and intrusion detection.

In this work we used the Gradient Boosting algorithm. The main objective of the Gradient Boosting method is to find an approximation of the function F(x), which associates the input instances x with their corresponding output values y in an iterative manner. In other words, the function F(x) gives the value predicted by the model of each instance x, by minimizing the value of a given loss function, L(yi,,F(x)) so that the predictions are as close as possible to the real values [20]. Each iteration refines the model by reducing the residual errors, which allows to progressively improve the accuracy of the predictions. The operating steps of Gradient Boosting are as follows:

Step 1: Initialization of the model $F_0(x)$:

$$F_0(x) = argmin_c \sum_{i=1}^{n} L(y_i, c) \quad (1)$$

Step 2: Iteration: For m=1→M:

Calculation of the residuals:

$$r_i^{(m)} = -\left[\frac{\partial L(Y_I, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}\right] \quad (2)$$

Training of the new weak estimator $h_m(x)$:

$$h_m(x) = argmin_h \sum_{i=1}^{m} L(Y_i, F_{m-1}(x_I) + h(x_i) \quad (3)$$

Update of the model $F_m(x)$:

$$F_m(x) = F_{m-1}(x) + \alpha\, h_m(x) \quad (4)$$

Step 3: Final Prediction:

$$\hat{y} = F_M(x) \quad (5)$$

or y_i: actual output value, $\hat{y}_i^{(m)}$: value predicted by the model for sample iii at iteration m, $F_m(x)$: Function predicted by the model at iteration m, $h_m(x)$: new weak estimator at iteration m, M: number of weak learners, L: loss function, α: learning rate.

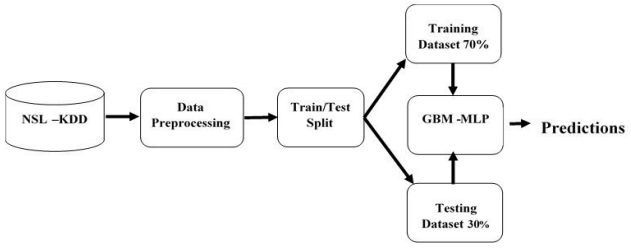


Fig.1. Proposed Method an intrusion detection method that combines Gradient Bossing and Multi-Layer Perception as a machine learning technique.

This paper proposes an intrusion detection method that combines Gradient Bossing and Multi-Layer Perceptron as a machine learning technique. GBM is used specifically for its performance in classifying samples in an ensemble, and MLP for its ability to extract complex features and capture nonlinear relationships in the data. Unlike this approach, Gradient Boosting uses a decision tree as the base learner; however, we employed an MLP as the base learner to further improve the initial prediction. The reason for combining a Boosting method with a Deep Learning approach derived from the classical neural network that is MLP is to help the model classify the different attack classes in the ensemble even more accurately. The dataset used to train and evaluate our model is NSL-KDD. Although NSL-KDD is often considered obsolete due to the lack of the latest cybersecurity threats, it remains a standard reference in intrusion detection. It presents a distribution of four types of attacks, which allows a choice evaluation namely binary or multi-class of our method. In addition, it is used in a large number of comparative works, which allow a fairly simple evaluation of our performances compared to other methods.

IV. Discussion of Results

In the experimentation phase, two types of classification were used namely binary classification and multi-class classification. We used a multi-layer neural network (MLP) with a simple architecture, designed to capture the most relevant features without increasing the training time. This MLP was then integrated into a Gradient Boosting classifier, which used it as a basis to build a classification model. The experiments were carried out on an HP Elite Book laptop equipped with an 8th generation Intel Core i5 processor and 8 GB of RAM. The implementation of the models was carried

out using the scikit-learn libraries for Gradient Boosting and Tensor Flow with Keras for MLP. The results obtained on the training and test sets in the case of binary classification are presented in Tables 1 and 2 respectively.

Metrics \ Models	GBM-MLP	GBM
Accuracy	99,48%	98,30%
Precision	99,48%	98,35%
Recall	99,48%	98,26%
F1-Score	99,48%	98,30%
DR	99,48%	98,30%
FPR	0,52%	1,69%
FNR	0,52%	1,69%

Table 1. Results of binary classification on the training set.

Metrics \ Models	GBM-MLP	GBM
Accuracy	99,26%	98,12%
Precision	99,26%	98,18%
Recall	99,26%	98,09%
F1-Score	99,26%	98,12%
DR	99,26%	98,12%
FPR	0,74%	1,87%
FNR	0,74%	1,87%

Table 2. Binary classification results on the test set.

Overall, the results in Tables 1 and 2, the confusion matrices of the different techniques and the classification reports show that the GBM-MLP model is superior to Gradient Boosting in all the underlying metrics for both the training and test sets. For the training set, GBM-MLP has 99.48% accuracy, while GBM has 98.30% accuracy alone. GBM-MLP has all other metrics such as precision, recall, F1-Score and Detection Rate at around 99.48% while GBM has all the classification metrics at around 98.30%.

V. CONCLUSION

In this work we proposed a model that combines the Multi-Layer Perceptron (MLP) and the Gradient Boosting (GBM). The MLP was used as a base model for the initial predictions of the GBM, in order to improve the prediction performance of the Boosting model. To evaluate this approach we used the NSL-KDD benchmark dataset. The results obtained on the test set outperformed the classical GBM in terms of accuracy, precision, F1-Score, Recall, FNR and FPR with the values of 99.26%, 99.26%, 99.26%, 99.26%, 0.74% and 0.74% respectively in the Binary classification. In the case of multi-class classification the model obtained an accuracy of 99.32%, a precision of 93.66%, a recall of 89.80%, an F1-Score of 91.47%, an FPR of 0.16% and an FNR of 0.67%.

Références

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, et M. Ayyash, « Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications », IEEE Commun. Surv. Tutor., vol. 17, no 4, p. 2347‑2376, 2015, doi: 10.1109/COMST.2015.2444095.

[2] A. Boukhalfa, A. Abdellaoui, N. Hmina, et H. Chaoui, « LSTM deep learning method for network intrusion detection system », Int. J. Electr. Comput. Eng. IJECE, vol. 10, no 3, p. 3315, juin 2020, doi: 10.11591/ijece.v10i3.pp3315-3322.

[3] P. Bedi, N. Gupta, et V. Jindal, « Siam-IDS: Handling class imbalance problem in Intrusion Detection Systems using Siamese Neural Network », Procedia Comput. Sci., vol. 171, p. 780‑789, 2020, doi: 10.1016/j.procs.2020.04.085.

[4] T. J. Lucas et al., « A Comprehensive Survey on Ensemble Learning-Based Intrusion Detection Approaches in Computer Networks », IEEE Access, vol. 11, p. 122638 ‑ 122676, 2023, doi: 10.1109/ACCESS.2023.3328535.

[5] C. Kolias, G. Kambourakis, et M. Maragoudakis, « Swarm intelligence in intrusion detection: A survey », Comput. Secur., vol. 30, no 8, p. 625‑642, nov.2011, doi: 10.1016/j.cose.2011.08.009.

[6] Y. Zhou, T. A. Mazzuchi, et S. Sarkani, « M-AdaBoost-A based ensemble system for network intrusion detection », Expert Syst. Appl., vol. 162, p. 113864, déc. 2020, doi: 10.1016/j.eswa.2020.113864.

[7] Y. Freund et R. E. Schapire, « A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting », J. Comput. Syst. Sci., vol. 55, no 1, p. 119‑139, août 1997, doi: 10.1006/jcss.1997.1504.

[8] J. H. Friedman, « Greedy Function Approximation: A Gradient Boosting Machine », Ann. Stat., vol. 29, no 5, p. 1189‑1232, 2001.

[9] T. Chen et C. Guestrin, « XGBoost: A Scalable Tree Boosting System », in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA: ACM, août 2016, p. 785‑794. doi: 10.1145/2939672.2939785.

[10] I. H. Sarker, « Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions », SN Comput. Sci., vol. 2, no 6, p. 420, nov.2021, doi: 10.1007/s42979-021-00815-1.

[11] M. Sajidet al., « Enhancing intrusion detection: a hybrid machine and deep learning approach », J. Cloud Comput., vol. 13, no 1, p. 123, juill. 2024, doi: 10.1186/s13677-024-00685-x.

[12] A. Sarıkaya, B. GünelKılıç, et M. Demirci, « GRU- GBM : A combined intrusion detection model using LightGBM and gated recurrent unit », Expert Syst., vol. 39, no 9, p. e13067, nov.2022, doi: 10.1111/exsy.13067.

[13] Q. Wang, W. Zhao, X. Wei, J. Ren, Y. Gao, et B. Zhang, « Intrusion Detection Algorithm Based on Convolutional Neural Network and Light Gradient Boosting Machine », Int. J. Softw. Eng. Knowl. Eng., vol. 32, no 08, p. 1229‑1245, août 2022, doi: 10.1142/S0218194022500462.

[14] C. Tang, N. Luktarhan, et Y. Zhao, « An Efficient Intrusion Detection Method Based on LightGBM and Autoencoder », Symmetry, vol. 12, no 9, p. 1458, sept. 2020, doi: 10.3390/sym12091458.

[15] G. Folinoet P. Sabatino, « Ensemble based collaborative and distributed intrusion detection systems: A survey », J. Netw. Comput. Appl., vol. 66, p. 1‑16, mai 2016, doi: 10.1016/j.jnca.2016.03.011.

[16] R. E. Schapire, « The strength of weak learnability ».

[17] L. Breiman, « Bagging predictors », Mach. Learn., vol. 24, no 2, p. 123‑140, août 1996, doi: 10.1007/BF00058655.

[18] G. Keet al., « LightGBM: A Highly Efficient Gradient Boosting Decision Tree ».

[19] P. Viola et M. Jones, « Rapid object detection using a boosted cascade of simple features », in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA: IEEE Comput. Soc, 2001, p. I-511-I‑518. doi: 10.1109/CVPR.2001.990517.

[20] C. Bentéjac, A. Csörgő, et G. Martínez-Muñoz, « A comparative analysis of gradient boosting algorithms », Artif. Intell. Rev., vol. 54, no 3, p. 1937‑1967, mars 2021, doi: 10.1007/s10462-020-09896-5.

[21] N. Anjumet M. R. Chowdhury, « International Journal of Advanced Research in Computer and Communication Engineering », SSRN Electron. J., 2024, doi: 10.2139/ssrn.4847308.

[22] [I. Sharafaldinet al., « Towards a Reliable Intrusion Detection Benchmark Dataset », Softw. Netw., vol. 2017, no 1, p. 177‑200, 2017, doi: 10.13052/jsn2445-9739.2017.009.

[23] A. Paya, S. Arroni, V. García-Díaz, et A. Gómez, « Apollon: A robust defense system against Adversarial Machine Learning attacks in Intrusion Detection Systems », Comput. Secur., vol. 136, p. 103546, janv. 2024, doi: 10.1016/j.cose.2023.103546.

[24] C. Xu, J. Shen, X. Du, et F. Zhang, « An Intrusion Detection System Using a Deep Neural Network With Gated Recurrent Units », IEEE Access, vol. 6, p. 48697 ‑ 48707, 2018, doi: 10.1109/ACCESS.2018.2867564.

[25] M. E. Elhamahmy, H. N. Elmahdy, et I. A. Saroit, « A New Approach for Evaluating Intrusion Detection System », 2010.

Complete process of developing a plant disease detection system using deep learning

Djawed Megaiz MED
Faculty of Electrical Engineering, Djillali Liabes University of Sidi Bel Abbes, Laboratory Intelligent Control and Electrical Power System (ICEPS), Algeria,
djawedmeg08@gmail.com.

Hocine CHEBI
Faculty of Electrical Engineering, Djillali Liabes University of Sidi Bel Abbes, Laboratory Intelligent Control and Electrical Power System (ICEPS), Algeria.
chebi.hocine@yahoo.fr, hocine.chebi@univ-sba.dz,
chebi.hocine@univ-boumerdes.dz.

***Abstract*—This paper presents the complete process of developing a plant disease detection system using deep learning. We first preprocessed the PlantVillage database to remove image backgrounds and isolate plants, thereby improving data quality. Then, we built a CNN model including multiple hidden layers, such as convolution operations, MaxPooling, and relu activation functions, to capture the distinctive features of the processed images. After dividing the data into training and testing sets, we trained the model on the training set and evaluated its performance on the testing set to measure its accuracy in detecting plant diseases. This step allowed us to verify the effectiveness of our approach. We then compared the results obtained with those of other similar works using the PlantVillage database, analyzing the accuracy rates, error rates, and learning times. This comparison provided us with a context to evaluate the competitiveness of our model compared to existing methods.**

Keywords—Plant disease detection, deep learning, Max Pooling, Deep learning.

I. INTRODUCTION

Plant disease detection in modern agriculture benefits from the integration of innovative technologies such as Artificial Intelligence (AI) and the Internet of Things (IoT). The use of AI for crop image analysis enables accurate identification of disease symptoms, facilitating early detection of early signs of disease and enabling proactive intervention to limit crop damage. By combining these technological advances with extensive interdisciplinary research, it is possible to transform the way plant diseases are detected and managed. In plant disease detection, the use of vision plays a crucial role. The work of Han, S et al [1] addresses in depth the importance of imaging for early identification of plant diseases.

By analyzing the different classes of points, represented in a three-dimensional space based on texture and color parameters. In the work of Cruz et al. [2] They developed a vision-based system to detect leaf blight symptoms on Olea europaea L. leaves infected with Xylella fastidiosa. This system uses deep learning and transfer learning techniques to automatically and quantitatively detect plant diseases, which could reduce diagnostic costs and time. The work of SS Harakannanavar et al [3] discusses the important role of current technology, including computer vision and machine learning, in plant disease detection. It emphasizes the use of analysis and detection processes to help farmers solve plant disease problems. The work divides the process into three stages: identification, analysis, and verification using an available database. Using machine learning and image processing tools, the proposed algorithm aims to automatically detect plant diseases, with a focus on early symptoms to enable farmers to take timely action. Furthermore, they highlight the importance of using techniques such as discrete wavelet transform, principal component analysis, and convolutional neural networks to extract and classify informative features from diseased leaf samples, thus providing a solution for early detection of plant diseases.

The role of vision for plant disease detection in the work of SinghV et al [4] is valuable. Imaging techniques and vision systems are used to acquire images of plants to detect signs of diseases. These images are then processed using various methods such as segmentation, thermal, hyperspectral, fluorescence, multispectral, and 3D analysis. In addition, advanced methods such as deep learning, Support Vector Machine (SVM), K-means clustering, and K-Nearest Neighbor (K-NN) are also employed for disease classification. These computer vision approaches help in early identification of diseases and provide effective means for detection and classification of plant diseases, which is essential for crop protection.

Deep learning (DL) is a form of machine learning based on artificial neural networks that has had a significant impact on several disciplines due to its exceptional data analysis capability. In the context of agricultural robotic vision, deep learning plays an important role in enabling autonomous systems to visualize, interpret, and act on visual information acquired in the agricultural environment. This modification of deep learning aims to transform field images and videos into actionable information that can be used to manage crops, inspect plants, and increase productivity while minimizing negative impacts on the environment. Let's discuss in detail the fundamental role of deep learning in agricultural robotic vision and the specific ways in which it has been adapted to address the unique challenges of this evolving sector.

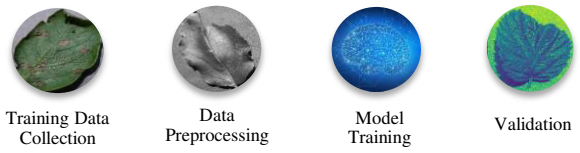


Fig. 1. Block diagram showing the different steps to be used for detecting plant diseases.

II. STATE OF THE ART ON DISEASE DETECTION FOR PLANTS

Computer vision, and object recognition in particular, has made tremendous progress in recent years, most recently computer vision for large-scale environments (VOGE) have been widely used as benchmarks for many problems related to computer vision visualization, including object classification.

In the agricultural domain, computer vision is needed at all costs of the application. In the work Mohanty et al [1], seems to be focused on the use of mobile technology and image data for the identification and diagnosis of agricultural problems, with a particular focus on plant diseases. The authors discuss the potential use of smartphones to capture image data, as well as location and time information, to help prevent yield losses. They also highlight the rapid advances in mobile technology and the potential for highly accurate diagnostics using smart phones, as well as the implementation of algorithms and the use of additional resources for deep learning and image analysis related to plant diseases.

III. DISEASE DETECTION METHODOLOGY BY DL

In this paper, we will use Deep Learning on plant images to detect diseases. DL is a subcategory of machine learning (ML), which in turn is a field of machine learning. It is distinguished by the use of artificial neural networks to learn from unstructured or unlabeled data. This technique is based on the use of artificial neural networks, which are mathematical models inspired by the functioning of the human brain. These networks are composed of several layers of interconnected neurons, allowing a hierarchical representation of the data, these layers allow a complex abstraction of the characteristics of the input data. It is effective for the automatic processing of unstructured data, such as images, sound and text, thanks to its ability to learn highly abstract representations.

DL is widely used in the field of computer vision, we will use it in our work to detect and classify diseased plants. We will use deep learning by participating in the release of PyTorch, a popular open-source machine learning library for developing deep learning-based plants. PyTorch offers us the ability to create custom neural network models, especially convolutional neural networks (CNNs) suitable for analyzing images of leaves or specific parts of plants affected by a particular disease, this will allow us to accurately classify different diseases. With its ability to dynamically compute complex problems, PyTorch allows for rapid experimentation and prototype development, which is essential for creating high-performance models. In addition, it benefits from documentation and a supportive community, both of which are essential for solving the specific problems associated with this application.

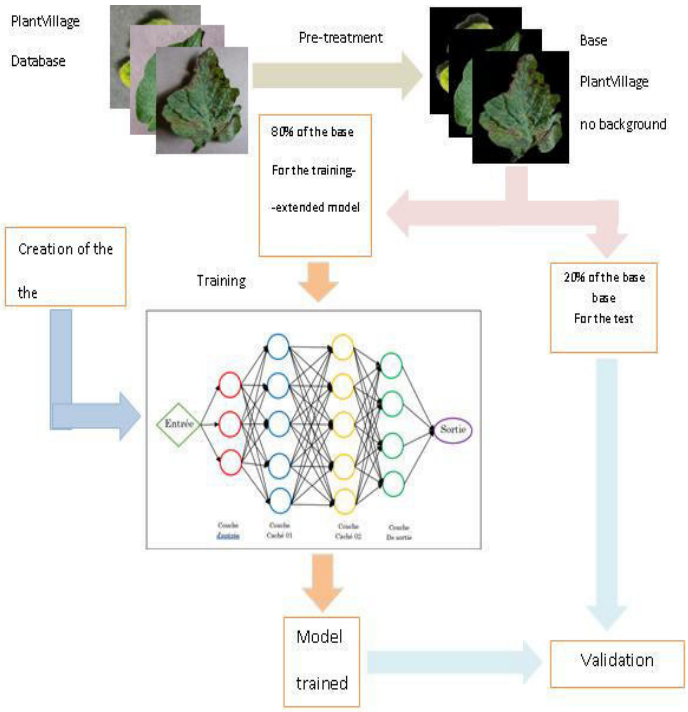


Fig. 2: Detailed flowchart of the method used.

IV. RESULTS AND DISCUSSIONS

We will start by exploring the database, we will see how many classes are present, the number of images in each class, the total number of images, the size and shape of the images.

We will perform the preprocessing of the PlantVillage database, This dataset consists of approximately 87,000 rgb images of healthy and diseased crop leaves, divided into 38 different classes. The dataset is divided into an 80/20 ratio between the training set and the validation set, while maintaining the structure of the there are 38 folders. A new folder containing 33 test images is created later for prediction purposes.

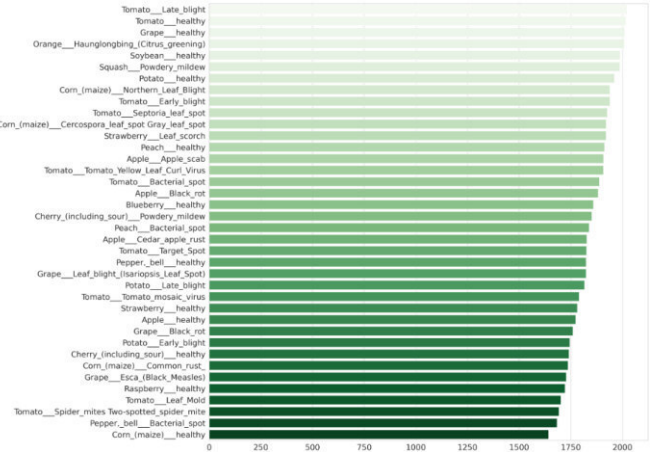


Fig. 1 : Diagram shows the distribution of images in classes.

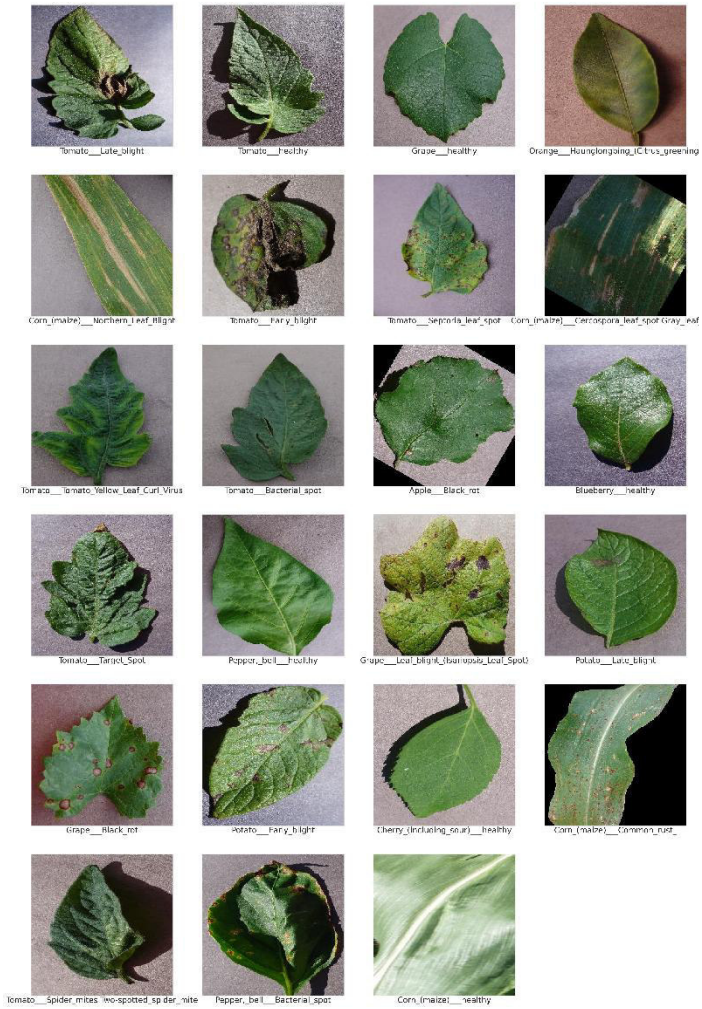


Fig. 2 : Des échantillons de chaque classe de notre base de données.

In the following table we will classify our accuracy results with the other two works, we notice that the three models give high accuracies, we can say that the three techniques work correctly and give the exact diseases of the plants.

Papers	Approach	Year	Accuracy (%)
Mohanty et al [3]	CNN Profonds	2016	99.43
Yao et al [10]	CNN AlexNet,	2024	99.57
	CNN MobileNetV2	2024	99.27
This Paper	**CNN ResNet**	**2024**	**99.73**

Tab. 3: Comparison of Accuracy of CNN Models for Plant Disease Classification

V. CONCLUSION

In conclusion, this paper presents the complete process of developing a plant disease detection system using deep learning. We first preprocessed the PlantVillage database to remove image backgrounds and isolate plants, thereby improving data quality. Then, we built a CNN model including multiple hidden layers, such as convolution operations, MaxPooling, and relu activation functions, to capture the distinctive features of the processed images.

Références

[1] 1] Han, S., & Cointault, F. (2013, June). Détection précoce de maladies sur feuilles par traitement d'images. In Orasis, Congrès des jeunes chercheurs en vision par ordinateur.

[2] [2] Cruz, A. C., Luvisi, A., De Bellis, L., & Ampatzidis, Y. (2017). Vision-based plant disease detection system using transfer and deep learning. In 2017 asabe annual international meeting (p. 1). American Society of Agricultural and Biological Engineers.

[3] [3] Harakannanavar, S. S., Rudagi, J. M., Puranikmath, V. I., Siddiqua, A., & Pramodhini, R. (2022). Plant leaf disease detection using computer vision and machine learning algorithms. Global Transitions Proceedings, 3(1), 305-310.

[4] [4] Singh, V., Sharma, N., & Singh, S. (2020). A review of imaging techniques for plant disease detection. Artificial Intelligence in Agriculture, 4, 229-242.

[5] [5] Chowdhury, M. E., Rahman, T., Khandakar, A., Ayari, M. A., Khan, A. U., Khan, M. S., ... & Ali, S. H. M. (2021). Automatic and reliable leaf disease detection using deep learning techniques. AgriEngineering, 3(2), 294-312.

[6] [1] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, 1419.

[7] [2] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Image net classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, eds F .Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates ,Inc.),1097–1105.

[8] [3] Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. ArXiv:1409.1556

[9] [4] Zeiler ,M. D., and Fergus, R.(2014). "Visualizing and understanding convolutional networks, "in Computer Vision – ECCV2014, eds D. Fleet, T. Pajdla, B. Schiele, And T. Tuytelaars (Springer), 818–833.

[10] [5] He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. arXiv :1512.03385.

[11] [6] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., etal. (2015). "Going deeper with convolutions," in Proceedings of the IEEE Conferenceon Computer Vision and Pattern Recognition.

[12] [10] Yao, J., Tran, S. N., Garg, S., & Sawyer, S. (2024). Deep Learning for Plant Identification and Disease Classification from Leaf Images: Multi-prediction Approaches. ACM Computing Surveys, 56(6), 1-37.

Cuckooo Algorithm for Maximum Power Point Tracking in Wind Turbines

Billel Naceri
Laboratory of Electrical Engineering (LGEB)
University of Biskra
Biskra, Algeria
billel.naceri@univ-biskra.dz

Ali Mazari
Laboratory Of The Applied and Automation and Industrial Diagnostic
University of Djelfa
Djelfa, Algeria
ali.mazari@univ-djelfa.dz

Younes Belhadjer
Laboratory Of The Applied and Automation and Industrial Diagnostic
University of Djelfa
Djelfa, Algeria
younes.belhadjer@univ-djelfa.dz

Okba Fergani
Laboratory of Electrical Engineering (LGEB)
University of Biskra
Biskra, Algeria
okba.fergani@univ-biskra.dz

Hamou Ait Abbas
Laboratory Of Materials and Durable Development-LM2D
University of Bouira
Bouira, Ageria
h.aitabbas@univ-bouira.dz

Kamel Aoues
Department of Mechanical Engineering-Biskra
University of Biskra
Biskra, Algeria
k.aoues@univ-biskra.dz

Abstract´ This paper examines the Cuckoo Search Algorithm (CSA) for Maximum Power Point Tracking (MPPT) in wind turbines, utilizing real wind speed data from a region known for its semi-arid climate and varied topography. The CSA, inspired by the brood parasitism behavior of cuckoo birds, is evaluated for its capability to optimize energy extraction under fluctuating wind conditions. Through Matlab/Simulink simulations, CSA's performance is compared with other popular MPPT techniques to assess its effectiveness. Results demonstrate that CSA delivers a notable improvement in power output and system stability, particularly in dynamic wind environments. Although CSA presents significant advantages in terms of efficiency and simplicity, the study also identifies areas for potential improvement in tracking accuracy and response time under extreme environmental conditions. The findings highlight the importance of utilizing advanced optimization algorithms to enhance the reliability of wind energy systems.

Keywords´ cuckooo Optimization Algorithm (WOA); Maximum Power Point Tracking (MPPT); Wind Turbines; Optimization Algorithms; Renewable Energy; Energy Efficiency; Computational Intelligence.

I. INTRODUCTION

Wind energy has emerged as a pivotal element in the global shift towards renewable energy sources, significantly contributing to the reduction of greenhouse gas emissions and the reliance on fossil fuels. The deployment of wind turbines (WTs) facilitates clean electricity generation, which is essential for combating climate change and promoting energy independence [1], [2]. However, the optimization of wind turbine performance is crucial, as it involves balancing energy output with mechanical strain and operational costs. This optimization is particularly challenging due to the dynamic and unpredictable nature of wind conditions, necessitating advanced control mechanisms to ensure that turbines operate at their maximum power point (MPP) [3], [4].

To address these challenges, Maximum Power Point Tracking (MPPT) techniques have been developed, which are designed to ensure that wind turbines operate at peak efficiency by continuously adjusting to changing wind conditions. Among the various MPPT methods, optimization algorithms such as the Cuckoo Search Algorithm (CSA) and Artificial Neural Networks (ANN) have shown considerable promise in enhancing energy extraction from wind turbines. CSA, inspired by the brood parasitism behavior of cuckoo birds, effectively balances exploration and exploitation within the search space, allowing for rapid convergence on optimal operating points even under fluctuating wind conditions [5], [6]. This adaptability makes CSA a robust choice for MPPT in wind energy systems, particularly when traditional methods may falter [7].

Conversely, ANN leverages machine learning techniques to predict and optimize turbine performance based on historical data. By learning from past wind conditions and operational data, ANN-based MPPT systems can provide high precision and flexibility in real-time adjustments, outperforming conventional techniques in complex and nonlinear environments [8], [9]. The continuous learning capability of ANN allows for better adaptation to the variable nature of wind, thereby improving the overall efficiency of wind energy systems [10], [11]. While other optimization methods, such as the Whale Optimization Algorithm (WOA), also present valuable solutions, this study primarily focuses on evaluating the performance of CSA and ANN for MPPT in wind turbines, highlighting their respective strengths in maximizing energy efficiency [12], [13].

The integration of advanced control mechanisms such as CSA and ANN into wind turbine operations is essential for optimizing energy output and enhancing the overall performance of wind energy systems. These methodologies not only improve the efficiency of energy extraction but also contribute to the sustainability of renewable energy sources, thereby playing a crucial role in addressing the environmental challenges posed by climate change [14], [15]. Future research should continue to explore and refine these techniques to further enhance the reliability and efficiency of wind energy systems.

II. WIND TURBINE MODELING

Wind turbine modeling consists of four parts: wind model, the model of the wind turbine, electrical and mechanical equations.

A. Wind Model

The wind speed at a point in space 侏鼹鼹is the resultant of two components: 侏鼹鼹varies slowly and regularly over longer periods at a given site, while the other 侏鼹鼹varies turbulently with time. Its expression is given by formula (1) [16].

侏鼹鼹 侏鼹鼹 侏鼹鼹 (1)

XXX-X-XXXX-XXXX-X/XX/$XX.00 ø 20XX IEEE

係 is a turbulent component of the wind which is a stationary random process (does not vary with the mean wind speed).

係 is a slow component: it is the average wind speed varying regularly over longer periods at a given site.

The aero turbine effectively filters high frequency fluctuations by employing a reconstituted low-pass filter specifically designed for the turbulence component. This process allows for the accurate reproduction of a characteristic closer to reality, as defined by the transfer function expressed in (2). [16]

$$係 = \frac{係}{係} \quad (2)$$

The time constant value is influenced by the rotor diameter, as well as the intensity of wind turbulence and the average wind speed.(係= 0.11375s) [6].

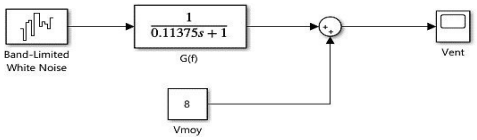


Fig. 1. Synoptic of wind reconstruction.

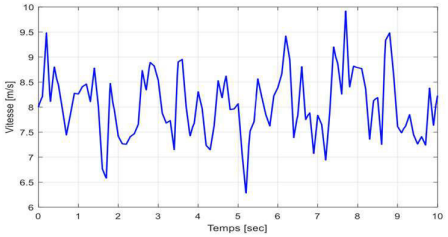


Fig. 2. Synthetic wind sequence.

The previous figure (2) represents a wind sequence generated by the Simulink model of figure (1) with an average wind speed 係 = 係 × 係.

The power coefficient is defined as the ratio between the power extracted from the wind (recoverable) and the total power theoretically available [16].

$$係 = \frac{係}{係} = \frac{係}{係} \quad (3)$$

The power coefficient reaches a theoretical maximum of 16/27, or approximately 0.59, known as the Betz limit (Figure 3). This limit defines the maximum amount of power that can be extracted from a given wind speed. However, in practice, this limit is never achieved, and each wind turbine operates with its own specific power coefficient. This coefficient is expressed as a function of the tip-speed ratio (≅), which is the ratio between the speed at the tip of the turbine blades and the wind speed [17]. From this, the aerodynamic efficiency of the turbine is derived:

$$係 = \frac{係}{係} 係 \quad (4)$$

In our case, the evolution of the power coefficient, which depends on both the tip-speed ratio (≅) and the blade pitch angle (∮), is represented by the following polynomial approximation:

係 = 係 + 係 + 係 + 係 + 係 + 係 + 係 + 係 + 係 + 係 + 係 + 係 + 係 (5)

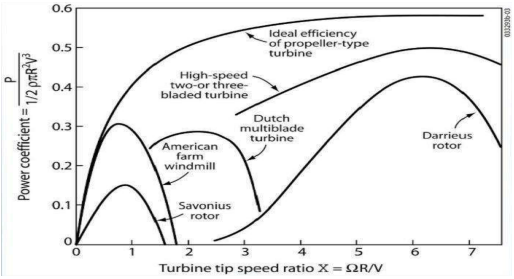


Fig. 3. Power coefficient for different types of wind turbines.

B. Model of the Wind Turbine

The turbine converts the kinetic energy of the wind into mechanical energy. It consists of three blades that rotate around an axis at an angle ∮. The wind speed V, acting on the blades, induces their rotation, generating mechanical power on the turbine's shaft [16].

Mechanical power 係 available on the wind turbine shaft, as well as that of its torque 係 are given as follows:

$$係 = 係 = \frac{係}{係} 係 \quad (6)$$

$$係 = \frac{係}{係} = \frac{係}{係} 係 \quad (7)$$

係 = Torque of the wind turbine (N.m).

The power coefficient 係 represents the aerodynamic efficiency of the wind turbine and is also influenced by the specific characteristics of the turbine. The output quantities of the turbine, such as power or torque, can be controlled by adjusting the previously mentioned input variables, including wind speed and blade pitch angle.

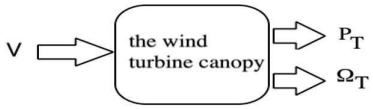


Fig. 4. Input-output of the wind turbine model.

The action of the moving air will result in forces applied at each point of the blade surface.

The block diagram of the studied blades is represented in (Figure 5).

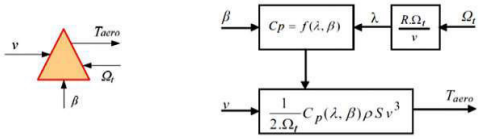


Fig. 5. Block diagram of the blade model.

The role of the gearbox is to convert the mechanical speed of the turbine into the appropriate speed for the generator and to transform the aerodynamic torque into the gearbox torque. This relationship is described by the following mathematical formula:

$$係 = \frac{係}{係}$$

Where: 係: Rotational speed after multiplier in (rad/s).

係: Torque after multiplier (N.m)

The fundamental equation of dynamics enables the determination of the mechanical speed's evolution from the total mechanical torque C_{em} applied to the rotor. This relationship is expressed by the following equation:

$$J\frac{d\Omega_t}{dt} = C_{em} - f_v\Omega_t - C_g$$

C_{em}: The electromagnetic torque developed by the generator.

C_g: The torque developed by the gearbox.

f_v The friction coefficient.

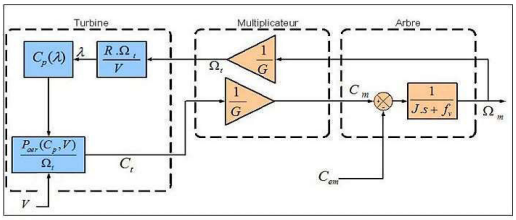


Fig. 6. Block diagram of the wind turbine model.

Classical MPPT facilitates the search for the maximum point of power, knowing the characteristic of the wind turbine, this method allows us to quickly approach the optimum with the help of simple measurements, internal to the mechanical-electrical converter, that is to say without the use of wind speed sensor [18].

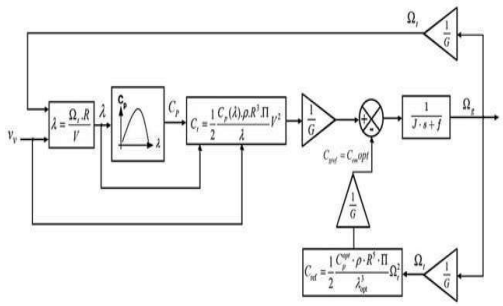


Fig. 7. scematic of the MPPT control strategy of the wind turbine.

C. Electrical equations

The electrical equations of the stator in a Permanent Magnet Synchronous Machine (PMSM) are derived from the generalized theorem of the generating convention. According to this convention, reversing the direction of the currents results in the following equations for the PMSM:

$$\begin{cases} V_d = R_s I_d + \frac{d\Phi_d}{dt} - \omega_e L_q I_q \\ V_q = R_s I_q + \frac{d\Phi_q}{dt} + \omega_e L_d I_d + \omega_e \Phi_f \end{cases} \quad (8)$$

The expression of the electromagnetic torque is written as follows:

$$C_{em} = \frac{3}{2}p\left[\Phi_f I_q + (L_d - L_q) I_d I_q\right] \quad (9)$$

We replace the flows by their expressions given in (9), we thus have:

$$C_{em} = \frac{3}{2}p\left[\Phi_f I_q + (L_d - L_q) I_d I_q\right] \quad (10)$$

Where:

V_d, V_q Represent respectively the voltage vectors related to the reference frame (d, q).

I_d, I_q Represent respectively the current vectors related to the reference frame (d, q).

ω_e: Represent the electrical speed of rotation of the rotor.

L_d, L_q: Represent respectively the stator inductances of the frame (d, q).

D. Mechanical Equations

The mechanical equation below explains the dynamics:

$$J\frac{d\Omega_t}{dt} = C_{em} - C_{mot} - f_v\Omega_t$$

Being aware of:

C_{mot}: applied motor torque to the generator.

J Inertia moment.

f_v Viscous friction coefficient.

III. Control Methodes Applied for MPPT

A. Sliding mode control

Sliding Mode Control (SMC) is a robust control strategy particularly suited for nonlinear systems, such as wind energy conversion systems, where it effectively manages uncertainties and fluctuations in wind speed and turbine characteristics. In the context of Maximum Power Point Tracking (MPPT) for wind turbines, SMC is employed to guide the system toward optimal power extraction despite varying conditions. The fundamental principle of SMC involves defining a "sliding surface" in the state space, which is based on the desired performance criteria related to the maximum power point (MPP) [19].

The sliding surface is typically formulated using the system's state variables, such as current, voltage, or power output, to ensure that these parameters align with the optimal operating conditions [20]. By driving the system dynamics to converge on this surface, SMC guarantees that the turbine operates near its maximum efficiency, effectively compensating for disturbances and uncertainties [19]. This approach has been shown to enhance the robustness of wind energy systems, allowing them to maintain optimal performance even under challenging operational conditions [21], [22].

The sliding surface s(t) can be expressed as:

$$s(x) = x_{ref} - x$$

Where:

x is the current state variable.

x_{ref} is the reference state variable corresponds to MPP.

The objective is to design the controller so that s(t)” 0, indicating that the system state has converged to the desired operating point [21], [22].

For the system to remain on the sliding surface and achieve stability, the time derivative of the sliding surface $\dot{s}(x)$ must be controlled. The control law is designed to make this derivative zero when the system state is on the sliding surface, ensuring convergence to the MPP. This approach is effective for rejecting disturbances and accommodating variations in

wind speed, as the control law continually adjusts the system to maintain optimal performance [21], [22].

The equivalent control law 係eq is derived based on the system dynamics and is formulated to keep the system state on the sliding surface:

係eq 唸 侷f(x) 戴 g(x) 喃 侵 戴 f(x)

Where: f(x) and g(x) represent the nonlinear dynamics of the system.

係eq is the component of the control that matches the system˜s dynamics.

However, due to the inherent uncertainties and external disturbances, the equivalent control alone is insufficient. An additional switching term is introduced to correct deviations and ensure robustness [21], [22].

To enhance the robustness of the control strategy, a switching function is added to the control law. This switching component helps the system stay on the sliding surface, even when there are disturbances or variations in system parameters:

俣唸 俣eq 営依佩øá姻 戴 戴

Where: K is a positive constant that determines the aggressiveness of the control response.

sign(s(t)) is the sign function, which provides a discrete adjustment to bring the system state back to the sliding surface [21], [22].

This combination of equivalent control and switching term enables the SMC to rapidly correct deviations, providing strong resilience to parameter variations and external disturbances commonly encountered in wind turbine systems.

B. ANN control

Neural networks (NN) are advanced mathematical systems designed to emulate the human brain's capability to learn and process information. Utilizing artificial neurons implemented in software, NNs facilitate the simulation of human-like intelligence, thereby promoting machine autonomy through artificial intelligence (AI). The architecture of NNs comprises multiple layers of parallel processors, allowing for complex data manipulation and analysis, as demonstrated in Figure 8 [23].

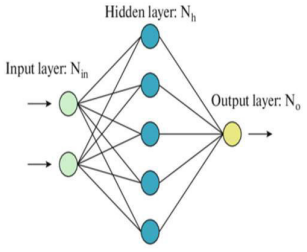


Fig. 8. NN structure.

In the realm of nonlinear systems control, NNs present an effective methodology for learning and approximating intricate dynamics. This capability enables the development of adaptive and precise control strategies, which are crucial for applications in various fields, including robotics and renewable energy. Traditional control methods often encounter significant challenges when addressing the complexities inherent in these systems. In contrast, NNs are favored for their exceptional ability to approximate continuous functions of real values, making them particularly well-suited for handling nonlinearities. The proposed NN controller for maximum power point tracking (MPPT) in wind turbines features a structure with two hidden layers. The inputs to the NN controller consist of the input current generated by the wind turbine and the output voltage from the boost converter. The primary output of the NN is the duty cycle (D), which modulates the operation of the boost converter to optimize energy extraction from the wind. The training of the NN was conducted using the (trainlm) function in Matlab, which employs the Levenberg-Marquardt (LM) algorithm [24]. This algorithm is renowned for its effectiveness in solving least-squares problems, thereby enhancing the training process of the neural network. Through this approach, the NN is expected to achieve a high level of performance in managing the complexities of MPPT in wind energy systems. In summary, the integration of neural networks in controlling nonlinear systems, particularly within the context of MPPT for wind turbines, represents a significant advancement in the field of artificial intelligence and control engineering. The ability of NNs to learn and adapt to complex dynamics not only improves the efficiency of energy systems but also paves the way for future innovations in control strategies across diverse applications [25].

C. cuckooo optimisaton

The Cuckoo Search Algorithm (CSA) is a nature-inspired optimization technique that simulates the brood parasitism behavior of certain cuckoo species to solve complex optimization problems. In CSA, a population of solutions, each representing a "cuckoo," is initialized within the search space. These cuckoos lay their eggs in the nests of other host birds, with the goal of having the host bird unknowingly raise the cuckoo's offspring. The algorithm selects the most promising solutions (nests) based on their quality, mimicking the survival strategy of cuckoo chicks [26].

CSA alternates between exploration and exploitation phases: during exploration, cuckoos perform L¶vy flights, a random search pattern that allows them to explore new areas in the search space and avoid local optima. During the exploitation phase, the algorithm focuses on refining the solutions by selecting the best nests and discarding poorly performing ones, leading to convergence towards the global optimum [27]. This process ensures a good balance between global search (exploration) and local refinement (exploitation).

The discovery of the best solution, or the "nest," depends on the fitness of the solutions, which is measured relative to the optimization objective. The L¶vy flight, which guides the cuckoos' movements, is mathematically represented by specific equations that dictate how cuckoos move in the search space [23]. The effectiveness of CSA has been demonstrated across various applications, including renewable energy systems and optimization tasks, showcasing its robustness and efficiency in navigating complex search spaces [28], [29].

The position update in CSA is modeled using the following equation:

侀(t+1) 唸 侀(t) 喃 偉佩 ¶vy姻 戴 戴

Where:

侪凯is the current position of the cuckoo, is a step size parameter, and L¶vy孀鼍is the L¶vy flight distribution controlling the random step length and direction of the cuckoo's movement.

In each iteration, a fraction of the worst-performing nests is replaced with new randomly generated solutions, ensuring diversity in the population. This mechanism enables CSA to balance exploration and exploitation, ultimately converging towards the global optimum.

By dynamically adjusting parameters like the step size and leveraging L¶vy flights, CSA provides a flexible and efficient framework for tackling complex optimization problems.

```
Initialize the cuckoo population X(i) (i
= 1, 2, ..., n)
Set the parameters pa, alpha, and other
parameters.
Calculate the fitness of each cuckoo
WHILE (t < Maximum of iterations)
  Generate new solutions by Lévy flights
  Evaluate the fitness of new solutions
   Choose a nest randomly
  IF (fitness of new solution > fitness
  of chosen nest)
    Replace the chosen nest with new
     Solutions
  END IF
  Quit a fraction (pa) of worse nests
  Keep the best solutions
  arrange the solutions and find the
   current best
  t = t + 1
END WHILE
RETURN the best solution
```

Fig. 9. cuckooo psudo code.

IV. Results and Interpretation

The purpose of this comparative study is to highlight the differences and effectiveness of the cuckooo optimization method and its effectiveness compared to previous methods:

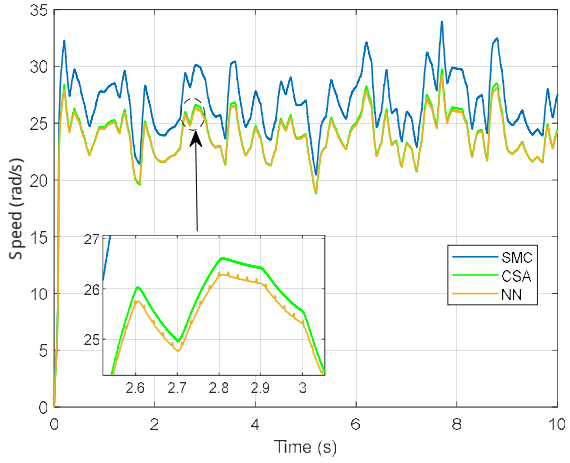


Fig. 10. Rotational speed 愠(rad/s).

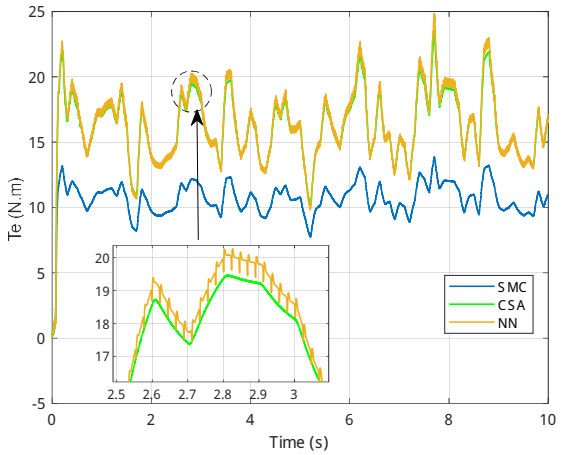


Fig. 11. Torques Ct and Cem (N.m).

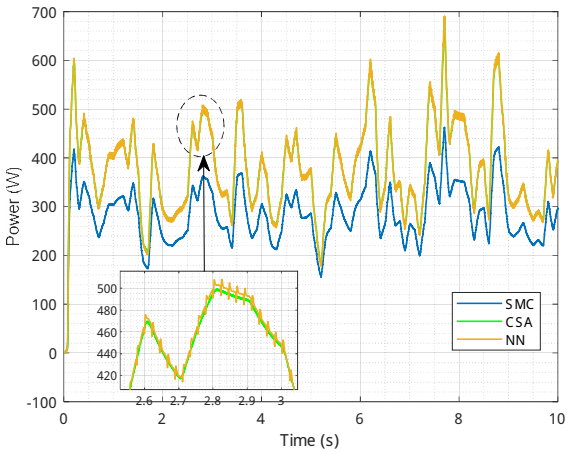


Fig. 12. Generated power (Watt).

A. Results Interpretation

I In low wind speed conditions, the power generated by all the control strategies is relatively similar due to the limited energy available. However, as wind velocity increases, the performance differences between the control strategies become more pronounced.

Sliding Mode Control (SMC): This approach exhibits moderate efficiency, with power output ranging between 160 W and 420 W. While power increases with wind speed, it does so less efficiently compared to Neural Networks (NN) or the Cuckoo Search Algorithm (CSA).

Neural Networks (NN): NN achieves a higher power output than SMC, ranging from 185 W to 690 W. This superior performance demonstrates the effectiveness of NN in harnessing energy from varying wind speeds, highlighting the significant improvements provided by its learning capability compared to traditional robust strategies like SMC.

Cuckoo Search Algorithm (CSA): CSA also achieves high power output in the range of 185 W to 690 W, closely matching the performance of NN. However, CSA shows slightly better results at higher wind speeds, delivering more stable and filtered power output.

This comparison illustrates that both NN and CSA outperform SMC, particularly in higher wind speed scenarios,

with CSA having a marginal advantage in terms of output stability.

V. Conclusion

The comparative analysis of control strategies for wind energy systems reveals significant differences in performance, particularly at higher wind speeds. While the Sliding Mode Control (SMC) strategy demonstrates moderate efficiency, it is outperformed by both Neural Networks (NN) and the Cuckoo Search Algorithm (CSA). NN's learning capability enables it to capture energy more efficiently, resulting in a notable increase in power output. Similarly, CSA offers competitive performance, slightly surpassing NN at higher wind speeds with more stable and filtered results. These findings suggest that advanced control strategies like NN and CSA provide substantial improvements in energy capture and efficiency over traditional methods such as SMC, especially in fluctuating wind conditions. Consequently, adopting these intelligent control techniques could lead to more effective wind energy harvesting, particularly in scenarios involving higher wind velocities.

References

[1] S. Muralidharan, "A comparative study of machine learning techniques for wind turbine performance prediction", E3s Web of Conferences, vol. 387, p. 04011, 2023. https://doi.org/10.1051/e3sconf/202338704011

[2] T. Demirdelen, P. Tekin, MAksu, & F. Ekinci, "The prediction model of characteristics for wind turbines based on meteorological properties using neural network swarm intelligence", Sustainability, vol. 11, no. 17, p. 4803, 2019. https://doi.org/10.3390/su11174803

[3] B. Zhang, "Study of cooperative wake control for multiple wind turbines under variable wind speeds/directions", Proceedings of the Institution of Mechanical Engineers Part a Journal of Power and Energy, vol. 237, no. 7, p. 1615-1627, 2023. https://doi.org/10.1177/09576509231176626

[4] S. Chouiekh, "Optimal control for wind turbine based on reinforcement learning", E3s Web of Conferences, vol. 469, p. 00057, 2023. https://doi.org/10.1051/e3sconf/202346900057

[5] A. TOKMAK, "Forecasting wind power generation using artificial neural network", International Journal of Pure and Applied Sciences, vol. 9, no. 1, p. 7-19, 2023. https://doi.org/10.29132/ijpas.1174444

[6] P. Qian, X. Ma, & Y. Wang, "Condition monitoring of wind turbines based on extreme learning machine",, 2015. https://doi.org/10.1109/iconac.2015.7313974

[7] B. Kim, D. Jung, Y. Jang, & K. Kang, "Allowable pitch angle of aerodynamic imbalance due to individual pitch movement for ultimate loads on offshore wind turbine using artificial neural network", Applied Sciences, vol. 12, no. 10, p. 5177, 2022. https://doi.org/10.3390/app12105177

[8] A. Asghar and X. Liu, "Online estimation of wind turbine tip speed ratio by adaptive neuro-fuzzy algorithm", International Journal of Advanced Computer Science and Applications, vol. 9, no. 3, 2018. https://doi.org/10.14569/ijacsa.2018.090306

[9] E. Assareh and M. Biglari, "A novel approach to capture the maximum power generation from wind turbines using hybrid mlp neural network and bees algorithm (hnnba)", Iete Journal of Research, vol. 62, no. 3, p. 368-378, 2015. https://doi.org/10.1080/03772063.2015.1083899

[10] S. Derakhshan, A. Tavaziani, & N. Kasaeian, "Numerical shape optimization of a wind turbine blades using artificial bee colony algorithm", Journal of Energy Resources Technology, vol. 137, no. 5, 2015. https://doi.org/10.1115/1.4031043

[11] Z. Sun and Z. Chen, "Power generation prediction method of offshore wind turbines based on cascaded deep learning", International Transactions on Electrical Energy Systems, vol. 2022, p. 1-10, 2022. https://doi.org/10.1155/2022/4404867

[12] K. AteQ "Estimation of short-term power of wind turbines using artificial neural network (ann) and swarm intelligence", Sustainability, vol. 15, no. 18, p. 13572, 2023. https://doi.org/10.3390/su151813572

[13] S. Lawan, W. Abidin, & M. Hotoro, "Development of an improved hybrid back propagation ann for low wind speed prediction and wind energy evaluation", Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, vol. 75, no. 1, p. 112-126, 2020. https://doi.org/10.37934/arfmts.75.1.112126

[14] M. Tsai, C. Tseng, Y. Hung, & K. Tsai, "A novel digital□signal□processor□based maximum□power□point tracking control design for a vertical□axis wind□turbine generation system using neural network compensator", Iet Electric Power Applications, vol. 16, no. 7, p. 761-775, 2022. https://doi.org/10.1049/elp2.12190

[15] F. Gemici and A. Hahin, "Estimation of wind speed with artificial neural networks method for isparta using meteorological measurement data", International Journal of Energy Applications and Technologies, vol. 8, no. 2, p. 65-69, 2021. https://doi.org/10.31593/ijeat.895362

[16] El Aimani, S., 2004. Mod¶lisation des diff¶rentes technologies d'¶pliennes int¶gr¶es dans un r¶seau de moyenne tension (Doctoral dissertation, Ecole Centrale de Lille).

[17] Fonseca, A., 2000. Comparaison de machines ¯ aimants permanents pour la traction de v¶hicules ¶lectriques et hybrides (Doctoral dissertation, Institut National Polytechnique de Grenoble-INPG).

[18] Kazmi, S.M.R., Goto, H., Guo, H.J. and Ichinokura, O., 2010, September. Review and critical analysis of the research papers published till date on maximum power point tracking in wind energy conversion system. In 2010 IEEE energy conversion congress and exposition (pp. 4075-4082). IEEE.

[19] X. Liu and Y. Han, "Sliding mode control for dfig-based wind energy conversion optimization with switching gain adjustment",, 2014. https://doi.org/10.1109/wcica.2014.7052892

[20] O. Zebraoui and M. Bouzi, "Robust sliding mode control based mppt for a pv/wind hybrid energy system", International Journal of Intelligent Engineering and Systems, vol. 11, no. 5, p. 290-300, 2018. https://doi.org/10.22266/ijies2018.1031.27

[21] Z. Zhang, H. Xu, J. Zou, & G. Zheng, "Tracking differentiator-based dynamic surface control for the power control of the wind generation system in the wind farm", Proceedings of the Institution of Mechanical Engineers Part I Journal of Systems and Control Engineering, vol. 227, no. 3, p. 275-285, 2013. https://doi.org/10.1177/0959651812468082

[22] H. Boulouiha, A. Allali, A. Tahri, A. Draou, & M. Denai, "A simple maximum power point tracking based control strategy applied to a variable speed squirrel cage induction generator", Journal of Renewable and Sustainable Energy, vol. 4, no. 5, 2012.

[23] T. Alzohairy, "Direct adaptive control of unknown nonlinear systems using radial basis function networks with gradient descent and k-means", International Journal of Computer Theory and Engineering, p. 775-784, 2011. https://doi.org/10.7763/ijcte.2011.v3.408

[24] Y. Bao, W. Hui, & J. Zhang, "Adaptive inverse control of variable speed wind turbine", Nonlinear Dynamics, vol. 61, no. 4, p. 819-827, 2010. https://doi.org/10.1007/s11071-010-9689-3

[25] H. Yang and S. Li, "Water quality prediction in sea cucumber farming based on a gru neural network optimized by an improved whale optimization algorithm", Peerj Computer Science, vol. 8, p. e1000, 2022. https://doi.org/10.7717/peerj-cs.1000

[26] Zaid, N., Mokhtar, M., Musirin, I., & Rahmat, N. (2015). Multi-objective optimization for sizing of distributed generation using cuckoo search algorithm. Applied Mechanics and Materials, 785, 34-37. https://doi.org/10.4028/www.scientific.net/amm.785.34

[27] Zaid, N., Mokhtar, M., Musirin, I., & Rahmat, N. (2015). Multi-objective optimization for sizing of distributed generation using cuckoo search algorithm. Applied Mechanics and Materials, 785, 34-37. https://doi.org/10.4028/www.scientific.net/amm.785.34

[28] Lakshmi, K., Visalakshi, N., Shanthi, S., & Parvathavarthini, S. (2017). Clustering categorical data using k-modes based on cuckoo search optimization algorithm. Ictact Journal on Soft Computing, 8(1), 1561-1566. https://doi.org/10.21917/ijsc.2017.0218

[29] Rajendra, K., Subramanian, S., Karthik, N., Naveenkumar, K., & Ganesan, S. (2023). State estimation for electric power system with load uncertainty and false data using cuckoo search algorithm. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4s), 209-214. https://doi.org/10.17762/ijritcc.v11i4s.6530

Whale Optimization Algorithm for Maximum Power Point Tracking in Wind Turbines

Billel Naceri
Laboratory of Electrical Engineering (LGEB)
University of Biskra
Biskra, Algeria
billel.naceri@univ-biskra.dz

Kamel Aoues
Department of Mechanical Engineering-Biskra
University of Biskra
Biskra, Algeria
k.aoues@univ-biskra.dz

Hamou Ait Abbas
Laboratory Of Materials and Durable Development-LM2D
University of Bouira
Bouira, Ageria
h.aitabbas@univ-bouira.dz

Okba Fergani
Laboratory of Electrical Engineering (LGEB)
University of Biskra
Biskra, Algeria
okba.fergani@univ-biskra.dz

Ali Mazari
Laboratory Of The Applied and Automation and Industrial Diagnostic
University of Djelfa
Djelfa, Algeria
ali.mazari@univ-djelfa.dz

Younes Belhadjer
Laboratory Of The Applied and Automation and Industrial Diagnostic
University of Djelfa
Djelfa, Algeria
younes.belhadjer@univ-djelfa.dz

Abstract´ This paper examines the Whale Optimization Algorithm (WOA) for Maximum Power Point Tracking (MPPT) in wind turbines, utilizing real wind speed data from a region known for its semi-arid climate and varied topography. The WOA, inspired by the hunting behavior of humpback whales, is evaluated for its capability to optimize energy extraction under fluctuating wind conditions. Through Matlab/Simulink simulations, WOA's performance is compared with other popular MPPT techniques to assess its effectiveness. Results demonstrate that WOA delivers a notable improvement in power output and system stability, particularly in dynamic wind environments. Although WOA presents significant advantages in terms of efficiency and simplicity, the study also identifies areas for potential improvement in tracking accuracy and response time under extreme environmental conditions. The findings highlight the importance of utilizing advanced optimization algorithms to enhance the reliability of wind energy systems.

Keywords´ Whale Optimization Algorithm (WOA); Maximum Power Point Tracking (MPPT); Wind Turbines; Optimization Algorithms; Renewable Energy; Energy Efficiency; Computational Intelligence.

I. INTRODUCTION

Wind energy has emerged as a pivotal element in the global shift towards renewable energy sources, significantly contributing to the reduction of greenhouse gas emissions and the reliance on fossil fuels. The deployment of wind turbines (WTs) facilitates clean electricity generation, which is essential for combating climate change and promoting energy independence [1], [2]. However, the optimization of wind turbine performance is crucial, as it involves balancing energy output with mechanical strain and operational costs. This optimization is particularly challenging due to the dynamic and unpredictable nature of wind conditions, necessitating advanced control mechanisms to ensure that turbines operate at their maximum power point (MPP) [3], [4].

To address these challenges, Maximum Power Point Tracking (MPPT) techniques have been developed, which are designed to ensure that wind turbines operate at peak efficiency by continuously adjusting to changing wind conditions. Among the various MPPT methods, optimization algorithms such as the Whale Optimization Algorithm (WOA) and Artificial Neural Networks (ANN) have shown considerable promise in enhancing energy extraction from wind turbines. WOA, inspired by the cooperative hunting strategies of humpback whales, effectively balances exploration and exploitation within the search space, allowing for rapid convergence on optimal operating points even under fluctuating wind conditions [5], [6]. This adaptability makes WOA a robust choice for MPPT in wind energy systems, particularly when traditional methods may falter [7].

Conversely, ANN leverages machine learning techniques to predict and optimize turbine performance based on historical data. By learning from past wind conditions and operational data, ANN-based MPPT systems can provide high precision and flexibility in real-time adjustments, outperforming conventional techniques in complex and nonlinear environments [8], [9]. The continuous learning capability of ANN allows for better adaptation to the variable nature of wind, thereby improving the overall efficiency of wind energy systems [10], [11]. While other optimization methods, such as the Cuckoo Search Algorithm (CSA), also present valuable solutions, this study primarily focuses on evaluating the performance of WOA and ANN for MPPT in wind turbines, highlighting their respective strengths in maximizing energy efficiency [12], [13].

The integration of advanced control mechanisms such as WOA and ANN into wind turbine operations is essential for optimizing energy output and enhancing the overall performance of wind energy systems. These methodologies not only improve the efficiency of energy extraction but also contribute to the sustainability of renewable energy sources, thereby playing a crucial role in addressing the environmental challenges posed by climate change [14], [15]. Future research should continue to explore and refine these techniques to further enhance the reliability and efficiency of wind energy systems.

II. WIND TURBINE MODELING

Wind turbine modeling consists of four parts: wind model, the model of the wind turbine, electrical and mechanical equations.

A. Wind Model

The wind speed at a point in space 係 is the resultant of two components: 係 varies slowly and regularly over longer periods at a given site, while the other 係 varies turbulently with time. Its expression is given by formula (1) [15].

係 係 係 (1)

XXX-X-XXXX-XXXX-X/XX/$XX.00 ø 20XX IEEE

is a turbulent component of the wind which is a stationary random process (does not vary with the mean wind speed).

is a slow component: it is the average wind speed varying regularly over longer periods at a given site.

The aero turbine filters the high frequency fluctuations. For this purpose, a low-pass filter is reconstituted for the turbulence component so that it reproduces a characteristic closer to reality, whose transfer function is given by the expression (2). [15]

$$G(f) = \frac{1}{0.11375s + 1} \quad (2)$$

The value of the time constant depends on the rotor diameter and also on the wind turbulence intensity and the average wind speed (= 0.11375s) [6].

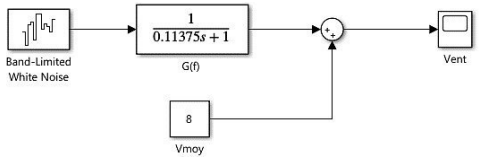


Fig. 1. Synoptic of wind reconstruction.

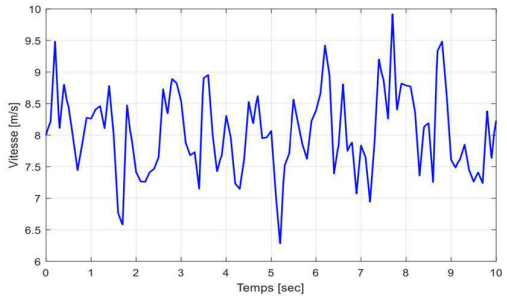


Fig. 2. Synthetic wind sequence.

The previous figure (2) represents a wind sequence generated by the Simulink model of figure (1) with an average wind speed =

The power coefficient is defined as the ratio between the power extracted from the wind (recoverable) and the total power theoretically available [16].

$$C_p = \frac{P_{recoverable}}{P_{theoretically\ available}} = \frac{P_{recoverable}}{S} \quad (3)$$

The power coefficient has a maximum of 16/27 meaning 0.59. It is this theoretical limit called Betz limit (figure 3) that sets the maximum extractable power for a given wind speed. In reality, this limit is never reached and each wind turbine is defined by its own power coefficient expressed as a function of the relative speed ≅ representing the relative speed is the ratio between the speed of the tip of the blades of the wind turbine and the wind speed. [17] The aerodynamic efficiency is then deduced:

$$\eta = \frac{P_{mec}}{P_{wind}} = C_p \quad (4)$$

In our case, the evolution of the power coefficient which is a function of ≅ and the wedge angle , is given by the following polynomial approximation:

$$C_p = \quad (5)$$

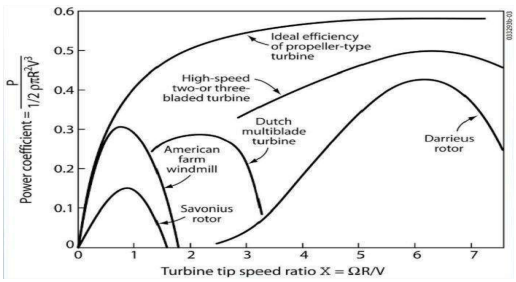


Fig. 3. Power coefficient for different types of wind turbines.

B. Model of the Wind Turbine

The turbine transforms the kinetic energy of the wind into mechanical energy. It is composed of three blades that rotate on the axes of an angle . The speed of wind V, applied on the blades of the wind turbine causes its rotation and creates a mechanical power on the shaft of the turbine [16].

The expression of the mechanical power available on the wind turbine shaft, as well as that of its torque are given as follows:

$$P_{mec} = C_p(\lambda, \beta) P_{wind} = \frac{1}{2} C_p(\lambda, \beta) \rho S V^3 \quad (6)$$

$$T_{mec} = \frac{P_{mec}}{\Omega_{mec}} = \frac{C_p(\lambda, \beta) \rho S V^3}{2\Omega_{mec}} \quad (7)$$

= Torque of the wind turbine (N.m).

The power coefficient represents the aerodynamic efficiency of the wind turbine and also depends on the turbine characteristic.

The output quantities of the turbine are the power or torque that can be controlled by varying the previous input quantities.

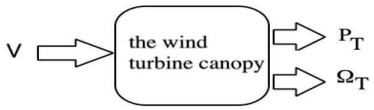


Fig. 4. Input-output of the wind turbine model.

The action of the moving air will result in forces applied at each point of the blade surface.

The block diagram of the studied blades is represented in (Figure 5).

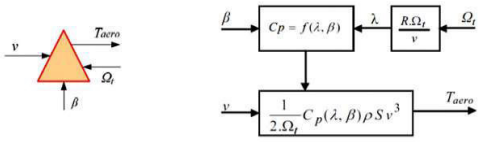


Fig. 5. Block diagram of the blade model.

The role of the gearbox is to transform the mechanical speed of the turbine into the speed of the generator, and the aerodynamic torque into the gearbox torque according to the following mathematical formula:

$$T_{mec} = \frac{T_{aero}}{G}$$

Where: : Rotational speed after multiplier in (rad/s).

: Torque after multiplier (N.m)

The fundamental equation of dynamics allows to determine the evolution of the mechanical speed from the total mechanical torque 嫣動納 嫡applied to the rotor shown in the equation below:

$$俯\frac{俥假}{俥妥} = 喵侃假_{告} 噫 俩_{告} 営俩_{動納}$$

$俩_{動納}$: The electromagnetic torque developed by the generator.

$俩_{告}$: The torque developed by the gearbox.

侃 The friction coefficient.

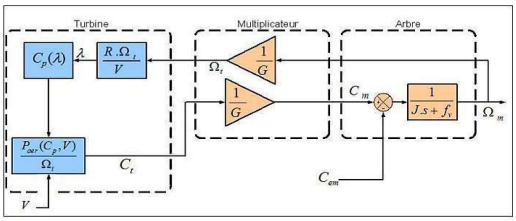


Fig. 6. Block diagram of the wind turbine model.

Classical MPPT facilitates the search for the maximum point of power, knowing the characteristic of the wind turbine, this method allows us to quickly approach the optimum with the help of simple measurements, internal to the mechanical-electrical converter, that is to say without the use of wind speed sensor [18].

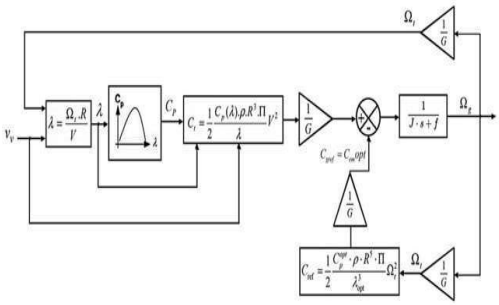


Fig. 7. scematic of the MPPT control strategy of the wind turbine.

C. Electrical equations

The electrical equations of the stator of a permanent magnet synchronous machine (PMSM) are described by The generalized theorem of the generating convention (PMSM), it˘s enough to reverse the direction of the currents; the (PMSM) equations are therefore written as follows:

$$\begin{cases} 俟_{告} 噫 営俸_{動告} 侵_{告} 営俯\frac{咨_{動告}}{咨贝} 営俯_{欲欲} 侵_{欲} 甚 \\ 俟_{欲} 噫 営俸_{動欲} 侵_{欲} 営俯\frac{咨_{動欲}}{咨贝} 営俯_{動磁} 俩_{告} 喵俣_{欲} 嫡 \end{cases} \quad (8)$$

The expression of the electromagnetic torque is written as follows:

$$俩_{動納} 噫 \frac{厦}{厦} 俯_{動磁} 侵_{告} 営俣_{欲} 侵_{欲} 嫡 \quad (9)$$

We replace the flows by their expressions given in (9), we thus have:

$$俩_{動納} 噫 \frac{厦}{厦} 俯_{動磁} 俩_{告} 営俯嫣_{動磁} 侵_{告} 営俣_{欲} 侵_{欲} 嗜 \quad (10)$$

Where:

$俟_{告}$, $俟_{欲}$ Represent respectively the voltage vectors related to the reference frame (d, q).

$侵_{告}$ $侵_{欲}$ Represent respectively the current vectors related to the reference frame (d, q).

甚: Represent the electrical speed of rotation of the rotor.

$俯_{告}$, $俯_{欲}$: Represent respectively the stator inductances of the frame (d, q).

D. Mechanical Equations

The mechanical equation below explains the dynamics:

$$俯\frac{俥假}{俥妥} 噫 俩_{納呐願} 営俩_{動納} 喵侃假$$

Being aware of:

$俩_{納呐願}$: applied motor torque to the generator.

俯 Inertia moment.

侃 Viscous friction coefficient.

III. Control Methodes Applied for MPPT

A. Sliding mode control

Sliding Mode Control (SMC) is a robust control strategy particularly suited for nonlinear systems, such as wind energy conversion systems, where it effectively manages uncertainties and fluctuations in wind speed and turbine characteristics. In the context of Maximum Power Point Tracking (MPPT) for wind turbines, SMC is employed to guide the system toward optimal power extraction despite varying conditions. The fundamental principle of SMC involves defining a "sliding surface" in the state space, which is based on the desired performance criteria related to the maximum power point (MPP) [19].

The sliding surface is typically formulated using the system's state variables, such as current, voltage, or power output, to ensure that these parameters align with the optimal operating conditions [20]. By driving the system dynamics to converge on this surface, SMC guarantees that the turbine operates near its maximum efficiency, effectively compensating for disturbances and uncertainties [19]. This approach has been shown to enhance the robustness of wind energy systems, allowing them to maintain optimal performance even under challenging operational conditions [21], [22].

The sliding surface s(t) can be expressed as:

$$俚_{動動} 噫 俄_{動動} 営俄_{動動}$$

Where:

$俄_{動動}$ is the current state variable.

$俄_{動動}$ is the reference state variable corresponds to MPP.

The objective is to design the controller so that s(t)” 0, indicating that the system state has converged to the desired operating point [21], [22].

For the system to remain on the sliding surface and achieve stability, the time derivative of the sliding surface $俚_{動動}$ must be controlled. The control law is designed to make this derivative zero when the system state is on the sliding surface, ensuring convergence to the MPP. This approach is effective for rejecting disturbances and accommodating variations in

wind speed, as the control law continually adjusts the system to maintain optimal performance [21], [22].

The equivalent control law 侮重太 is derived based on the system dynamics and is formulated to keep the system state on the sliding surface:

侮重太 喻 侵戴道描做戴道用戏

Where: f(x) and g(x) represent the nonlinear dynamics of the system.

侮重太 is the component of the control that matches the system˜s dynamics.

However, due to the inherent uncertainties and external disturbances, the equivalent control alone is insufficient. An additional switching term is introduced to correct deviations and ensure robustness [21], [22].

To enhance the robustness of the control strategy, a switching function is added to the control law. This switching component helps the system stay on the sliding surface, even when there are disturbances or variations in system parameters:

俣喻 俣重太 营依佣ó à 嬲嬲嬲

Where: K is a positive constant that determines the aggressiveness of the control response.

sign(s(t)) is the sign function, which provides a discrete adjustment to bring the system state back to the sliding surface [21], [22].

This combination of equivalent control and switching term enables the SMC to rapidly correct deviations, providing strong resilience to parameter variations and external disturbances commonly encountered in wind turbine systems.

B. ANN control

Neural networks (NN) are advanced mathematical systems designed to emulate the human brain's capability to learn and process information. Utilizing artificial neurons implemented in software, NNs facilitate the simulation of human-like intelligence, thereby promoting machine autonomy through artificial intelligence (AI). The architecture of NNs comprises multiple layers of parallel processors, allowing for complex data manipulation and analysis, as demonstrated in Figure 8 [23].

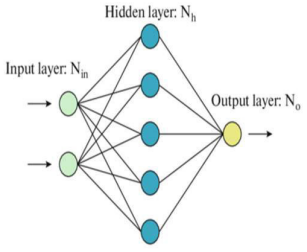


Fig. 8. NN structure.

In the realm of nonlinear systems control, NNs present an effective methodology for learning and approximating intricate dynamics. This capability enables the development of adaptive and precise control strategies, which are crucial for applications in various fields, including robotics and renewable energy. Traditional control methods often encounter significant challenges when addressing the complexities inherent in these systems. In contrast, NNs are favored for their exceptional ability to approximate continuous functions of real values, making them particularly well-suited for handling nonlinearities. The proposed NN controller for maximum power point tracking (MPPT) in wind turbines features a structure with two hidden layers. The inputs to the NN controller consist of the input current generated by the wind turbine and the output voltage from the boost converter. The primary output of the NN is the duty cycle (D), which modulates the operation of the boost converter to optimize energy extraction from the wind. The training of the NN was conducted using the (trainlm) function in Matlab, which employs the Levenberg-Marquardt (LM) algorithm [24]. This algorithm is renowned for its effectiveness in solving least-squares problems, thereby enhancing the training process of the neural network. Through this approach, the NN is expected to achieve a high level of performance in managing the complexities of MPPT in wind energy systems. In summary, the integration of neural networks in controlling nonlinear systems, particularly within the context of MPPT for wind turbines, represents a significant advancement in the field of artificial intelligence and control engineering. The ability of NNs to learn and adapt to complex dynamics not only improves the efficiency of energy systems but also paves the way for future innovations in control strategies across diverse applications [25].

C. Whale optimisaton

The Whale Optimization Algorithm (WOA) is a nature-inspired optimization technique that simulates the hunting behavior of humpback whales to solve complex optimization problems. Initially, WOA initializes a population of solutions, each representing a "whale" in the search space. These whales adjust their positions relative to the best-known solution, termed the "prey," by encircling it in a spiral motion, mimicking the whales' hunting strategy Yang & Li [26]. The algorithm alternates between exploration and exploitation phases: during exploration, whales perform a random search for new solutions to escape local optima, while in the exploitation phase, they refine their positions based on the current best solution, converging towards the global optimum [27].

The position of the prey is crucial for the encircling behavior, as it is assumed to represent the optimal solution or be close to it. This encircling mechanism is mathematically represented by specific equations that dictate how whales adjust their positions based on the prey's location [23]. The effectiveness of WOA has been demonstrated across various applications, including renewable energy systems and optimization tasks, showcasing its robustness and efficiency in navigating complex search spaces [28], [29].

侖喻 |侵佣 ^ 营侥 − 侥 |

where D is the distance between the whale and the prey, C is a coefficient that controls the spiral motion, 侥 is the position of the prey (the best-known solution), and X is the current position of the whale.

侥喻 侥 − 营律侵侖

Where A is another coefficient that influences the movement of the whale. This equation updates the position of the whale based on the distance to the prey and the current best position.

The values of A and C are dynamically adjusted during the algorithm's execution, allowing the whales to balance exploration and exploitation effectively. The equations governing the WOA facilitate a structured approach for the whales to navigate the search space, ultimately aiming to converge toward the global optimum.

```
Initializing the whale's population X(i)
(i = 1, 2, ..., n)
Initializing a, A, C, l, and p
Calculating the fitness of each search
operative
X* = best search operative
WHILE (t < Maximum iterations)
  FOR each search operative
    Update a, A, C, l, and p
      IF (p < 0.5)
        IF (|A| < 1)
          Update the location of the
           current search operative
        ELSE IF (|A| >= 1)
          Select a random search operative
          Update the location of the
           current search operative
      ELSE IF (p >= 0.5)
        Update the location of the current
         search operative
  END FOR
  Check if any search operative goes
   outside the search space and adjust it
  Calculate the fitness of each search
   operative
  Update X* if a better solution is
   available
    t = t + 1
END WHILE
RETURN X*
```

Fig. 9. Whale psudo code.

IV. Results and Interpretation

The purpose of this comparative study is to highlight the differences and effectiveness of the whale optimization method and its effectiveness compared to previous methods:

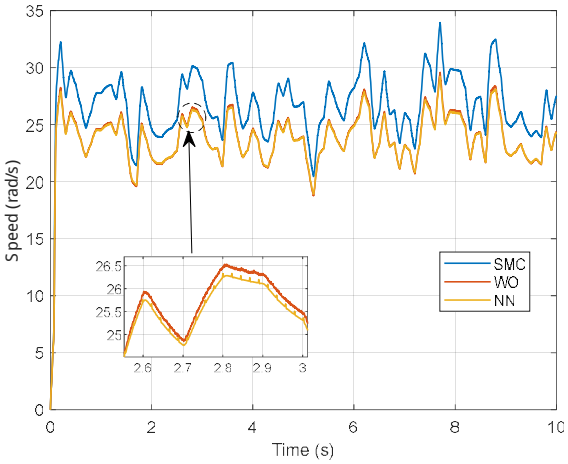


Fig. 10. Rotational speed 僅(rad/s).

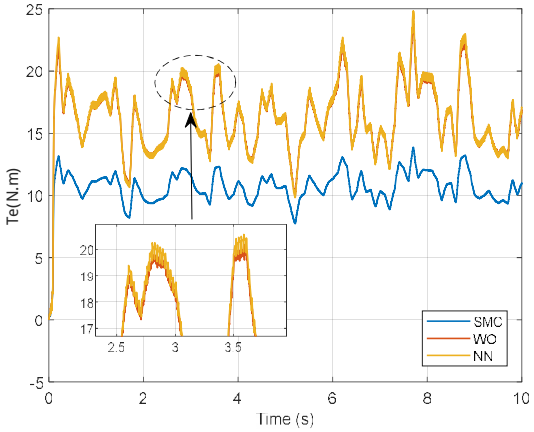


Fig. 11. Torques Ct and Cem (N.m).

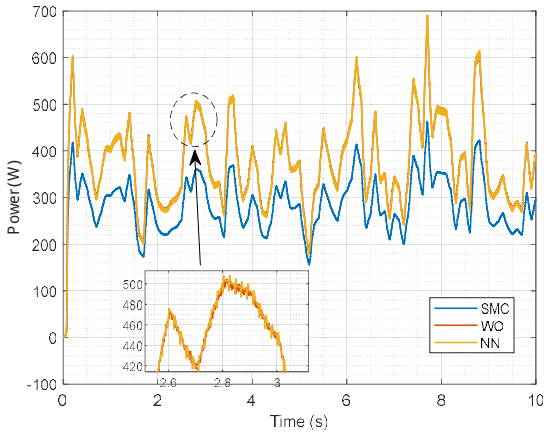


Fig. 12. Generated power (Watt).

A. Results Interpretation

c In low wind speed conditions, the power generated by all the control strategies is relatively similar due to the limited energy available. However, as wind velocity increases, the performance differences between the control strategies become more pronounced.

Sliding Mode Control (SMC): This approach exhibits moderate efficiency, with power output ranging between 160 W and 420 W. While power increases with wind speed, it does so less efficiently compared to Neural Networks (NN) or the Whale Optimization Algorithm (WOA).

Neural Networks (NN): NN achieves a higher power output than SMC, ranging from 185 W to 690 W. This superior performance demonstrates the effectiveness of NN in harnessing energy from varying wind speeds, highlighting the significant improvements provided by its learning capability compared to traditional robust strategies like SMC.

Whale Optimization Algorithm (WOA): WOA also achieves high power output in the range of 185 W to 690 W, closely matching the performance of NN. However, WOA shows slightly better results at higher wind speeds, delivering more stable and filtered power output.

This comparison illustrates that both NN and WOA outperform SMC, particularly in higher wind speed scenarios, with WOA having a marginal advantage in terms of output stability.

V. CONCLUSION

the comparative analysis of control strategies for wind energy systems reveals significant differences in performance, particularly at higher wind speeds. While the Sliding Mode Control (SMC) strategy demonstrates moderate efficiency, it is outperformed by both Neural Networks (NN) and the Whale Optimization Algorithm (WOA). NN's learning capability enables it to capture more energy efficiently, resulting in a notable increase in power output. Similarly, WOA offers competitive performance, slightly surpassing NN at higher wind speeds with more stable and filtered results. These findings suggest that advanced control strategies like NN and WOA provide substantial improvements in energy capture and efficiency over traditional methods such as SMC, especially in fluctuating wind conditions. Consequently, adopting these intelligent control techniques could lead to more effective wind energy harvesting, particularly in scenarios involving higher wind velocities.

REFERENCES

[1] S. Muralidharan, "A comparative study of machine learning techniques for wind turbine performance prediction", E3s Web of Conferences, vol. 387, p. 04011, 2023. https://doi.org/10.1051/e3sconf/202338704011

[2] T. Demirdelen, P. Tekin, MAksu, & F. Ekinci, "The prediction model of characteristics for wind turbines based on meteorological properties using neural network swarm intelligence", Sustainability, vol. 11, no. 17, p. 4803, 2019. https://doi.org/10.3390/su11174803

[3] B. Zhang, "Study of cooperative wake control for multiple wind turbines under variable wind speeds/directions", Proceedings of the Institution of Mechanical Engineers Part a Journal of Power and Energy, vol. 237, no. 7, p. 1615-1627, 2023. https://doi.org/10.1177/09576509231176626

[4] S. Chouiekh, "Optimal control for wind turbine based on reinforcement learning", E3s Web of Conferences, vol. 469, p. 00057, 2023. https://doi.org/10.1051/e3sconf/202346900057

[5] A. TOKMAK, "Forecasting wind power generation using artificial neural network", International Journal of Pure and Applied Sciences, vol. 9, no. 1, p. 7-19, 2023. https://doi.org/10.29132/ijpas.1174444

[6] P. Qian, X. Ma, & Y. Wang, "Condition monitoring of wind turbines based on extreme learning machine",, 2015. https://doi.org/10.1109/iconac.2015.7313974

[7] B. Kim, D. Jung, Y. Jang, & K. Kang, "Allowable pitch angle of aerodynamic imbalance due to individual pitch movement for ultimate loads on offshore wind turbine using artificial neural network", Applied Sciences, vol. 12, no. 10, p. 5177, 2022. https://doi.org/10.3390/app12105177

[8] A. Asghar and X. Liu, "Online estimation of wind turbine tip speed ratio by adaptive neuro-fuzzy algorithm", International Journal of Advanced Computer Science and Applications, vol. 9, no. 3, 2018. https://doi.org/10.14569/ijacsa.2018.090306

[9] E. Assareh and M. Biglari, "A novel approach to capture the maximum power generation from wind turbines using hybrid mlp neural network and bees algorithm (hnnba)", Iete Journal of Research, vol. 62, no. 3, p. 368-378, 2015. https://doi.org/10.1080/03772063.2015.1083899

[10] S. Derakhshan, A. Tavaziani, & N. Kasaeian, "Numerical shape optimization of a wind turbine blades using artificial bee colony algorithm", Journal of Energy Resources Technology, vol. 137, no. 5, 2015. https://doi.org/10.1115/1.4031043

[11] Z. Sun and Z. Chen, "Power generation prediction method of offshore wind turbines based on cascaded deep learning", International Transactions on Electrical Energy Systems, vol. 2022, p. 1-10, 2022. https://doi.org/10.1155/2022/4404867

[12] K. AteQ "Estimation of short-term power of wind turbines using artificial neural network (ann) and swarm intelligence", Sustainability, vol. 15, no. 18, p. 13572, 2023. https://doi.org/10.3390/su151813572

[13] S. Lawan, W. Abidin, & M. Hotoro, "Development of an improved hybrid back propagation ann for low wind speed prediction and wind energy evaluation", Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, vol. 75, no. 1, p. 112-126, 2020. https://doi.org/10.37934/arfmts.75.1.112126

[14] M. Tsai, C. Tseng, Y. Hung, & K. Tsai, "A novel digital□signal□ processor□based maximum□power□point tracking control design for a vertical□axis wind□turbine generation system using neural network compensator", Iet Electric Power Applications, vol. 16, no. 7, p. 761-775, 2022. https://doi.org/10.1049/elp2.12190

[15] F. Gemici and A. Hahin, "Estimation of wind speed with artificial neural networks method for isparta using meteorological measurement data", International Journal of Energy Applications and Technologies, vol. 8, no. 2, p. 65-69, 2021. https://doi.org/10.31593/ijeat.895362

[16] El Aimani, S., 2004. Mod¶lisation des diff¶rentes technologies d'¶pliennes int¶gr¶es dans un r¶seau de moyenne tension (Doctoral dissertation, Ecole Centrale de Lille).

[17] Fonseca, A., 2000. Comparaison de machines ¯ aimants permanents pour la traction de v¶hicules ¶lectriques et hybrides (Doctoral dissertation, Institut National Polytechnique de Grenoble-INPG).

[18] Kazmi, S.M.R., Goto, H., Guo, H.J. and Ichinokura, O., 2010, September. Review and critical analysis of the research papers published till date on maximum power point tracking in wind energy conversion system. In 2010 IEEE energy conversion congress and exposition (pp. 4075-4082). IEEE.

[19] X. Liu and Y. Han, "Sliding mode control for dfig-based wind energy conversion optimization with switching gain adjustment",, 2014. https://doi.org/10.1109/wcica.2014.7052892

[20] O. Zebraoui and M. Bouzi, "Robust sliding mode control based mppt for a pv/wind hybrid energy system", International Journal of Intelligent Engineering and Systems, vol. 11, no. 5, p. 290-300, 2018. https://doi.org/10.22266/ijies2018.1031.27

[21] Z. Zhang, H. Xu, J. Zou, & G. Zheng, "Tracking differentiator-based dynamic surface control for the power control of the wind generation system in the wind farm", Proceedings of the Institution of Mechanical Engineers Part I Journal of Systems and Control Engineering, vol. 227, no. 3, p. 275-285, 2013. https://doi.org/10.1177/0959651812468082

[22] H. Boulouiha, A. Allali, A. Tahri, A. Draou, & M. Denai, "A simple maximum power point tracking based control strategy applied to a variable speed squirrel cage induction generator", Journal of Renewable and Sustainable Energy, vol. 4, no. 5, 2012.

[23] T. Alzohairy, "Direct adaptive control of unknown nonlinear systems using radial basis function networks with gradient descent and k-means", International Journal of Computer Theory and Engineering, p. 775-784, 2011. https://doi.org/10.7763/ijcte.2011.v3.408

[24] Y. Bao, W. Hui, & J. Zhang, "Adaptive inverse control of variable speed wind turbine", Nonlinear Dynamics, vol. 61, no. 4, p. 819-827, 2010. https://doi.org/10.1007/s11071-010-9689-3

[25] H. Yang and S. Li, "Water quality prediction in sea cucumber farming based on a gru neural network optimized by an improved whale optimization algorithm", Peerj Computer Science, vol. 8, p. e1000, 2022. https://doi.org/10.7717/peerj-cs.1000

[26] M. Paramasivam, "Whale optimization algorithm and internet of things for horizontal axis solar tracker-basedload optimization", Indonesian Journal of Electrical Engineering and Computer Science, vol. 32, no. 3, p. 1278, 2023. https://doi.org/10.11591/ijeecs.v32.i3.pp1278-1287

[27] A. Mittal, "Enhancing solar power forecasting using grasshopper optimization and whale optimization algorithm", jes, vol. 20, no. 3, p. 2054-2059, 2024. https://doi.org/10.52783/jes.4005

[28] S. Shukla, A. Choudhary, S. Pandey, V. Sachan, & P. Kumar, "Whooph : whale optimization based optimal placement of hub node within a wban",, 2022. https://doi.org/10.21203/rs.3.rs-2126065/v1

[29] J. Tang, "A whale optimization algorithm based on atom-like structure differential evolution for solving engineering design problems", Scientific Reports, vol. 14, no. 1, 2024.

Performance Assessment of Photovoltaic Power Plant of Naàma

HABBATI BELLIA Assia
Faculty of Technology
Tahri Mohamed University
Bechar, Algeria
bellia_abdeljalil@yahoo.fr

MOULAY Fatima
Faculty of Technology
Djilali Liabes University
Sidi bel abbes, Algeria
moulayfatima@yahoo.fr

***Abstract*—This study is devoted to the evaluation of a large-scale photovoltaic power plant (LS-PVPP) in the high plains of Algeria. The plant, which is located near the city of Naàma, known for its altitude exceeding 1000 m, is designed to provide up to 20 MW in an arid climate. The simulations were carried out with the PVsyst software, which allowed making a comparison between the simulated values and the real values of some parameters obtained during a visit to the plant. Among all the losses recorded, the loss due to the field temperature remains the highest -11.8% due to the limitations of polycrystalline cell technology but the results indicate a performance ratio of 81.4%,. A similar station installed in Adrar saw production drop by 40% due to the high temperature in summer. Although the main objective of the work is on the performance of the station, a comparison with another station similar in power is necessary to discuss some parameter.)**

Keywords— Large-scale photovoltaic power plant, arid climate, Performance assessment, grid connected

I. INTRODUCTION

Man isalwayslooking for new forms of clean renewableenergy, withlow impact on the environment and from inexhaustible sources. Wind and solarenergymeet all theserequirementswith an additionalcriterion of size and whichisabundance.According to the International RenewableEnergy Agency (IRENA), global renewableenergycapacitynearlydoubledfrom 2019 to 2020 to reach 260 gigawatts (GW) [1]. The mostauthoritative global source of energyanalysis and projections examines how the contours of a new, more electrifiedenergy system are cominginto focus as global electricitydemandsoars, growingattwice the pace of overallenergydemand over the last decade. [2]

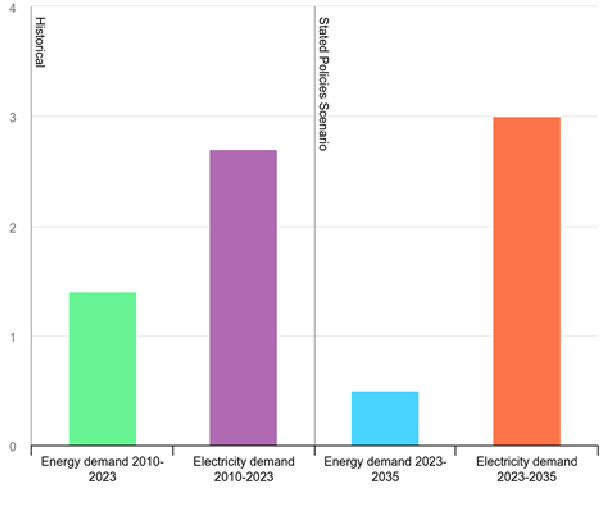


Fig. 1. Annual energy / electricity demand growth, historical and in the Stated Policies Scenario, 2010-2035 [2]

This encourages progress in the field of solarenergy, whichrepresents a veryabundant source in Algeria. The country has one of the highestsolardeposits in the world. The duration of irradiance over almost the entire national territoryexceeds 2000 hoursannually and canreach 3900 hours (high plateaus and Sahara). The energyreceivedannually on a horizontal surface of 1m² isnearly 3 KWh/m² in the north and exceeds 5.6 KWh/m in the Far South.[3].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

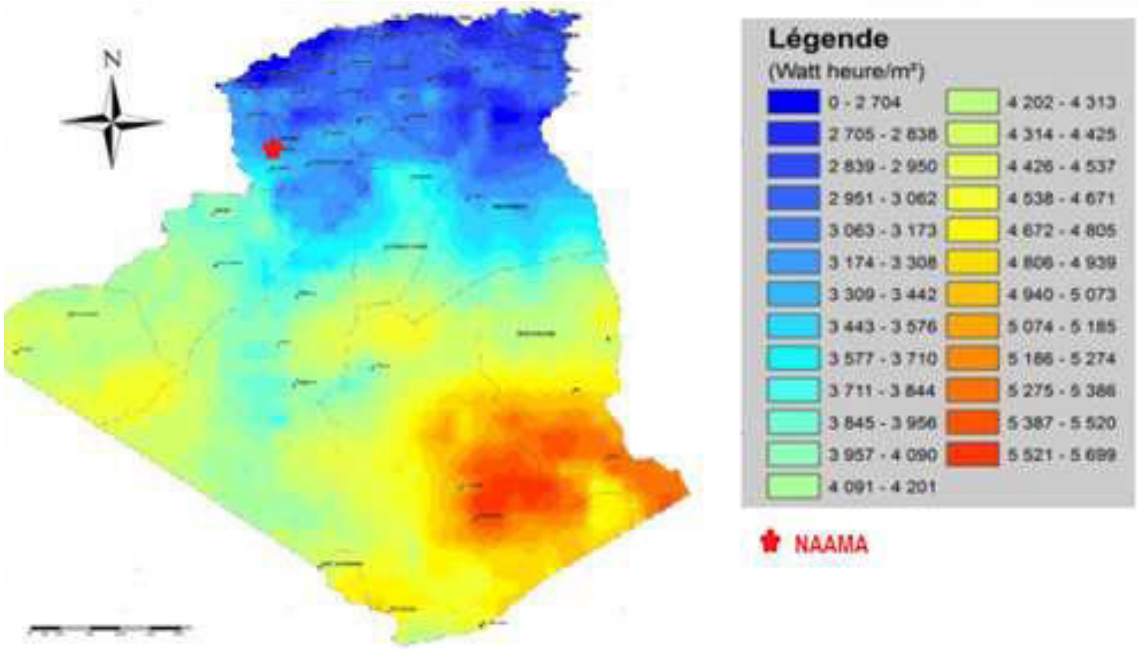


Fig. 2. Map of Average Annual Direct Global Irradiation (Period 2002-2011) [3]

However, without a strongpoliticalwill, this inexhaustible and widespreadpotentialmayremainunexploiteduntil all gas and oildeposits are exhausted. According to the report of the 2020-2024 economicrecovery plan, drawn up by the Government, naturalgasis the main source of energy in Algeriawith a rate of 65% of the energy mix, whileoilonlyrepresents 35%. This energy (naturalgas) is the main source of electricity production with a rate of 99%, in addition to its high use in industrialactivities. As a result, national consumptionisincreasingsharplyunder the influence of population growthwhile production isstagnating [3].

The energy transition isbased on the development of a national sector of renewableelectricity production equipment, particularlysolar. The energy transition canonlybebased on the development of a real national industry of renewableelectricity production equipment, particularlysolar. Importing a solar panel, for example, meansimporting the energyused to produceit. Real energyindependencewouldthereforeconsist of producingone'sownmeans of energy production. Finally, in 2011, the country'sdecision-makerslaunched a plan for the development of renewableenergies and energyefficiency, whichwasupdatedin 2015. Amongotherthings, the plan focuses on the deployment of large-scalephotovoltaicsolar installations. According to the Ministry of Energywebsite, the objective is to install 22,000 MW by 2030. The projectincludes large solar power plants connected to the gridwith a total capacity of 13,500 MW. [3,4]

The first part of thispapersimulates and discusses the system in terms of total generatedenergy, performance ratio, capacity factor and efficiency. The configuration adoptedis the one presented in the PVsyst software report, used to simulate the production of Naama station. The second part discusses the impact of the fixed tilt angle of the solar panels in the station on the generation of differentlosses and consequently, on the monthly and annual production of the energyproduced.

II. DESCRIPTION OF NAÀMA POWER PLANT

The study focuses on the 20 MWc photovoltaic power plant (LS PVPP) of Naàma. The city is located on the high plains of western Algeria, a region with high solar potential and an altitude exceeding 1000 m.

The plant is part of a program carried out by the renewable energy subsidiary of the public electricity supplier Sonelgaz (SKTM).

The program includes photovoltaic stations covering the entire country with a total energy capacity of 400 MW.[4]

The solar park is located in SedretLeghzel, at the west of the city, with coordinates 33.27 N and 0.37 W, occupying an area of 31.3 ha. The plant consists of 79680 photovoltaic modules whose characteristics are presented in the table (1). The station also has 20 inverters and 10 transformers. The structure has a fixed inclination of 15 ° C, oriented sound.

TABLE I. ELECTRICAL DATA AT STC

Electrical Data CS6P 250P	
Nominal Max. Power (Pmax)	250 W
Opt. Operating Voltage (Vmp)	30.1 V
Opt. Operating Current (Imp)	8.30 A
Open Circuit Voltage (Voc)	37.2 V
Short Circuit Current (Isc)	8.87 A
Module Efficiency	15.54%
Operating Temperature	-40°C ~ +85°C

The AC electrical network has a configuration withtwoloops of 5 SKIDs, each of the SKIDsbeingcomposed of twoparallelinverter modules and one transformer. Each of the loopscollects the power of 5 transformers via 5 RMU units in the collection station of the PV installation.

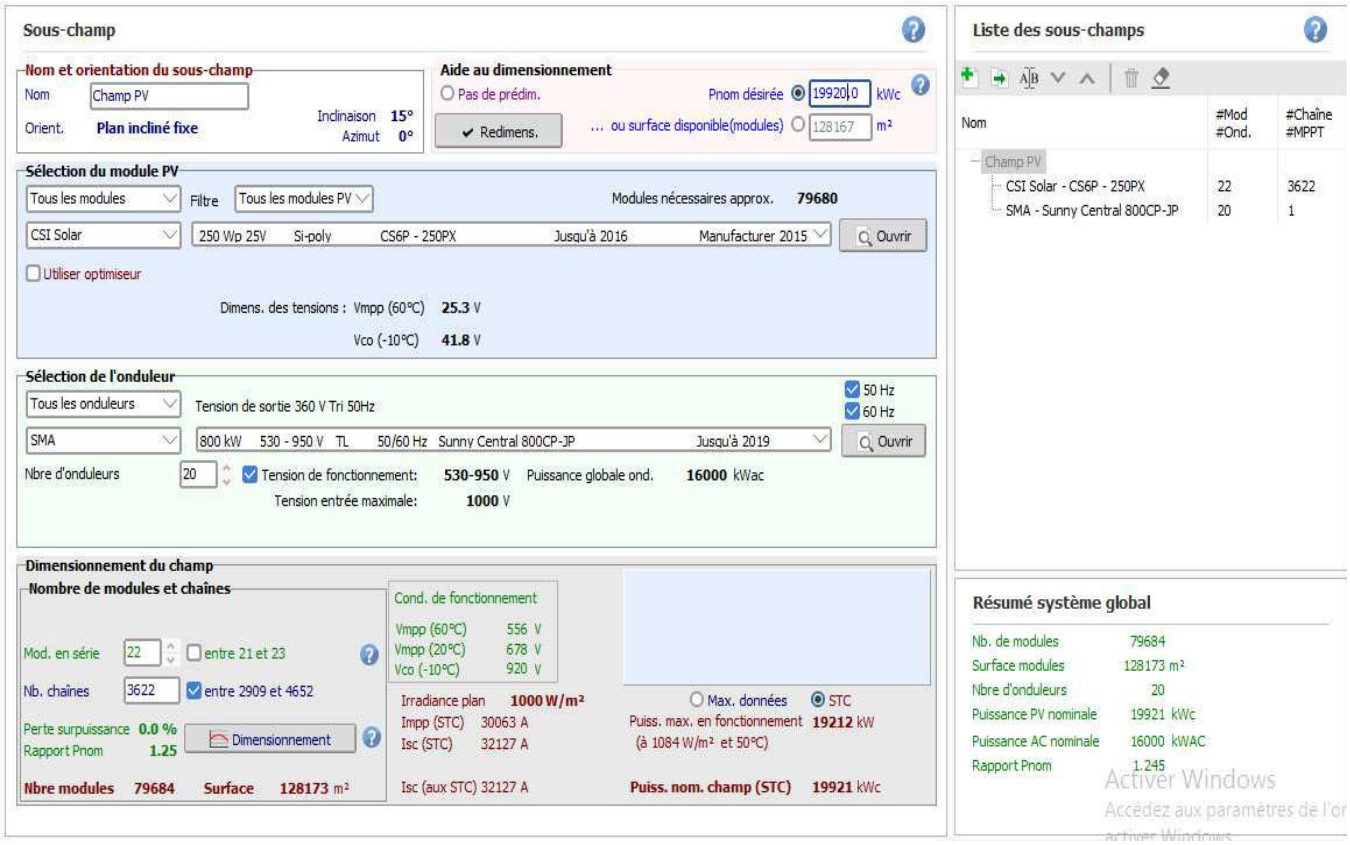


Fig. 3. PV system design parameters (from PVsyst 7.4 software simulation

Fig. 4 shows the trajectory of the sunwhichis the periodbetweensunrise and sunset, duringwhich the sunreaches a maximum heightat a time that varies depending on the days of the year. For the case of the Naama site, the maximum heightisreachedthroughout the year, approximatelyaround 1 p.m.

Note that the maximum height corresponds to the time when the system reachesits maximum efficiency, but does not correspond throughout the year to the maximum temperature of the day.

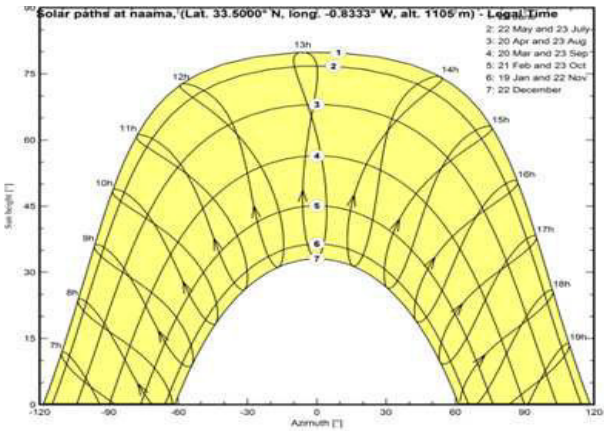


Fig. 4. Solar paths

III. QUANTITATIVE PERFORMANCE ASSESSMENT

Severalfactors are actors in determining the photovoltaicsystem's performance, such as geographic location of the site, solarpotential, energyefficiency, and losses. The IEC 61724 standard uses a quantitative evaluationmethodologybased on specificparameters to evaluate the performance of a GCPV system[5-7].

A. *Reference yield (Yr)*

The ratio between the total solar radiation incident on the surface of the PV solar panels Ht (kWh/m2) and the reference radiation quantity G0 (1 kW/m2).

$$Y_r = \frac{H_t}{G_0} \quad \text{(hours)}$$

G0 : is the reference irradiance at STC (1000 W/m²)

B. *Final yield (Yf)*

The ratio between the total energyproduced by the system Eac (kWh), and the installed nominal power P0 (kWp).

$$Y_f = \frac{E_{ac}}{P_0} \quad \text{(hours)}$$

.

C. *Performance ratio (PR)*

A parameterthatindicates the overalleffect of the losses on the energy production of the rows of a PV system, and isdefined by the ratio between the final yieldYf and the referenceyieldYr.

$$PR = \frac{Y_f}{Y_r}(\text{hours})$$

D. Capacity factor (CF)

The ratio between the actualenergy output in AC current and the amountthatcouldbegenerated if itoperatedat full nominal capacitycontinuouslythroughout the year (8760 h).

$$CF = \frac{E_{ac}/P_{MPP\,rated}}{24 * nuber\ of\ day\ in\ month} * 100(\%)$$

.

E. PV module efficiency (η_pv)

The ratio betweenEdc power and solarenergycaptured. It canbecalculated for the day or the month.

$$\eta_{pv} = \frac{E_{dc}}{H * A_m} * 100\ \ (\%)$$

H : The solar energy received by the PV system, and $A_m$its area.

F. Inverter efficiency (η_inv)

The ratio between DC power output Edc of the system and itsenergy output Eac. The monthlyinverterefficiencyis:

$$\eta_{in} = \frac{E_{dc}}{E_{ac}} * 100\ \ (\%)$$

G. System efficiency (η_sys)

$$\eta_{pv} = \frac{E_{ac}}{H * A_m} * 100\ \ (\%)$$

H. Losses

Many components in PV systemscanbe the source of energylosses. The mostcommonphenomenathatcan cause DC sidelosses are malfunction, shading, fouling, and component failure. AC sidelosses are mainlycaused by inefficiencies in transformers and inverters. Losses, which influence the production of the PV system, canbesummarized as system losses (inverterloss, inverter-threshold power loss, inverterloss on voltage, night consumption) and capture losses (ohmiclosses of wiring, loss due to fieldtemperature).

TABLE II. COMPARATIVETABLE OF EQUIPMENTS

	Characteristics	
Equipment & the design parameters	**Naàma**	**Adrar**
Module type	Poly-crystalline silicon	Poly-crystalline silicon
Photovoltaic module efficiency	15.54%	15%
Orientation and tilt	15◦ , south	26.5◦ , south
Installation type	Fixed	Fixed
Distance between photovoltaic rows	6 m	8 m
Inverters	20*(800 kWac)	40*(500 kW)
Transformers	20*(1800 kVA/900kVA, 900kVA, 30kV/ 360V ,360V)	20*(1.250 kVA/360kVA, 315 V/31.5 kV)
modules	79680* (250Wp)	81 840* (245 Wp)

IV. COMPARISON OF NAÀMA AND ADRAR 20 MWP PHOTOVOLTAIC POWER PLANT PARAMETERS

As the Adrar and Naama stations are part of the same SKTM program, it is interesting to compare their equipment and design parameters.

This comparison will help explain any changes in energy production, relative to meteorological data. Table II summarizes the Characteristics of all equipment and design parameters. Data were provided in the photovoltaic station of Naàma. They are compared by data collected in papers whose authors worked on Adrar photovoltaic station.[5-7]

TABLE III. COMPARISON OF WEATHER AND TILT DEPENDENT GENERATION

	GlobHor kWh/m²	DiffHor kWh/m²	TAmb °C	α= 15°				α = 33.5°			
				GlobInc kWh/m²	GlobEff kWh/m²	EArray kWh	E_Grid kWh	GlobInc kWh/m²	GlobEff kWh/m²	EArray kWh	E_Grid kWh
Jan	102.6	30.44	7.68	136.1	131.0	2443106	2403193	166.3	163.3	2964643	2914734
Feb	116.2	29.05	9.66	144	139.4	2542099	2501187	166.7	163.5	2916802	2868299
Mar	169.0	45.15	14.30	192.5	186.8	3279223	3226308	206.2	201.2	3483245	3425947
Apr	202.4	54.20	18.03	214.2	207.6	3549777	3493416	212.2	205.5	3497932	3442023
May	220.6	65.16	22.83	221.8	215.6	3585752	3525729	206.6	199.5	3341973	3285804
Jun	234.7	67.34	27.83	230.2	223.9	3617793	3559716	208.7	201.2	3292068	3239280
Jul	234.4	74.46	32.42	233.8	227.5	3551442	3494687	215.6	208.4	3291634	3238720
Aug	215.7	70.43	30.71	224.3	218.4	3462491	3408148	217.6	211.3	3351940	3299270
Sep	172.8	51.25	25.23	191.2	185.4	3075239	3027032	198.9	193.5	3174934	3124506
Oct	146.3	38.27	20.35	175.9	170.6	2936831	2889750	198.2	194.3	3276902	3223392
Nov	112.4	27.32	12.71	146.7	141.5	2559499	2518869	176.9	173.7	3057949	3007731
Dec	95.2	23.97	8.88	129.7	125.1	2320279	2281231	161.4	158.9	2859823	2809858
Year	2022.3	577.04	19.28	2240.3	2172.9	36923531	36329265	2335.2	2274.3	38509844	37879564

V. EFFECT OF TILT ANGLE

In the literature, it is recommended for the northern hemisphere to take the value of the latitude as the value of the tilt angle of the photovoltaic modules. The table (2) represents the variations over the year of the global irradiance, the effective irradiance, the energy produced and the energy injected into the network in the real case of 15° of tilt and the theoretically optimal case of 33.5°. The results show that the tilt angle influences the global incident irradiance because this irradiance is computed from the Horizontal Global and Diffuse irradiances in hourly values. It depends on the geographical coordinates. It is the full irradiance as received by the tilted plane.

However, since the position of the sun changes depending on the time and the season, it is noted that for the applied inclination (15°), the values are maximum from April to August with a significant difference between the performance ratio at this period of the year and the remaining months as shown on fig (5). On the other hand, the difference becomes less important for a theoretically optimal inclination (33.5°) and the hot months show a decrease in production against an increase during the cold months as seen on fig (6).

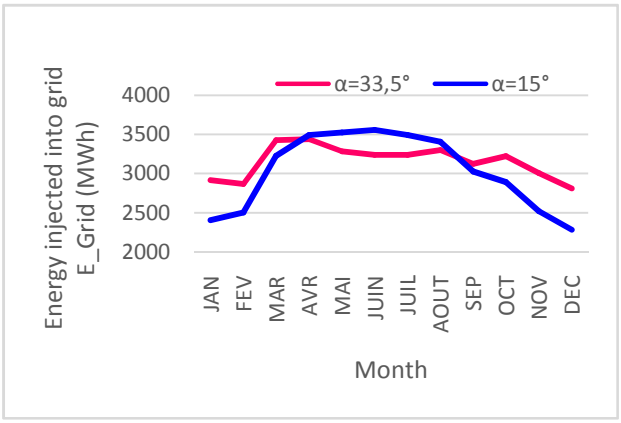


Fig. 5. Energy injected into the grid for α =15° and α = 33.5°

VI. COMPARAISONOF LOSS DIAGRAMS

Although the losses due to temperature are higher for the 33.5° inclination, -12.07% against -11.8% for 15°, the energy injected to the network is higher for the first case because of the global incident in collector plane which is 15.5% for 33.5° against +10.8% for 15°. The losses due to IAM factor on global are smaller for the 33.5° inclination than for the 15° inclination. The rest of the losses are without significant difference as in fig. 6

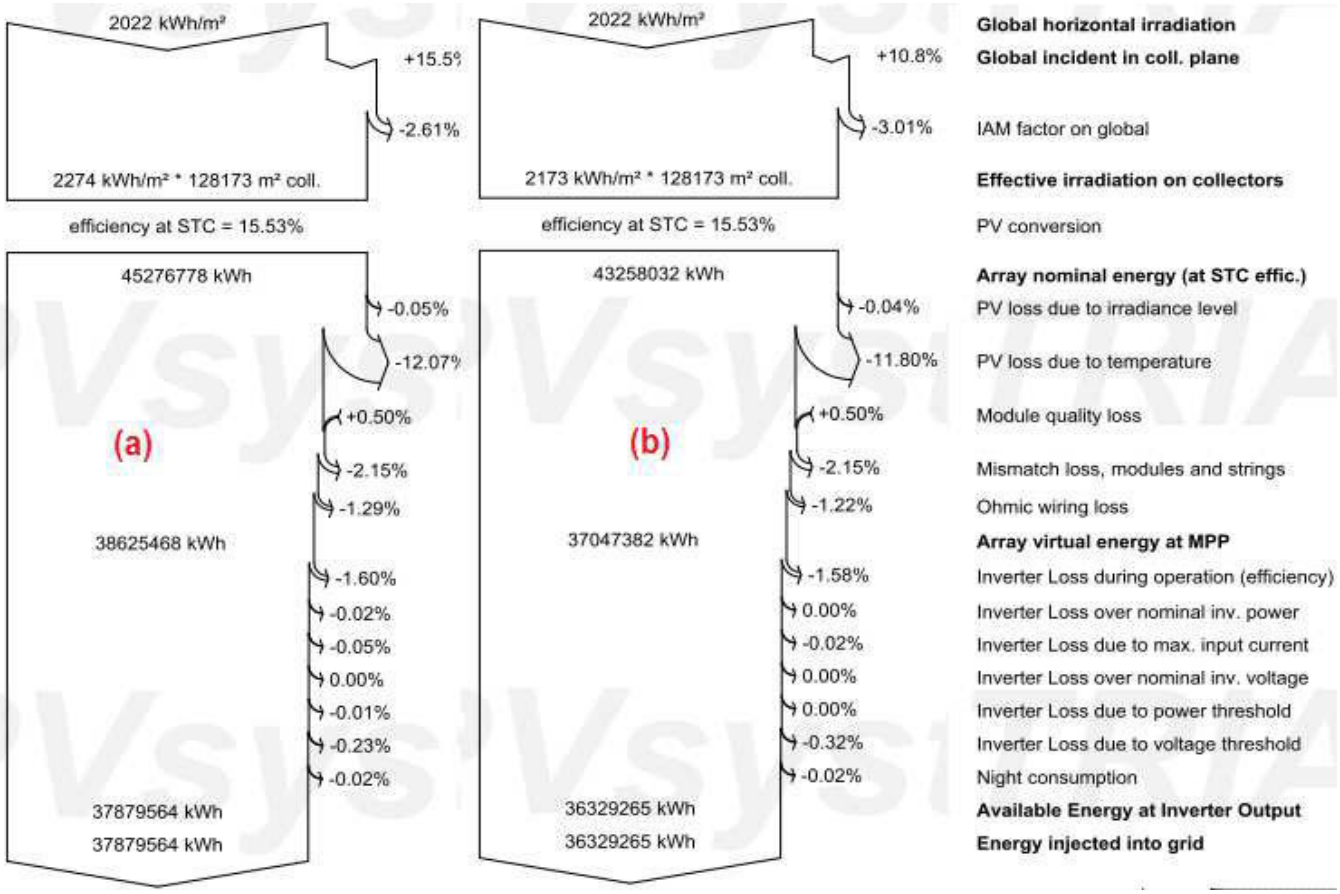


Fig. 6. Loss diagrams, (a) for tilt angle α = 33.5°, (b) for α = 15°

VII. CONCLUSION

The Naama photovoltaic power plant of type (LS-PVPP), injects up to 20 MW into the MT network and operates in an arid climate. Its performance is characterized by the influence of sunshine, air temperature, and photovoltaic cell technology as demonstrated in the Adrar station which is very similar to that of Naama and which resulted in underperformance. The tilt angle influences significantly the energy injected into the grid, mainly due to the global incident in coll plane, which isthe result of the transposition from horizontal tothe tilted plane.

No experimental study on an academic scale has been found on the physical behavior of the material used in the production of the cell (polycrystalline silicon or other) in the studied station.

But a study was carried out in the Adrar station and the simulation showed that the performance of thin-film solar cells is better in hot climate conditions. Although the climate in Naama is not classified as hot but rather arid, the climate change that the region is experiencing as well as the whole world, dictates to take into consideration the studies of Adrar not as a comparative but rather as an advisory one.

REFERENCES

[1] Renewable-Capacity-Statistics-2020, Mars 2020, https://www.irena.org/publications

[2] The International Energy Agency (IEA), https://www.iea.org

[3] Energies Nouvelles, Renouvelables et Maitrise de l'Energie, https://www.energy.gov.dz

[4] Les réalisations de SKTM dans le domaine des Energies Renouvelables, https://era.dz/event/

[5] S.Bentouba, M.Bourouis, N. Zioui, A.Pirashanthan, D. Velauthapillai, ''Performance assessment of a 20 MW photovoltaic power plant in a hot climate using real data and simulation tools,'' Energy Reports, 2021, pp.7297–7314.

[6] I. Bendaas, K. Bouchouicha, S.Semaoui, A. Razagui, S.Bouchakour, S. Boulahchiche, "Performance evaluation of large-scale photovoltaic power plant in Saharan climate of Algeria based on real data," Energy for Sustainable Development, 76 (2023), 101293

[7] R. Ihaddadene, M. E. H. Jed, N. Ihaddadene, A. D. Sou, "Analytical assessment of Ain Skhouna PV plant performance connected to the grid under a semi-arid climate in Algeria," Solar Energy, 232 (2022) 52-62.

[8] M. Baqir, H. K. Channi, "Analysis and design of solar PV system using Pvsyst software," Materials Today: Proceedings, 48 (2022) 1332-1338.

[9] K. R. Oyshei , K. M. Sazid Hasan, N. Sadat , Md. A. Hoque, "Comparative analysis of solar module configuration and tracking systems for enhanced energy generation in South Sakucia Union, Bhola, Bangladesh: A software based analysis," Heliyon, 10 (2024) e33884.

[10] R. Redwan, M. Hasan, A. Nadia, M. S.Khan, N. A. Chowdhury, N. R. Chowdhury, ''Design analysis and techno-economic assessment of a photovoltaic-fed electric vehicle charging station at Dhaka-Mawa expressway in Bangladesh', Energy Conversion and Management: X, 24 (2024) 100737

Extract Parameters of Photovoltaic System Using Metaheuristic Algorithm

1st Elmamoune Halassa
Electrical Engineering Department
Ziane Achour University
Djelfa, Algeria
halassa.elmamoune@univ-djelfa.dz

2nd Abdellatif Seghiour
second cycle Department
Ecole Supérieure en Génie Electrique et Énergétique
Oran, Algeria
seghiour_abdellatif@esgee-oran.dz

***Abstract*—Photovoltaic (PV) systems harness solar energy to produce electricity, offering a cleaner and more sustainable alternative to fossil fuel-based power generation. As concerns over environmental impacts and energy sustainability grow, PV systems are becoming increasingly important in global energy strategies. The performance of these systems is often illustrated through the current-voltage (I-V) characteristic curve, a representation influenced by several critical electrical parameters, such as series resistance, shunt resistance, and diode ideality factor. Proper identification of these parameters is essential, as they directly affect PV system efficiency, stability, and energy output. This paper focuses on the application of the Dandelion Optimization (DO) algorithm, a nature-inspired computational approach, for identifying unknown parameters within PV cells and modules. The DO algorithm seeks to minimize a predefined objective function by simulating current to achieve the most accurate representation of the PV system's behavior. Through iterative search processes, the algorithm identifies optimal parameter values that closely match experimental data. Results indicate that DO provides a reliable and effective means of parameter extraction, outperforming conventional methods in terms of accuracy and computational efficiency. This research demonstrates that DO can enhance the modeling and optimization of PV systems, contributing to improved performance and supporting the transition toward renewable energy.**

***Index Terms*—photovoltaic system, single diode model, parameter extraction, Dandelion optimizer**

I. Introduction

The growing demand for electricity from both domestic and industrial users, coupled with the limitations and environmental harm of conventional energy sources, it has necessitated the shift towards alternative, eco-friendly energy solutions. Renewable energy, particularly PV systems, offers a sustainable option due to its low cost, infinite availability, global accessibility, and clean energy output. Solar PV systems lead the renewable energy sector, with around 115 GW of new capacity added in 2019 [1]. With the rising adoption of PV systems, accurate modelling and parameter estimation are essential to analyse their static and dynamic behavior in power systems [2]. This study utilizes the Dandelion Optimizer to determine key parameters for PV cells and modules. Proper modelling and characterization not only help assess performance but also enhance cell design, optimize manufacturing processes, and support quality control [3] [4]. Simulations using accurate models are crucial for system evaluation. Among the various PV cell models such as single, double, and triple diode models the single diode model is widely used due to its balance of accuracy and simplicity [5].

II. Single Diode Model

The single diode model, illustrated in Figure 1, consists of a constant current source I_{ph}, which is proportional to solar radiation, connected in parallel with a diode. The diode incorporates an ideality factor to account for recombination within the space charge region. Series resistance Rs represents losses due to connections with other cells and modules, while parallel resistance R_{sh} models leakage currents caused by impurities and imperfections in the p-n junction. This model, known for its simplicity and accuracy, involves five unknown parameters for effective representation of PV modules [5] [6].

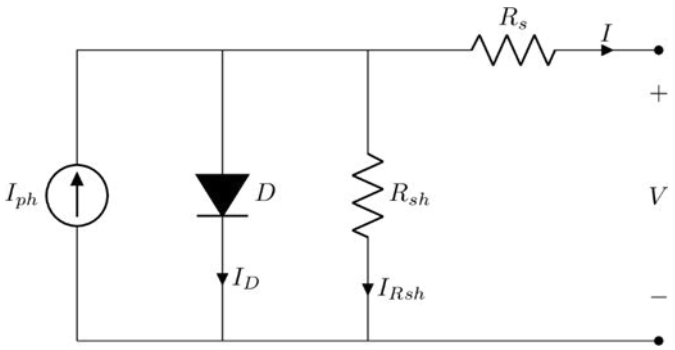


Fig. 1. Single Diode Model.

Equation 1 describes the output current of single diode model presented in Fig. 1

$$I = I_{ph} - I_0 \left[\exp \frac{q\,(V + R_s I)}{AkT} - 1\right] - \frac{V + R_s I}{R_{sh}} \qquad (1)$$

Where :

I : Output current (A)
I_{ph} : Photo-generated current (A)
I_0 : Dark saturation current (A)
R_s : Series resistance (Ω)
R_{sh} : Shunt resistance (Ω)
V : Output voltage (V)

q : Elementary charge ($1.60217662 \times 10^{-19} C$)
k : Boltzmann constant ($1.380649 \times 10^{-23} J/K$)
T : Ambient temperature (K)
A : Diode ideality factor

III. IDENTIFICATION'S METHODS

Many studies have addressed the challenge of accurately identifying the parameters of solar cells. Research in [7] shows that linear identification techniques alone are insufficient for solving this inherently nonlinear problem. Various nonlinear algorithms have been proposed, but each comes with trade-offs related to complexity, precision, speed, or convergence limitations [8]. These approaches can be categorized as follows:

A. Analytical Approaches

These rely on mathematical operations to address the non-linearity between parameters. They primarily focus on key points of the I-V curve, such as maximum power current, maximum power voltage, short circuit current, and open circuit voltage [9] [10] [11] [12] [13].

B. Numerical Approaches

Known as exact methods, these consider all measured data points from the I-V curve and use iterative algorithms [12] [14]. These methods require extensive computation and precise initial assumptions to ensure convergence during iterations [9].

C. Metaheuristic Approaches

Metaheuristic algorithms treat parameter extraction as an optimization problem by minimizing the difference between theoretical and experimental current values through a defined objective function [12] [13] [14] [15] [16]. Various algorithms, including Genetic Algorithm (GA) [17], Particle Swarm Optimization (PSO) [18], and Artificial Bee Colony Optimization (ABC) [19], have been explored for identifying the unknown parameters of PV cells.

IV. DANDELION OPTIMIZER OVERVIEW

The Dandelion Optimization (DO) algorithm, introduced by Shijie Zhao in 2022, is an optimization technique inspired by the flight of dandelion seeds carried by the wind. Dandelions disperse their seeds over long distances, with some traveling dozens of kilo meters if the wind conditions are favourable. The algorithm is based on three key stages of the seed's journey: Rising Stage, Descending Stage and Landing Stage [20].

V. DANDELION OPTIMIZER ALGORITHM MATHEMATICAL MODEL

A. Initialization phase

DO begins with the evolution of a population and iterative optimization based on initial population or seed generation. Multiple potential solutions, represented as X_i, are generated to form the population, which is structured as a matrix Every potential solution X_i is randomly generated between the upper bound (UB) and lower bound (LB) as follows:

$$X_i = rand \times (UB - LB) + LB \tag{2}$$

where $rand$ is a random number between 0 and 1.

B. Rising stage

During the rising stage, dandelion seeds must reach a certain height before they can float away from their parent. The heights to which the seeds rise are influenced by wind speed, air humidity, and other factors.

C. Descending Stage:

In this phase, the DO algorithm emphasizes exploration. Dandelion seeds' motion is modeled using Brownian motion after rising to a certain height. Brownian motion allows individuals to explore a wider range of search communities as it follows a normal distribution at each step. The average position information after the rising stage, denoted as D_{mean_t}, guides the population towards promising communities. The descent process is mathematically represented as:

$$D_{t+1} = D_t - \alpha \times \beta_t \times (D_{mean_t} - \alpha \times \beta_t \times D_t) \tag{3}$$

where β_t is a random number drawn from the normal distribution to denote Brownian motion.

D. Landing Stage:

During the exploitation phase, the dandelion seed selects its landing spot based on the results of the previous stages. The algorithm aims to reach the global optimal solution. This process is represented by:

$$D_{t+1} = D_{elite} + levy(\lambda) \times \alpha \times (D_{elite} - D_t \times \delta) \tag{4}$$

where D_{elite}" represents the dandelion seed's optimal position as shown in equation 5, and $levy(\lambda)$ is the Levy flight function calculated using Equation 6. The parameter σ is a linearly increasing function between [0, 2].

$$\begin{cases} D_{elite} = D(find(f_{best} = f(D_i))) \\ f_{best} = min(f(D_i)) \end{cases} \tag{5}$$

$$levy(\lambda) = s \times \frac{\omega \times \sigma}{|t|^{1/\beta}} \tag{6}$$

VI. DANDELION OPTIMIZER IMPLEMENTATION

The single diode model parameters were identified using the DO algorithm which was implemented MATLAB. The procedure of parameter extraction may be thought of as an optimization problem, and the vector's solution can be seen as follows :

$$\alpha = [I_{ph}, I_0, A, R_s, R_{sh}] \tag{7}$$

A. DO algorithm objective function

The root means square error (RMSE) between the computed current and the measured (experimental) current can be used to define the objective function. The following formula may be used to determine the RMSE value:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} c_i\left(I, V, \alpha\right)^2} \tag{8}$$

Where N is the size of the measured data (number of experimental I-V points) and c is defined as follows:

$$c\left(I, V, \alpha\right) = I - \left(I_{ph} - I_0\left[\exp\frac{q\left(V + R_s I\right)}{AkT} - 1\right] - \frac{V + R_s I}{R_{sh}}\right) \tag{9}$$

B. DO algorithm parameters

The parameters used are summarized in the table I

TABLE I
DO ALGORITHM PARAMETERS FOR ISOFOTON 106/12

Parameters	Values
Population size	50
Number of iterations	1500

C. Algorithm results

The extracted parameters of the single diode model are summarized in Table II. The convergence speed is illustrated in Figure 2.

TABLE II
DO ALGORITHM RESULTS FOR ISOFOTON 106/12

Parameters	Values
$I_{ph}\ (A)$	5.0675
$I_0\ (A)$	1.5078×10^{-5}
A	1.7259
$R_s\ (\Omega)$	0.0034
$R_{sh}\ (\Omega)$	3.6659
RMSE	0.0178

VII. CONCLUSION

This paper focuses on parameter identification using the Dandelion Optimization (DO) algorithm. The DO algorithm was applied to extract the parameters of the single diode model based on experimental data, ensuring accuracy and reliability. The results confirm that the DO algorithm is a powerful tool for precise parameter extraction, enabling accurate modeling and simulation of PV systems. This study offers valuable insights into optimizing system performance, contributing to the improved design and operation of solar energy systems.

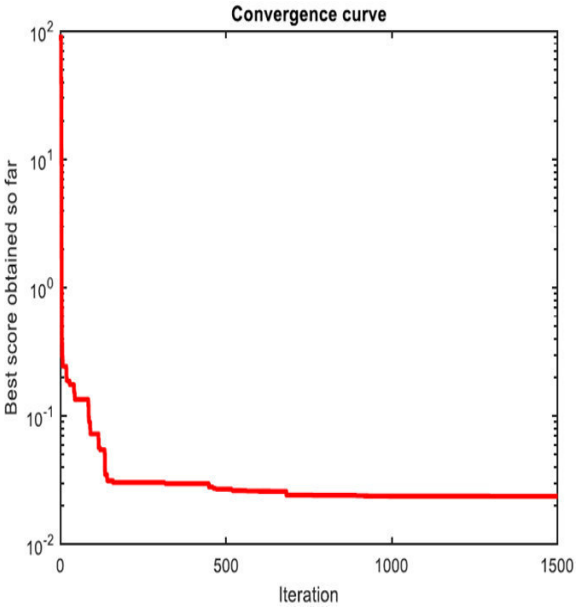


Fig. 2. DO algorithm convergence curve.

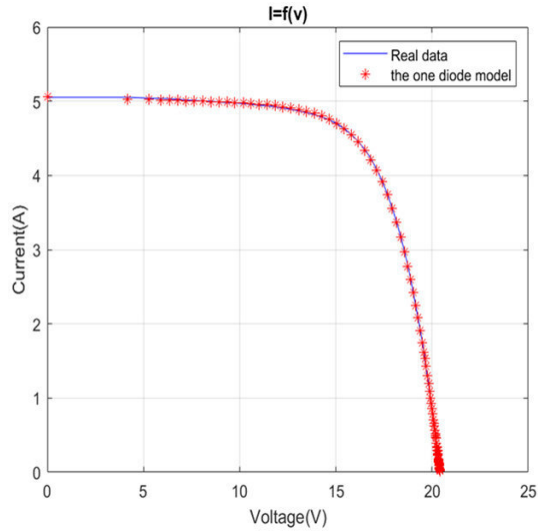


Fig. 3. Measured and calculated I-V curve.

REFERENCES

[1] Renewables 2020 Global Status Report. Available online: https://www.ren21.net/gsr-2020 (accessed on 1 August 2021).

[2] Yu, S., Zhang, L., Lu, H. H. C., Fernando, T., & Wong, K. P. (2017). A DSE-based power system frequency restoration strategy for PV-integrated power systems considering solar irradiance variations. IEEE Transactions on Industrial Informatics, 13(5), 2511-2518.

[3] Jordehi, A. R. (2016). Parameter estimation of solar photovoltaic (PV) cells: A review. Renewable and Sustainable Energy Reviews, 61, 354-371.

[4] Tamrakar, R., & Gupta, A. (2015). A Review: extraction of solar cell modelling parameters. International journal of innovative research in electrical, electronics, instrumentation and control engineering, 3(1), 55-60.

[5] Nayak, B., Mohapatra, A., & Mohanty, K. B. (2019). Parameter estimation of single diode PV module based on GWO algorithm. Renewable Energy Focus, 30, 1-12.

[6] Villalva, M. G., Gazoli, J. R., & Ruppert Filho, E. (2009). Comprehensive approach to modeling and simulation of photovoltaic arrays. IEEE Transactions on power electronics, 24(5), 1198-.
[7] Lim, L. H. I., Ye, Z., Ye, J., Yang, D., & Du, H. (2015). A linear method to extract diode model parameters of solar panels from a single I–V curve. Renewable Energy, 76, 135-142.
[8] Ortiz-Conde, A., Ma, Y., Thomson, J., Santos, E., Liou, J. J., Sánchez, F. G., ... & Layman, P. (1999). Direct extraction of semiconductor device parameters using lateral optimization method. Solid-State Electronics, 43(4), 845-848.
[9] Toledo, F. J., & Blanes, J. M. (2016). Analytical and quasi-explicit four arbitrary point method for extraction of solar cell single-diode model parameters. Renewable Energy, 92, 346-356.
[10] Pindado, S., & Cubas, J. (2017). Simple mathematical approach to solar cell/panel behavior based on datasheet information. Renewable energy, 103, 729-738.
[11] Cubas, J., Pindado, S., & Victoria, M. (2014). On the analytical approach for modeling photovoltaic systems behavior. Journal of power sources, 247, 467-474.
[12] Abbassi, R., Abbassi, A., Jemli, M., & Chebbi, S. (2018). Identification of unknown parameters of solar cell models: A comprehensive overview of available approaches. Renewable and Sustainable Energy Reviews, 90, 453-474.
[13] Merchaoui, M., Sakly, A., & Mimouni, M. F. (2018). Particle swarm optimisation with adaptive mutation strategy for photovoltaic solar cell/module parameter extraction. Energy Conversion and Management, 175, 151-163.
[14] Ayodele, T. R., Ogunjuyigbe, A. S. O., & Ekoh, E. E. (2016). Evaluation of numerical algorithms used in extracting the parameters of a single-diode photovoltaic model. Sustainable Energy Technologies and Assessments, 13, 51-59.
[15] Easwarakhanthan, T., Bottin, J., Bouhouch, I., & Boutrit, C. (1986). Nonlinear minimization algorithm for determining the solar cell parameters with microcomputers. International journal of solar energy, 4(1), 1-12.
[16] Ferhat-Hamida, A., Ouennoughi, Z., Hoffmann, A., & Weiss, R. (2002). Extraction of Schottky diode parameters including parallel conductance using a vertical optimization method. Solid-State Electronics, 46(5), 615-619.
[17] Jervase, J. A., Bourdoucen, H., & Al-Lawati, A. (2001). Solar cell parameter extraction using genetic algorithms. Measurement science and technology, 12(11), 1922.
[18] Ye, M., Wang, X., & Xu, Y. (2009). Parameter extraction of solar cells using particle swarm optimization. Journal of Applied Physics, 105(9), 094502.
[19] Oliva, D., Cuevas, E., & Pajares, G. (2014). Parameter identification of solar cells using artificial bee colony optimization. Energy, 72, 93-102.
[20] Halassa, E., Mazouz, L., Seghiour, A., Chouder, A., & Silvestre, S. (2023). Revolutionizing photovoltaic systems: an innovative approach to maximum power point tracking using enhanced dandelion optimizer in partial shading conditions. Energies, 16(9), 3617.

Modeling and Analysis of Shading Effects on Solar Photo-Voltaic Systems

1st Med Mounir Abbes
LEER Laboratory Department of Electrical Engineering
University of Souk-Ahras
Souk-Ahras, Algeria
m.abbes@unvi-soukahras.dz

2ndZouhir Boumous
LEER Laboratory, Department of Electrical Engineering
University of Souk-Ahras
Souk-Ahras, Algeria
zohir.boumous@univ-soukahras.dz

3rd Samira Boumous
LEER Laboratory, Department of Electrical Engineering
University of Souk-Ahras
Souk-Ahras, Algeria
samira.boumous@univ-soukahras.dz

4th Rayane Yasmine Mahfoudi
LEER Laboratory, Department of Electrical Engineering
University of Souk-Ahras
Souk-Ahras, Algeria
r.mahfoudi@univ-soukahras.dz

***Abstract*—This paper investigates the effects of partial shading on the performance of solar photovoltaic (PV) plants and modules, focusing on the resulting power losses and temperature variations. A simulation model of a solar plant, developed using Matlab Simulink , is used to analyze how non-uniform irradiance across PV cells and modules impacts overall efficiency. The study considers different shading scenarios to observe changes in power output and PV cell junction temperatures. To address these shading effects, the model includes bypass and blocking diodes within the Solar Plant block, which help maintain power levels and prevent overheating by controlling current flow through shaded areas. Users can adjust irradiance and temperature settings to replicate real-world shading conditions, enabling a comprehensive evaluation of shading mitigation strategies. The insights gained from this work offer guidance on optimizing PV systems for improved resilience and performance under partial shading, contributing to enhanced efficiency and reliability in solar power installations.**

***Index Terms*—Partial Shading , photovoltaic plants, diode, efficiency.**

I. Introduction

The study of shading effects on photovoltaic (PV) systems has evolved as solar energy technologies have become more widespread and sophisticated [1] . In early PV system research, shading was identified as a key factor impacting solar energy output, especially in environments where clouds, trees, buildings, or other obstacles caused non-uniform irradiance. Initial studies primarily observed shading effects on small-scale PV modules, noting that even minor shading could lead to significant power losses, as shaded cells act as resistive loads within the circuit [2-3]. With advancements in semiconductor and PV technology, researchers began examining ways to mitigate shading effects. By the late 20th century, the introduction of bypass diodes in PV modules became a standard approach to reduce power losses by allowing current to bypass shaded cells, preventing them from creating hotspots that could damage the panel. However, bypass diodes could only partially address shading challenges, particularly in large interconnected systems where non-uniform irradiance caused more complex issues [4-6] . The early 2000s saw a surge in PV system simulation and modeling capabilities, facilitated by computational tools like MATLAB/Simulink, which allowed researchers to simulate shading impacts on large PV arrays and analyze power losses under varied environmental conditions. With the integration of Simscape and similar modeling languages, researchers could design and test sophisticated PV systems with realistic shading scenarios, leading to new insights into shading-induced energy losses, temperature effects, and current imbalances [7].

Recent studies have increasingly focused on optimizing PV arrays under partial shading, using techniques like maximum power point tracking (MPPT) algorithms, microinverters, and advanced diodes to enhance efficiency. Today, the study of shading impacts on PV systems is central to the design of resilient and efficient solar installations, with ongoing research dedicated to creating models that accurately replicate real-world shading conditions and assess protective technologies to optimize energy yield [8-10]. This paper builds on these advancements, using Simulink to model shading effects and assess the role of bypass and blocking diodes in large PV systems under partial shading conditions. PV Systems Modelisation and Shading Effects: This section details the modeling approach for analyzing shading effects in photovoltaic (PV) systems. The model is designed to capture the impact of shading on PV cells and modules within interconnected systems, allowing for a detailed study of power output variations, temperature changes, and overall efficiency in response to non-uniform irradiance. Figure (1) present a general PV cell model using single diode. estimation.

Identify applicable funding agency here. If none, delete this.

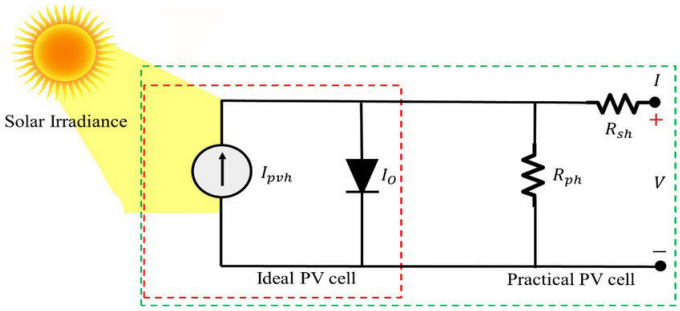


Fig. 1. PV cell model

II. Shading Mechanism and Parameter Settings

Shading in a PV system arises when some cells or modules receive less sunlight than others, often due to shadows from clouds, buildings, or trees. This uneven irradiance can create hotspots in shaded cells, causing them to operate as resistors and reduce the current flow through the system. To model this effect, shading parameters are defined by setting irradiance and temperature values for specific cells or modules. Adjusting these parameters allows the simulation of partial, full, or dynamic shading, offering insight into the performance variations under different conditions.

III. Bypass and Blocking Diodes

The model incorporates bypass and blocking diodes to mitigate shading effects. Bypass diodes are placed across individual PV cells or groups of cells to allow current to ”bypass” shaded cells. When a cell is shaded, the bypass diode enables the current to reroute around it, reducing energy losses and preventing the formation of hotspots. Blocking diodes are used to prevent reverse current flow, protecting the modules from potential damage when irradiance levels vary between modules. These diodes play a crucial role in maintaining stable power output and minimizing thermal stress on shaded cells.

IV. Mathematical Formulation and Simulation Equations

The electrical characteristics of PV cells, including current-voltage (I-V) and power-voltage (P-V) relationships, are modeled based on the single-diode equivalent circuit. It represents the PV cell as an equivalent circuit consisting of a current source, a diode, a series resistor (Rs), and a parallel resistor (Rp). The current-voltage (I-V) characteristic equation for the single-diode model is shown in equation 1:

$$I = I_{\text{ph}} - I_0 \left(\exp\left(\frac{V + IR_s}{nV_t} \right) - 1 \right) - \frac{V + IR_s}{R_p} \qquad (1)$$

Where: I is the output current (A).
V is the output voltage (V).
Iph is the photocurrent (A), proportional to irradiance.
I0 is the saturation current of the diode (A).
n is the ideality factor of the diode.
Vt=VTq=qkT is the thermal voltage (V),
where K is the Boltzmann constant,
T is the temperature (K), and q is the charge of an electron.
Rs is the series resistance (Ω).
Rp is the parallel (shunt) resistance (Ω).

V. Results and Discussion

The study Solar Plant block comprises 4 parallel-connected strings. Each string comprises 5 series-connected solar PV modules. Case 1: No protection diode. This subsection examines the impact of shading on a PV system in the absence of protection diodes, providing insight into how shading affects power output, current flow, and cell temperature when bypass and blocking diodes are not present. The graph presented

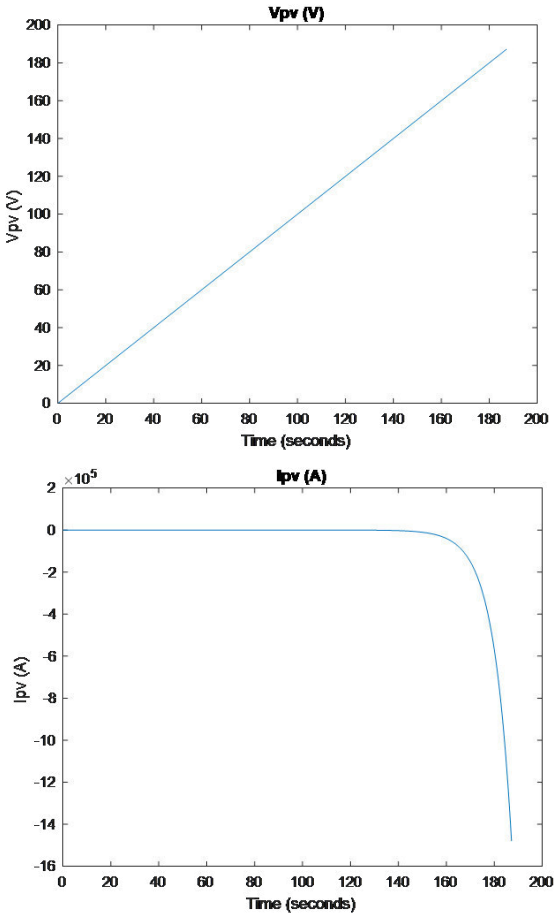


Fig. 2. The characteristic curve of the photovoltaic panel

in figure (2) shows the I-V and P-V characteristics of a solar photovoltaic module at 20°C and 45°C, illustrating how temperature impacts performance. At higher temperatures, the open-circuit voltage and maximum power point (MPP) decrease, resulting in reduced efficiency. Without protection diodes, such as bypass or blocking diodes, the system is more susceptible to significant power losses, especially under partial shading or non-uniform conditions, as there is no mechanism to bypass shaded cells or prevent reverse current flow. This can lead to increased risks of hotspots and potential damage, compromising overall system reliability and efficiency. Case

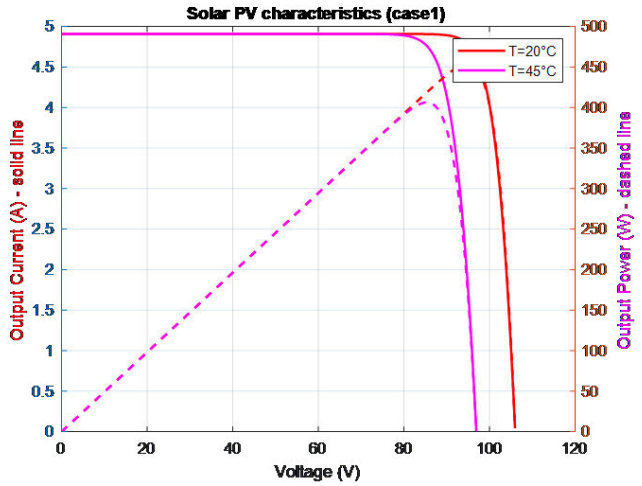


Fig. 3. System output in case 1.

2 : With only the Baypass diode Figure (3) shows the I-

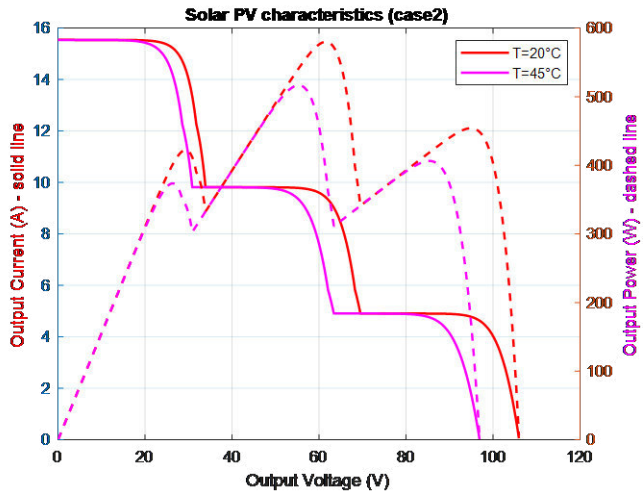


Fig. 4. System output in case 2.

V (current-voltage) and P-V (power-voltage) characteristics of a solar photovoltaic module at temperatures of 20°C and 45°C, with the inclusion of bypass diodes, The current drops in distinct steps rather than sharply, indicating that portions of the module are bypassed when shading or non-uniform irradiance occurs. This behavior is a result of the bypass diodes redirecting current around shaded or underperforming sections. At both 20°C and 45°C, the output current decreases more gradually compared to the curve without diodes, reflecting partial bypassing of shaded sections. The power output also drops in steps, corresponding to the activation of bypass diodes as different parts of the array experience shading or lower performance. At 45°C, the overall power output is still lower than at 20°C, showing the temperature effect, but the bypass diodes help to maintain some power output even when sections are shaded.

Case 3 : with both bypass and blocking diode Figure (4)

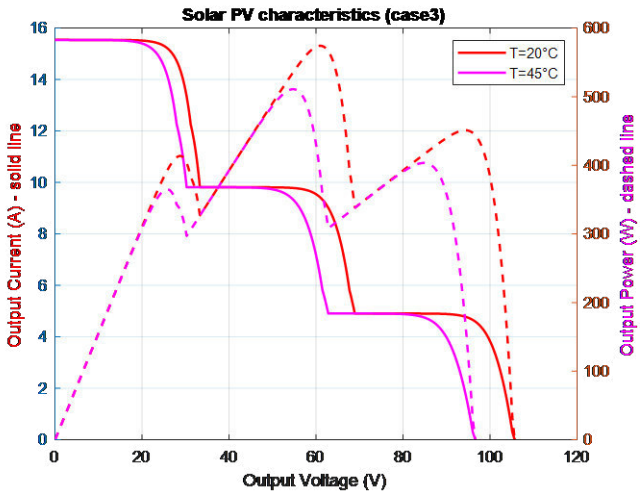


Fig. 5. System output in case 3.

displays the I-V and P-V characteristics of a solar photovoltaic module at 20°C and 45°C with bypass diodes included. At 20°C, the current initially holds steady near 16 A before experiencing drops at around 35 V and 75 V, where the bypass diodes activate to bypass shaded sections, maintaining partial output. The power curve peaks approximately 500 W, followed by distinct drops as shaded sections are bypassed. At 45°C, the maximum current drops to about 14 A, with the voltage and power curves peaking lower at approximately 450 W, indicating temperature-induced efficiency losses. The bypass diodes still create steps in the output, helping maintain some power despite partial shading, but the overall system performance decreases as temperature rises.

VI. Conclusion

This study illustrates the performance of a solar photovoltaic (PV) system under different conditions and the impact of temperature and protection diodes. Without Protection Diodes: The first graph showed a rapid drop in current and power as voltage increased, indicating that the system was highly susceptible to power losses when exposed to partial shading or temperature effects. At 20°C, the maximum power was around 450 W, while at 45°C, the power output reduced significantly, showing that temperature adversely affects performance without protective measures. With Only Bypass Diodes: The second graph demonstrated that adding bypass diodes mitigated the effects of partial shading. The current and power curves featured stepped drops, indicating that the diodes effectively isolated shaded sections and maintained current flow through unshaded parts. At 20°C, the peak power output was approximately 500 W, while at 45°C, it was reduced to about 450 W due to temperature effects. This showed that while bypass diodes improved the system's response to shading, temperature still had a notable impact. Detailed Performance with Bypass Diodes: The third graph provided a more comprehensive look at the PV system with bypass

diodes at both temperatures. At 20°C, the current started at 16 A and dropped in stages around 35 V and 75 V, with a maximum power output of about 500 W. At 45°C, the current peaked at 14 A, and power output was lower at around 450 W. The stepped drops highlighted the activation of bypass diodes, maintaining partial performance under shading but showing clear temperature-induced efficiency losses.

Effectiveness of Bypass Diodes: The bypass diodes help maintain the flow of current by creating alternative paths around shaded sections, thereby preventing the entire string from being impacted by the drop in current. This reduces the impact of partial shading on overall system power output and prevents damage due to overheating or hotspots. Partial Shading Management: The stepped nature of the curves indicates that bypass diodes effectively isolate shaded sections, allowing the unshaded parts to continue contributing power. Temperature Dependence: Even with bypass diodes, higher temperatures (e.g., 45°C) still result in reduced power and voltage compared to cooler conditions (20°C), demonstrating that temperature management is crucial for optimal performance. Finaly, it is evident that temperature significantly affects PV system performance, reducing peak power output from 500 W at 20°C to about 450 W at 45°C. The use of bypass diodes helps manage partial shading by maintaining power output in unshaded sections, but it does not counteract the negative effects of high temperatures. Both bypass diodes and effective temperature management are essential for optimal system design.

References

[1] Sun, C., Zou, Y., Qin, C. et al. Temperature effect of photovoltaic cells: a review. Adv Compos Hybrid Mater 5, 2675–2699 (2022). https://doi.org/10.1007/s42114-022-00533-z.

[2] Tang H, Zhou Z, Jiao S, Zhang Y, Li S, Zhang D, Zhang J, Liu J, Zhao D (2022) Radiative cooling of solar cells with scalable and high-performance nanoporous anodic aluminum oxide. Sol Energy Mater Sol Cells 235:111498.

[3] Elminshawy NAS, Mohamed AMI, Osama A, Amin I, Bassam AM, Oterkus E (2022) Performance and potential of a novel floating photovoltaic system in Egyptian winter climate on calm water surface. Int J Hydrog Energy 47(25):12798–12814.

[4] Zhang J, Zhai H, Wu ZH, Wang YY, Xie HQ (2020) Experimental investigation of novel integrated photovoltaic-thermoelectric hybrid devices with enhanced performance. Sol Energy Mater Sol Cells 215:110666.

[5] Shaker, L.M., Al-Amiery, A.A., Hanoon, M.M. et al. Examining the influence of thermal effects on solar cells: a comprehensive review. Sustainable Energy res. 11, 6 (2024). https://doi.org/10.1186/s40807-024-00100-8.

[6] Zhou, J., Yi, Q., Wang, Y., & Ye, Z. (2015). Temperature distribution of photovoltaic module based on finite element simulation. Solar Energy, 111, 97–103.

[7] Silva M, Roberts JJ, Prado PO. Calculation of the Shading Factors for Solar Modules with MATLAB. Energies. 2021; 14(15):4713. https://doi.org/10.3390/en14154713.

[8] Numan, A.H.; Dawood, Z.S.; Hussein, H.A. Theoretical and experimental analysis of photovoltaic module characteristics under different partial shading conditions. Int. J. Power Electron. Drive Syst. 2020, 11, 1508–1518.

[9] Trejos Grisales, L.A.; Ramos Paja, C.A.; Saavedra Montes, A.J. Techniques for modeling photovoltaic systems under partial shading. Tecnura 2016, 20, 171–183.

[10] Boubaker, O. MPPT techniques for photovoltaic systems: a systematic review in current trends and recent advances in artificial intelligence. Discov Energy 3, 9 (2023). https://doi.org/10.1007/s43937-023-00024-2.

Wind Energy Conversion System IGBT Open Circuit Fault Compensation

Mokhtar Mahmoud Mohammedi
Laboratory of Electrical Drives Development (LDEE) of Electrical Department
University of Sciences & Technology of Oran Mohamed Boudiaf
Oran, Algeria
mokhtar.mohammedi@yahoo.fr

Hocine Guentri
Department of Mechanic and Electromechanics, Institute of Science and Technologies, Abdelhafid Boussouf University centre of Mila,
Mila, Algeria
hguentri2005@centre-univ-mila.dz

Noureddine Messaoudi
LIST Laboratory of Engineering of Electrical Systems Department
University M'Hamed Bougara
Boumerdes, Algeria
n.messaoudi@univ-boumerdes.dz

Zoubir Gadouche
Laboratory of Electrical and Computer Engineering (L2GEGI) of Electrical Department
University Ibn Khaldun of Tiaret
Tiaret, Algeria
gadouche.zoubir@gmail.com

Benabied Djamal
Lab of Research on Electromechanical and Dependability of Mechanical Engineering Department
University of Souk Ahras
Souk Ahras, Algeria
bendjamel2018@gmail.com

Samah Bouanik
Lab of Electromechanical Systems of Sciences Engineering Faculty
University of 20 Aout 1955, innovation academy
Skikda, Algeria
s.bouanik@univ-skikda.dz

***Abstract*— the work presented in this paper is about Doubly Fed Induction Generator (DFIG) Wind Energy Conversion System (WECS) under Rotor Side Converter (RSC) IGBT open circuit fault compensation proposed technique. First a classical indirect PI controller based on the pole compensation technique is presented and considered as a reference well known approach, then a new IGBT open circuit fault compensation technique is associated to our DFIG-WECS in the aim of guaranteeing the electrical power generation under the fault occurrence. The results obtained from the proposed fault compensation technique under MATLAB-Simulink gives value to it through the faulty condition.**

Keywords*— *fault compensation technique, IGBT open circuit fault, DFIG, WECS, MPPT, PI controller.

I. INTRODUCTION

Unexpected IGBT open circuit fault appearance in the Rotor Side Converter (RSC) of the Doubly Fed Induction Generator (DFIG) Wind Energy Conversion System (WECS) can lead to electrical power generation disturbance [1]. Hence, to remove the failure effect, fault compensation techniques [2], [3] have been implemented especially in sensitive applications such as WECS and other renewable energetic systems processes [4], [5], where the power devices inverters includes fault compensation topologies. The main used topologies are the neutral four leg redundant topology and the four leg phase redundant topology [6]. Among the constraints of these topologies is the use of additional fast acting fuses or semiconductor switches to isolate the faulty leg, hence resulting in the increase of the losses in the inverter [6], [7].

In this paper, first a classical indirect PI controller based on the pole compensation technique is presented and considered as a reference well known approach. Then the proposed fault compensation technique is used, this does not take into consideration the fault isolation step from the fault compensation techniques process, where the proposed IGBT open circuit fault compensation scheme is based on a hardware redundancy and corresponding Fault Tolerant Control (FTC), which delivers evaluated and controlled voltages to the faulty inverter voltages in the IGBT open circuit fault occurrence. The main contributions of this work lies in the reduced inverter losses and the minimal fault compensation computational time. This is due to the enhanced implementation of the proposed fault compensation technique compared to the conventional ones.

II. DFIG MODELLING AND WECS CONTROL

The DFIG state model is:

$$\begin{cases} \dot{\Phi}_{sd} = -\frac{R_s}{L_s}\Phi_{sd} + \omega_s \Phi_{sq} + \frac{R_s M_{sr}}{L_s} i_{rd} + V_{sd} \\ \dot{\Phi}_{sq} = -\omega_s \Phi_{sd} - \frac{R_s}{L_s}\Phi_{sq} + \frac{R_s M_{sr}}{L_s} i_{rq} + V_{sq} \\ \dot{i}_{rd} = \frac{R_s M_{sr}}{\sigma L_r L_s^2}\Phi_{sd} - \frac{M_{sr}}{\sigma L_r L_s}\omega\Phi_{sq} - \left[\frac{R_r}{\sigma L_r} + \frac{R_s M_{sr}^2}{\sigma L_r L_s^2}\right] i_{rd} + \\ +(\omega_s - \omega) i_{rq} - \frac{M_{sr}}{\sigma L_r L_s} V_{sd} + \frac{1}{\sigma L_r} V_{rd} \\ \dot{i}_{rq} = \frac{M_{sr}}{\sigma L_r L_s}\omega\Phi_{sd} + \frac{R_s M_{sr}}{\sigma L_r L_s^2}\Phi_{sq} - (\omega_s - \omega) i_{rd} + \\ -\left[\frac{R_r}{\sigma L_r} + \frac{R_s M_{sr}^2}{\sigma L_r L_s^2}\right] i_{rq} - \frac{M_{sr}}{\sigma L_r L_s} V_{sq} + \frac{1}{\sigma L_r} V_{rq} \\ \dot{\omega} = \frac{P^2 M_{sr}}{L_s J}(\Phi_{sq} i_{rd} - \Phi_{sd} i_{rq}) - \frac{P}{J} Cr - \frac{f}{J}\omega \end{cases} \quad (1)$$

Where: σ–Blondel dispersion coefficient given as:

$$\sigma = 1 - \left(M_{sr}^2 / L_r . L_s\right) \quad (2)$$

Whith: R_s, R_r -stator and rotor resistances; L_s, L_r -stator and rotor inductances; M_{sr} -mutual inductance between stator and rotor; ω_s, ω_r -stator and rotor angular speed; σ-Blondel dispersion coefficient.

"Fig. 1" depicts the RSC IGBT5 open circuit proposed fault compensation technique block diagram associated to the RSC PI control and to the Maximum Power Point Tracking (MPPT) control.

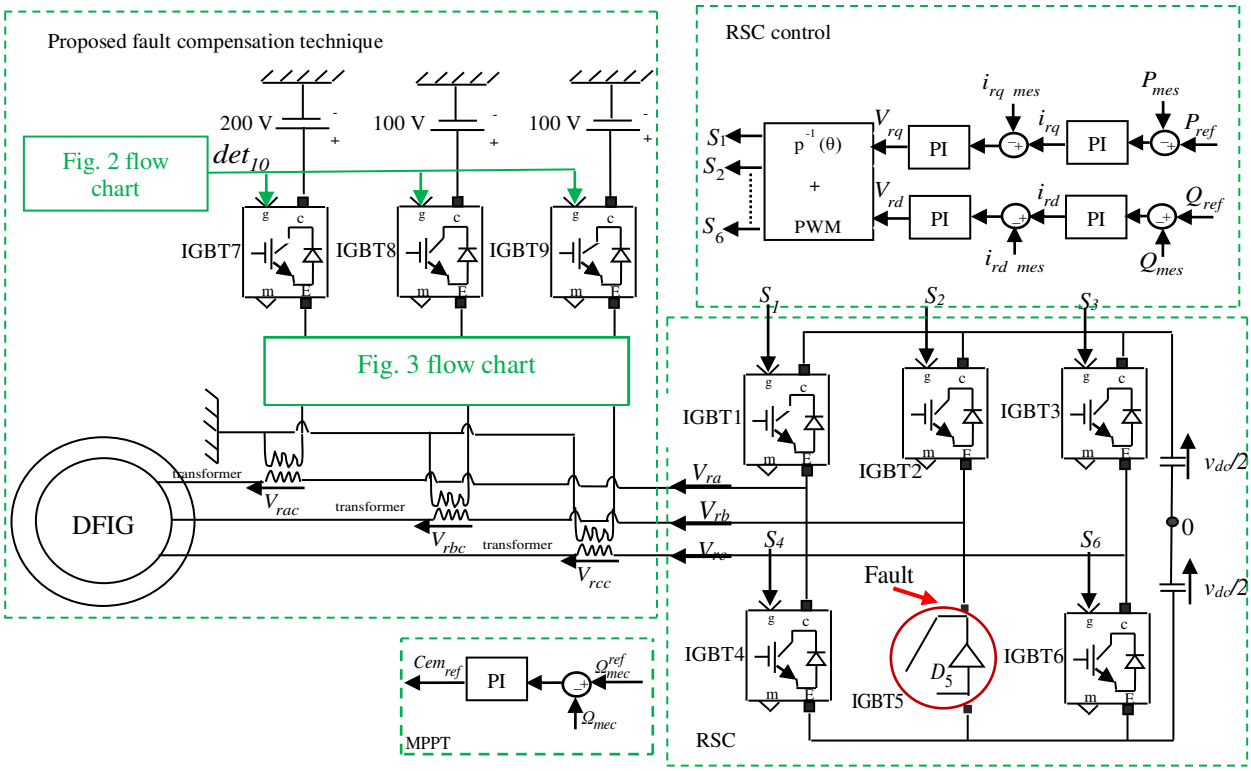


Fig. 1. IGBT5 open circuit proposed fault compensation technique block diagram associated to the PI and MPPT control techniques.

II.2 Converters control

Both of GSC and RSC are controlled by Pulse width modulation (PWM) technique. The GSC is related to the grid by the Inductor–Resistor (L_f, R_f) filter and to the RSC by the DC bus. The DC bus voltage is controlled by a PI controller at a constant value of V_{dc}=600V.

We facilitate the stator active and reactive powers decoupling by the alignment of the stator flux on (*d*) axis of (*d, q*) frame. That's leads us to equation (3) RSC vector control voltages [1]:

$$\begin{cases} V_{rd} = R_r i_{rd} + \left(L_r - \frac{M_{sr}^2}{L_s}\right)\dot{i}_{rd} - g\omega_s\left(L_r - \frac{M_{sr}^2}{L_s}\right)i_{rq} \\ V_{rq} = R_r i_{rq} + \left(L_r - \frac{M_{sr}^2}{L_s}\right)\dot{i}_{rq} + g\omega_s\left(L_r - \frac{M_{sr}^2}{L_s}\right)i_{rd} + g\frac{M_{sr}V_s}{L_s} \end{cases} \quad (3)$$

II.3 DFIG-WECS PI regulator gains determination:

The PI regulator gains k_p, k_i represented by Table I are determined by the pole compensation technique:

TABLE I. DFIG-WECS PI regulator gains

Control	*kp*	*ki*
MPPT	$k_{P\Omega mec} = \frac{J}{\tau_1}$	$k_{i\Omega mec} = \frac{F}{\tau_1}$
Rotor currents i_{rd}, i_{rq}	$k_{pc} = \frac{L_r.\sigma}{\tau_2}$	$k_{ic} = \frac{R_r}{\tau_2}$
Stator powers *P, Q*	$k_{pp} = \frac{L_r.\sigma}{\tau_3.k_{pc}}$	$k_{ip} = \frac{1}{\tau_3}$
Filter currents i_{fd}, i_{fq}	$k_{pf} = \frac{L_f}{\tau_4}$	$k_{if} = \frac{R_f}{\tau_4}$
DC bus voltage V_{dc}	$k_{pv} = 2.\xi.C.w_0$	$k_{iv} = C.w_0^2$

Where: *F*–overall friction coefficient; w_0 –Cutoff frequency; ξ –Damping factor and *J*–generator rotor overall inertia.

III. PROPOSED FAULT COMPENSATION TECHNIQUE

The signal resulting from equation (4) of the IGBT open circuit fault detection [8] is used in the proposed IGBT open circuit fault tolerant control (FTC) of the proposed fault compensation technique [2], [9]:

$$\varepsilon_{k0} = V_{k0es} - V_{k0m} \quad (4)$$

V_{k0m} and V_{k0es} are the measured and the estimated pole voltages with k (leg number) = {1, 2, 3}.

The identification of the fault compensation technique voltages is based on equation (5) [2], [9]:

$$V_{\text{compensation}} = V_{\text{healthy}} - V_{\text{faulty}} \quad (5)$$

Therefore, filling Table II through equation (5), based on the healthy case voltages model represented by equation (6), in terms of the switching functions, substituted from each voltages model of the open circuit fault case of the same equation (6), gives us the compensation voltages of the proposed fault compensation technique [2], [9]. The compensation voltages are mounted in series with the inverter output voltages [2], [9].

$$\begin{cases} V_{ra} = \frac{V_{dc}}{3}(2.S_1 - S_2 - S_3) \\ V_{rb} = \frac{V_{dc}}{3}(-S_1 + 2.S_2 - S_3) \\ V_{rc} = \frac{V_{dc}}{3}(-S_1 - S_2 + 2.S_3) \end{cases} \quad (6)$$

In real case, table II can be realized in a hardware redundancy as shown in "Fig. 1" block diagram. This hardware redundancy is composed of three IGBTs and three DC voltages sources; one of them with a value of 200 V and

the two other ones with values of 100 V each. These values are deduced from table II and therefore, six voltages can be obtained as a result (the positive and its reverse negative voltage), these voltages cover all table II fault compensation technique voltages cases.

The IGBT open circuit fault: FTC and voltages right combination choice are respectively illustrated by "Fig. 2" and "Fig. 3" flow charts.

TABLE II. IGBTs OPEN CIRCUIT FAULTS COMPENSATION VOLTAGES [2]

Compensation voltages	IGBT1	IGBT2	IGBT3	IGBT4	IGBT5	IGBT6
V_{rac}	$2.S_1.\frac{V_{dc}}{6}$	$-S_2.\frac{V_{dc}}{6}$	$-S_3.\frac{V_{dc}}{6}$	$-2.S_4.\frac{V_{dc}}{6}$	$S_5.\frac{V_{dc}}{6}$	$S_6.\frac{V_{dc}}{6}$
V_{rbc}	$-S_1.\frac{V_{dc}}{6}$	$2.S_2.\frac{V_{dc}}{6}$	$-S_3.\frac{V_{dc}}{6}$	$S_4.\frac{V_{dc}}{6}$	$-2.S_5.\frac{V_{dc}}{6}$	$S_6.\frac{V_{dc}}{6}$
V_{rcc}	$-S_1.\frac{V_{dc}}{6}$	$-S_2.\frac{V_{dc}}{6}$	$2.S_3.\frac{V_{dc}}{6}$	$S_4.\frac{V_{dc}}{6}$	$S_5.\frac{V_{dc}}{6}$	$-2.S_6.\frac{V_{dc}}{6}$

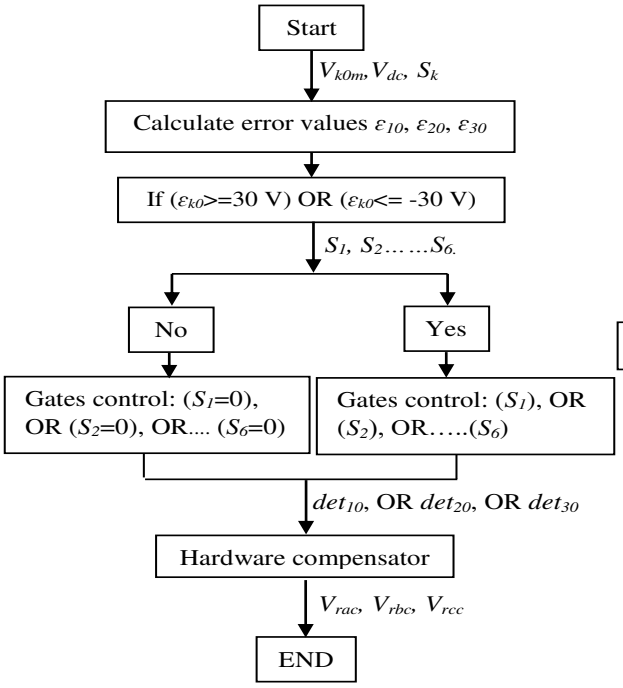


Fig. 2. IGBT open circuit FTC.

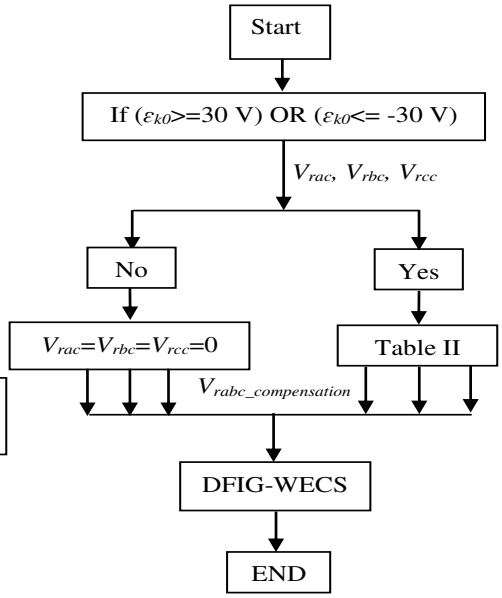


Fig. 3. Fault compensation technique voltages combination choice.

The compensation voltages are mounted in series with the inverter voltages as shown in "Fig. 1" block diagram.

IV. DFIG-WECS Simulation results:

The DFIG (7.5 KW) is linked to the wind turbine (10 KW) associated by its rotor the converters (2KHZ) and by its stator to the grid (50 HZ, 220/380V). β=0 is the angle of orientation of the blades. In this section, all the simulation results are presented.

IV.1.1 IGBT5 open-circuit fault

IV.1 proposed fault compensation technique Simulation results

In this case the wind speed profile reference is constant. At time t=1s we apply the open circuit fault in the RSC IGBT5 switch, where the faulty legs correspond directly to the current *irb* and to the voltage *Vrb*.

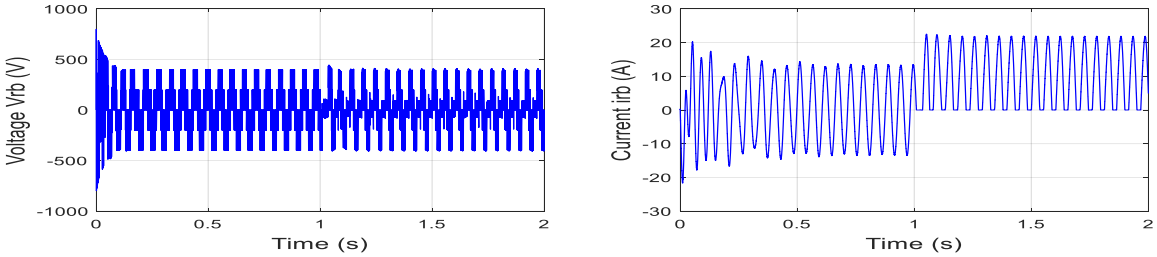


Fig. 4. *Vrb* voltage, *irb* current -under fault without fault compensation technique-.

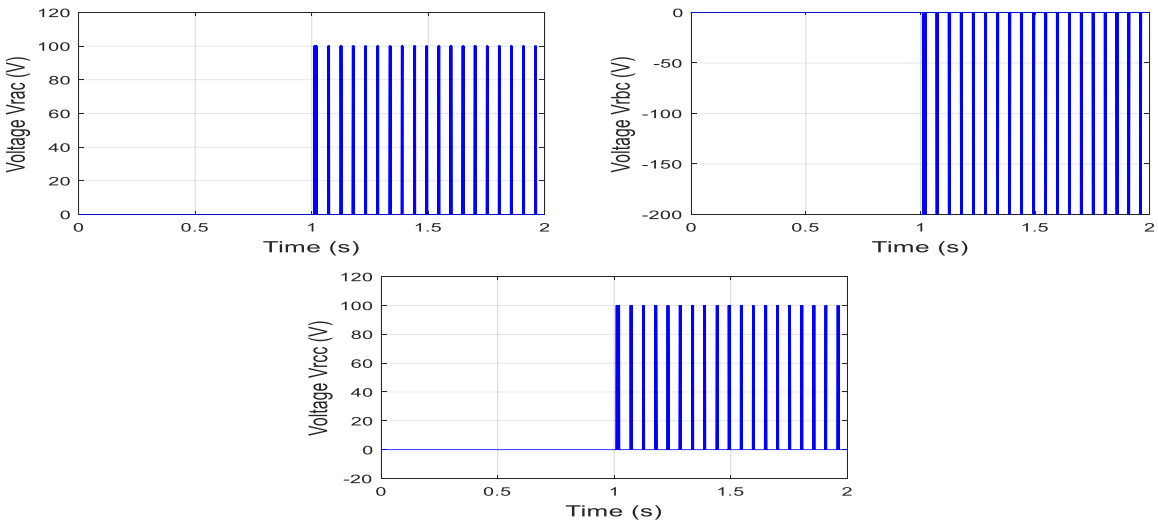


Fig. 5. Fault compensation technique voltages *Vrac, Vrbc, Vrcc.*

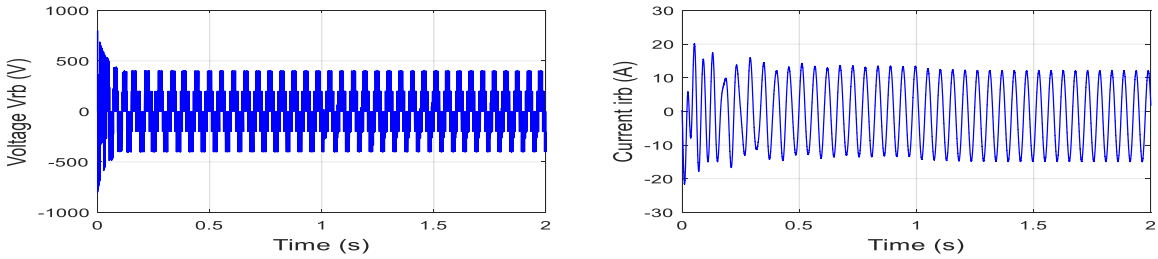


Fig. 6. Compensated: *Vrb* voltage, *irb* current.

"Fig. 4" shows the RSC IGBT5 open circuit fault performance without fault compensation technique of the *irb* current and *Vrb* voltage.

"Fig. 5" represent the IGBT5 open-circuit proposed fault compensation technique voltages.

"Fig. 6" depict the *irb* current and the *Vrb* voltage comeback to the healthy case state in the proposed fault compensation technique presence.

V. CONCLUSION

The paper proposes the use of the PI controller and the proposed fault compensation technique applied to the DFIG RSC under IGBT open-switch fault. The PI controller regulator gains are determined by the pole compensation technique considered as a reference well known approach. It is found throught simulation results that the PI controller shows a good performance and that the proposed fault compensation technique compared to the faulty leg isolation process from the four leg redundant topologies does not need to isolate the faulty leg resulting in minimal of inverter losses and of fault compensation computational time. The proposed fault compensation technique voltages are obtained from separated three DC voltages sources and three IGBTs hardware redundancy. The fault compensation technique: IGBTs FTC and voltages right combination choice, are based on algorithms. The obtained simulation results confirm the efficiency of the proposed fault compensation technique where the electrical power generation is successfully guaranteed.

[1] M. M. Mohammedi, A. Bendiabdellah, T. Allaoui, B. D. E. Cherif, "Wind energy conversion system control robustness based on current analysis of IGBT open-circuit fault", Przegląd Elektrotechniczny, Vol. 10, pp. 14–24, October 2020.

[2] M. M. Mohammedi, A. Bendiabdellah, T. Allaoui, "Voltage reconfiguration fault-tolerant technique for two-level three-phase inverter integrated in a wind energy system", International Journal of Modelling, Identification and Control , Vol. 40, pp. 315–326, September 2022.

[3] Z. Jian-Jian, C. Yong, C. Zhang-Yong, Z. Anjian, "Open-Switch Fault Diagnosis Method in Voltage-Source Inverters Based on Phase Currents", IEEE access Vol 7, pp. 63619–63625, April 2019.

[4] H. Chunxiao, L. Xisheng, "Linear discrete-time modelling and hybrid fault-tolerant controller design for complex systems with component and sensor faults", International Journal of Modelling, Identification and Control, Vol. 34, No. 4, pp.342–349, january 2021.

[5] Z. Gadouche, C. Belfedal, T. Allaoui, M. Denai, M. Bey, "Hybrid Renewable Energy System Controlled with Intelligent Direct Power Control", Journal Européen des Systèmes Automatisés, Vol. 55, pp. 467-475, August 2022.

[6] B. Tabbache, M. Benbouzid, A. Kheloui, J. Bourgeot, A. Mamoune, "An improved fault-tolerant control scheme for PWM inverter-fed induction motor-based EVs", ISA Trans, vol.52, pp.862-869, November 2013.

[7] B. H. Mahdhi, A. Djerioui, H. Ben Azza, M. Jemli, "Experimental investigation of an open-switch fault diagnosis approach in the IGBT-based power converter connected to permanent magnet synchronous generator-DC system", Int Trans Electr Energ Syst. e12436, April 2020.

[8] S. Karimi, P. Poure, S. Saadate, "FPGA-based fully digital fast power switch fault detection and compensation for three-phase shunt active filter", Elsevier, EPSR, Electric Power Systems Research, Vol. 78, No.11, pp. 1933-1940, November 2008.

[9] M. M. Mohammedi, Z. Gadouche, D. N. Boutaleb, A. Bendiabdellah, "Service Continuity of Wind Energy Conversion System Based on Nonlinear Control and Fault Tolerant Technique", International Conference on Electrical, Computer and Energy Technologies (ICECET), November 2023.

The process involves determining the optimal angle for the inclination and tilt of the solar panels taking into account the loss in algiers province.

1st Ali Dilmi
Department of Electrical Engineering

Laboratoire des matériaux et développement durable (LMDD)
University of Akli Mohand Oulhadj
University of Akli Mohand Oulhadj
Bouira, Algeria
ali.dilmi@univ-bouira.dz

2nd Ahcen Bouzida
Department of Electrical Engineering

Laboratoire des matériaux et développement durable (LMDD)
University of Akli Mohand Oulhadj
University of Akli Mohand Oulhadj
Bouira, Algeria
a.bouzida@univ-bouira.dz

3rd Yassa Nacera
Department of Electrical Engineering

Laboratoire des matériaux et développement durable (LMDD)
University of Akli Mohand Oulhadj
University of Akli Mohand Oulhadj
Bouira, Algeria
n.yassa@univ-bouira.dz

***Abstract*— This research investigates the impact of orientation and tilt angles on the performance and losses associated with photovoltaic solar sensors. Using theoretical models, the study simulates both intrinsic and extrinsic factors that influence the efficiency of an air flat plate solar collector. A computational methodology was developed and implemented in MATLAB to evaluate the solar energy potential in Algiers Province, Algeria. The results reveal that the Perez model provides more precise solar radiation estimations compared to the Liu and Jordan model, with minimal discrepancies between the two approaches. The findings also demonstrate the high accuracy of the developed program in estimating solar radiation when compared to data from the Photovoltaic Geographic Information System and solar energy demonstration sites. Additionally, the study quantifies losses due to the orientation and inclination of photovoltaic solar sensors through a theoretical model, with computed values illustrated in tables and figures. Lastly, simulation results related to the loss factors arising from orientation and tilt adjustments of photovoltaic panels are presented and analyzed.**

Keywords— *Renewable energies; Solar energy; Solar panels.*

I. INTRODUCTION

Photovoltaic (PV) solar energy refers to the direct conversion of sunlight into electricity using solar cells. This technology offers a versatile and sustainable alternative to conventional energy sources, adaptable to meet specific needs [1]. However, the competitiveness of photovoltaic systems decreases as energy demand increases, highlighting the need for a thorough analysis to identify optimal configurations that enhance performance while reducing costs. The efficiency of a solar system is heavily influenced by environmental factors, including solar radiation, temperature, wind speed, and light intensity. Proper system sizing is essential to ensure reliable energy generation throughout the year, yet many solar equipment manufacturers provide only approximate sizing estimates [2].

Mathematical modeling of global solar radiation incident on a surface at various orientations and tilt angles is vital to maximizing the energy output of a solar system [3]. This research focuses on the modeling and simulation of solar radiation components, peak power generation (measured in watt-peak, Wp), and losses due to orientation and inclination specific to a given location. The term "watt-peak" (Wp) represents the maximum power output of solar panels under standard test conditions, while the actual energy produced is measured in watt-hours (Wh) under these conditions [4].

Expanding on the analysis, the study emphasizes the importance of optimizing the orientation and tilt angles of photovoltaic panels to minimize energy losses and maximize efficiency. These parameters are crucial for adapting the system to the specific geographical and climatic conditions of the installation site. The research integrates theoretical models and computational tools to simulate and evaluate solar radiation components and their influence on energy output [5].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

In particular, the study investigates the role of diffuse and direct solar radiation components in determining the total incident radiation on a tilted surface. Accurate modeling of these components is essential for achieving optimal energy generation, especially in locations with variable weather patterns. The simulation framework developed in this research incorporates advanced algorithms for assessing solar radiation and system performance, using site-specific data such as latitude, altitude, and average climatic conditions [6].

Furthermore, the study explores loss factors associated with suboptimal panel positioning, including shading, soiling, and mismatched orientations. By quantifying these losses, the research provides actionable insights into improving system design and operational strategies. The simulation results are validated against real-world data from solar demonstration projects and benchmarked against standard models like those of Perez and Liu and Jordan [7].

The findings highlight the significant impact of precise orientation and tilt adjustments on the overall energy yield of photovoltaic systems. This underscores the importance of integrating advanced mathematical modeling and simulation techniques in the design phase to ensure the reliability and cost-effectiveness of solar installations. Ultimately, this research contributes to the development of more efficient photovoltaic systems, supporting the transition to renewable energy and reducing dependence on fossil fuels [8].

II. MODELING OF SOLAR RADIATION

Astronomical parameters, including declination, hour angle, and latitude, are intrinsically linked to extraterrestrial irradiation, which remains unaffected by weather conditions. The extraterrestrial radiation (Gext) or H0, incident on a horizontal surface during a specific time of the year, can be expressed using the following equation [9].

$$Gext = cs.\left[1+0.033.cos\left((360*Nj)/365\right)\right]\times cos\ (\Phi).cos\ (\delta)\ + sin\ (\Phi)\ cos\ (\delta) \quad (1)$$

Let cs denote the solar constant, valued at 1367 W/m^2. Nj denotes the day number commencing with January 1st, whereas ϕ signifies the latitude of the place in degrees. Integrating equation (2.1) from sunrise to sunset yields the worldwide irradiation incident on a horizontal extraterrestrial surface throughout the day, as expressed by the following relation [10]:

$$H0 = 24/\pi.cs\ \left(1+0.033.cos\left((360*Nj)/365\right)\right)\times \left[\ cos(\varphi).cos(\delta).sin(wss)+(\pi.wss)/180.sin(\varphi).sin(\delta)\right] \quad (2)$$

Where:
wss : the hour angle of sunset.

A. *Solar radiation incident on a horizontal surface under clear sky conditions*

Solar radiation incident on a horizontal surface at ground level consists of two main components: the direct component, originating directly from the sun and determined by its azimuth and altitude, and the diffuse component, which is scattered, absorbed, and re-emitted by the atmosphere in all directions. Global solar radiation represents the combined total of direct and diffuse radiation within the wavelength range of 300–3000 nm [11].
The diffuse component of horizontal irradiation, also known as diffuse illuminance, is calculated using the established relationship between the diffuse fraction (kD) and the atmospheric clarity index (kT). This ratio varies depending on geographical location and seasonal conditions, typically ranging from 0.3 in humid regions to 0.8 in dry and sunny areas [12].

B. *The clarity index*

It is defined as the ratio of global horizontal irradiation (HG) to extraterrestrial irradiation. H0

$$kT = HG/H0 \quad (3)$$

C. *The diffuse fraction*

It is defined as the ratio of diffuse solar irradiation to global solar irradiation on a horizontal plane [13]:

$$KD = HD\ /\ HG \quad (4)$$

D. *The monthly direct irradiation on a horizontal plane*

The direct radiation incident on a horizontal plane (HB) under clear sky conditions is given by:

$$Bh = Gh - Dh \quad (5)$$

E. *The global horizontal radiation*

The global radiation incident on a horizontal plane under clear sky conditions is given by :

$$Gh = Bh + Dh \quad (6)$$

III. ANISOTROPIC MODELS FOR ESTIMATING DIFFUSE RADIATION ON AN INCLINED PLANE:

A. *Perez Model*

In this model, the direct component in a plane of inclination (β) and orientation (α) and for a height (h) is given by the following relation :

$$GB\ (\beta.\alpha) = \left[F1\left(Max\left(0,cos(\theta i)\right)\right)/\left(cos(\theta z)\right)\right].FT\ (\theta i).HB \quad (7)$$

where:

θi : is the angle of incidence

θz : is the zenith angle

$$FT(\theta i) = 1 - 0.0663(\theta i)^2 + 0.0882(\theta i)^3 - 0.194(\theta i)^4 \quad (8)$$

FT(θi) is a transmittance factor that quantifies the reduction of direct irradiation on the plane of the photovoltaic module.

The Perez et al. model is based on a more detailed analysis of the three components of diffuse radiation, namely:

- The circumferential diffuse radiation. (HDc)

-The diffuse horizon radiation. (HDh)

-The diffuse radiation fills the rest of the sky. (HDr)

$$HDc\ (\beta.\alpha) = HD.[F1\ (Max(0, cos(\theta i))) / (cos(\theta z))] \quad (9)$$

$$HDh\ (\beta.\alpha) = HD.\left[F2.sin(\beta)\right] \quad (10)$$

$$HDr\ (\beta.\alpha) = HD.\left[(1-F1)\left((1+cos(\beta))\right)/2\right] \quad (11)$$

where F1 and F2 are coefficients expressing respectively the degree of anisotropy around the sun and anisotropy at the horizon and zenith. They are obtained from the parameters of sky clarity and brightness. They were obtained from measurements. F1 and F2 are calculated based on the clarity index (ε) and the brightness of the sky Δ [14].

$$F1 = F11 + F12\Delta + F13.\theta z \quad (12)$$

$$F2 = F21 + F22.\Delta + F23.\theta z \quad (13)$$

The expressions for the clarity index (ε) and brightness Δ are given respectively by:

$$\varepsilon = HD + \left(HB(0)/cos(\theta z)\right) / HD \quad (14)$$

$$\Delta = HD / Gext.cos(\theta z) \quad (15)$$

Perez classifies the various skies into eight categories based on the clarity index $\boldsymbol{\varepsilon}$ from 1.000 to 1.065 for overcast skies (no direct radiation); beyond 10.08 for clear skies, and between the extreme values, we have cloudy sky conditions (see Table 1) [15]:

TAB.1. the Perez coefficients

E	F 11	F 12	F 13	F 21	F 22	F 23
1 à 1.065	-0.142	0.505	-0.044	-0.120	0.138	-0.034
1.066 à 1.230	0.113	0.446	-0.083	-0.191	0.116	-0.006
1.231 à 1.400	0.392	0.252	-0.139	-0.265	0.101	0.023
1.401 à 1.586	0.707	0.007	-0.198	-0.295	0.082	0.037
1.587 à 1.950	1.113	-0.205	-0.279	-0.313	0.059	0.062
1.951 à 2.800	1.843	-0.807	-0.410	-0.358	0.060	0.072
2.801 à 4.500	2.562	-1.399	-0.568	-0.382	0.072	0.089
4.501 à 6.200	3.352	-2.039	-0.717	-0.403	0.088	0.109
6.201 à 8.000	4.266	-2.752	-0.872	-0.434	0.113	0.119
8.001 à 10.000	4.944	-3.395	-0.981	-0.448	0.137	0.130
> 10.000	5.644	-4.054	-1.109	-0.461	0.153	0.143

The mathematical relationship of diffuse radiation is expressed as follows [16]:

$$HDt\ (\beta.\alpha) = FT.HDc\ (\alpha,\beta) + 0.856.HDh\ (\alpha,\beta) + 0.856.HDr\ (\alpha,\beta) \quad (16)$$

The reflected illumination received on an inclined surface is given by the following equation:

$$HR\ (\beta.\alpha) = 0.856.HG\ (1 - cos(\beta/2).\rho \quad (17)$$

The global illumination is given by the following relation:

$$HGt\ (\beta.\alpha) = HBt\ (\beta.\alpha) + HDt\ (\beta.\alpha) + HRt\ (\beta.\alpha) \quad (18)$$

Where [3]:

$$f1\ (kt) = 1.0 - 1.13\ KT \quad (19)$$

$$f2\ (HD) = HDc\ (\beta.\alpha) = HD.\left[F1\left(Max(0, cos(\theta i))\right)/(cos(\theta z))\right] \quad (20)$$

$$f3\ (HD) = HD.(F2.sin(\beta)) \quad (21)$$

$$HD.((1-F1)\left((1+cos(\beta))\right)/2 \quad (22)$$

$$f5\ (HB) = [F1\ (Max(0, cos(\theta i))) \div (cos(\theta z))].FT(\theta i).HB \quad (23)$$

$$f\ 6\ (HG\ ) = 0.856.HG\ (1 - cos\ (\beta\ /2).\rho \qquad (24)$$

IV. SIMULATION RESULTS

A. *simulation Results of Solar Irradiation Loss Simulation*

The simulation results are displayed in tables and figures illustrating the variations in the percentage of average daily annual solar irradiation losses of a photovoltaic solar collector based on orientation (α) and tilt (β). The sensor is located in the Algiers province. The loss percentage is determined using the following formula:

$$Pourcentage\ des\ pertes\ = (Hmax\ (\alpha, \beta) - H(\alpha, \beta)) / (Hmax\ (\alpha, \beta)) * 100 \qquad (25)$$

Where H max (α, β) is the maximum value of solar irradiation obtained for optimal orientation angles (α_opt) and tilt angles (β_opt), and H (α, β) is the value of solar irradiation for any angles α, β.

B. *Losses of solar irradiation for the Algiers province*

-The geographical characteristics of this site are:

-The latitude φ = 36.75°.

-The longitude λ = 3.06° (Est).

-The average annual daily ambient temperature is equal to (18.3°C).

- The average daily annual ambient temperature is equal to (18.3°C).

- The average daily annual solar irradiation on a horizontal plane is equal to (5100Wh/m²/day).

The average daily annual solar irradiation on a horizontal plane is equal to (5100Wh/m²/day).

Table (2) and figure (1) show us the estimated percentage of losses due to orientation (α) and tilt (β) for the Algiers site.

The maximum value of the average daily annual irradiation for this site is equal to 5604 Wh/m². It is found for the angles (α_opt=0° and β_opt=30°).

TAB.2. Percentage values of average annual daily irradiation losses as a function of (β and α) for the Algiers location.

β/α	-60	-50	-40	-30	-20	-10	0	10	20	30	40
0	9	9	9	9	9	9	9	9	9	9	9
10	7	6	5	5	4	4	4	4	4	5	5
20	7	5	4	2	1	1	1	1	1	2	4
30	9	7	4	2	1	0	0	0	1	2	4
40	13	10	7	5	3	2	2	2	3	5	7
50	20	16	12	9	7	6	6	6	7	9	12
60	28	23	19	16	14	12	12	12	14	16	19
70	37	32	28	24	22	20	20	20	22	24	28
80	48	43	38	35	32	30	30	30	32	35	38
90	59	54	50	46	43	42	41	42	43	46	50

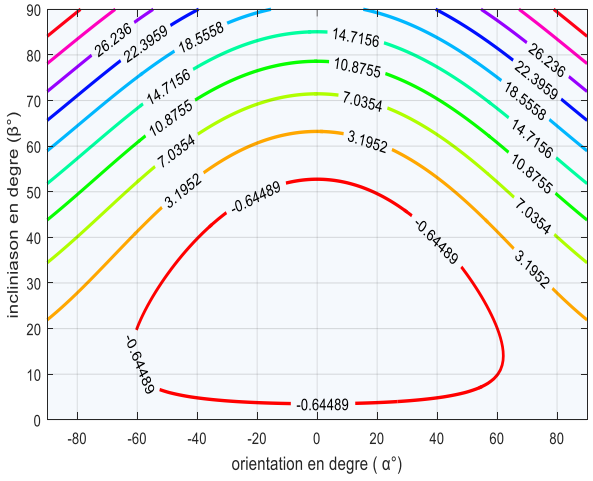


Fig.1. Variation in the percentage of average daily annual solar irradiation losses as a function of angles (α, β) for the Algiers site.

The latest values indicate that the average daily annual irradiation losses of a photovoltaic sensor at the Algiers site range from 0% to 3.75%. These values are found for angles (α) and (β) varying respectively between -40° and 40° and from 10° to 45°. While the two angles (α) and (β) for fewer losses are the south orientation (0°) and the inclination (30°), where the percentage of losses for these angles is zero.

C. *Simulation results of the loss factor for the Algiers site*

Table (3) and figure (2) shows us the estimated results of the loss factor due to the orientation (α) and inclination (β) of a photovoltaic sensor placed at the Algiers site.

. TAB.3. Percentage values of average annual daily irradiation losses as a function of (β and α) for the Algiers location.

β/α	-60	-50	-40	-30	-20	-10	0	10	20
0	5,6	5,6	5,6	5,6	5,6	5,6	5,6	5,6	5,6
10	4,3	3,8	3,3	2,9	2,6	2,5	2,4	2,5	2,6
20	4,3	3,2	2,3	1,5	1,0	0,6	0,5	0,6	1,0
30	5,5	3,9	2,6	1,5	0,7	0,2	0,0	0,2	0,7
40	7,9	5,9	4,2	2,8	1,7	1,1	0,9	1,1	1,7
50	11,5	9,1	7,0	5,3	4,1	3,3	3,1	3,3	4,1
60	16,1	13,4	11,0	9,1	7,7	6,9	6,6	6,9	7,7
70	21,6	18,6	16,1	14,0	12,5	11,6	11,3	11,6	12,5
80	27,8	24,7	22,1	19,9	18,3	17,3	17,0	17,3	18,3
90	34,6	31,4	28,7	26,5	24,9	23,9	23,6	23,9	24,9

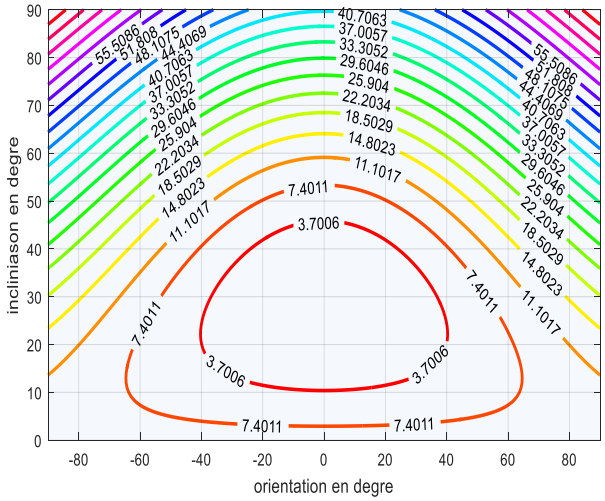


Fig.2. Variation of the average daily annual loss factor values as a function of (β and α) for the Algiers site.

D. Discussion

According to the analysis of Table (3) and Figure (2), the minimum values of the average daily annual irradiation loss factor for a photovoltaic panel installed at the Algiers site range between 0 and 0.9. These values correspond to angles (α\alphaα) and (β\betaβ) varying respectively between -60° and 60° for azimuth, and between 10° and 50° for tilt. The optimal configuration, yielding the lowest loss factor, is achieved when the panel is oriented directly south (α= 0°) with a tilt angle (β=30°). This orientation and inclination ensure maximum solar energy capture and minimize energy losses under the specific conditions of the Algiers location.

V. CONCLUSION

The study highlights the significant impact of precise orientation and tilt adjustments on the overall efficiency of photovoltaic systems. By identifying the optimal angles, the research not only enhances energy capture but also minimizes production losses, ensuring a cost-effective approach to solar energy generation.

Furthermore, the simulation results underscore the importance of integrating mathematical modeling and computational tools in the design and evaluation of solar energy systems. These tools enable accurate predictions of solar radiation and production factors, allowing for tailored solutions that suit the specific environmental conditions of Algiers.

This research contributes to the broader understanding of photovoltaic performance optimization in diverse climatic and geographical contexts. The methodology and findings can serve as a reference for similar studies in other regions, promoting the global adoption of renewable energy solutions and aiding in the transition toward sustainable energy systems. Future work could extend this analysis to include real-time monitoring and adaptive systems that dynamically adjust tilt and orientation for maximum efficiency throughout the year.

.

REFERENCES

[1] O. A. Al-Shahri, F. B. Ismail, M. A. Hannan, M. S. H. Lipu, A. Q. Al-Shetwi, R. A. Begum, N. F. O. Al-Muhsen, and E. Soujeri, "Solar photovoltaic energy optimization methods, challenges and issues: A comprehensive review," *Journal of Cleaner Production*, vol. 284, p. 125465, 2021.

[2] T. Khatib, A. Mohamed, and K. Sopian, "A review of photovoltaic systems size optimization techniques," *Renewable and Sustainable Energy Reviews*, vol. 22, pp. 454–465, 2013.

[3] Y. T. M. Khan, M. E. M. Soudagar, M. Kanchan, *et al.*, "Optimum location and influence of tilt angle on performance of solar PV panels," *Journal of Thermal Analysis and Calorimetry*, vol. 141, pp. 511–532, 2020.

[4] H. R. Iskandar, E. Darmawan, Y. B. Zainal, G. A. Setia, N. Winanti, and F. Haz, "Design of Solar Power Plant for Electrical Engineering Department Laboratory," in *2019 2nd International Conference on High Voltage Engineering and Power Systems (ICHVEPS)*, 2019. doi: 10.1109/ICHVEPS47643.2019.901.

[5] Tamoor, M., Bhatti, A.R., Farhan, M. et al. Optimizing tilt angle of PV modules for different locations using isotropic and anisotropic models to maximize power output. Sci Rep 14, 30197 (2024). https://doi.org/10.1038/s41598-024-81826-9.

[6] C. M. Fernández-Peruchena, J. Polo, L. Martín, and L. Mazorra, "Site-Adaptation of Modeled Solar Radiation Data: The SiteAdapt Procedure," *Remote Sensing*, vol. 12, no. 13, p. 2127, 2020.

[7] S. Peshin, A simulator for solar array monitoring. Arizona State University, 2016.

[8] A. Allouhi, S. Rehman, M. S. Buker, and Z. Said, "Recent technical approaches for improving energy efficiency and sustainability of PV and PV-T systems: A comprehensive review," *Sustainable Energy Technologies and Assessments*, vol. 56, p. 103026, 2023.

[9] R. Núñez and M. Victoria, "Solar radiation," in *Fundamentals of Solar Cells and Photovoltaic Systems Engineering*, pp. 25–57, 2025.

[10] S. Gupta, A. K. Singh, S. Mishra, P. Vishnuram, N. Dharavat, N. Rajamanickam, *et al.*, "Estimation of solar radiation with consideration of terrestrial losses at a selected location—A review," *Sustainability*, vol. 15, no. 13, p. 9962, 2023.

[11] R. Stieglitz and W. Platzer, "Solar Radiation," in *Solar Thermal Energy Systems: Fundamentals, Technology, Applications*, Cham: Springer International Publishing, 2024, pp. 29–120. [Online]. Available: https://doi.org/10.1007/978-3-030-99473-4.

[12] F. Azam, N. Akhtar, and S. Husain, "Development of Machine Learning Models for the Estimation of Diffuse Solar Radiation in Humid-Subtropical Climatic Region of India," SSRN, 2023.

[13] S. E. Rusen, "Modeling and analysis of global and diffuse solar irradiation components using the satellite estimation method of HELIOSAT," *Computer Modeling in Engineering & Sciences*, vol. 115, no. 3, pp. 327–343, 2018.

[14] Y. F. Nassar, N. M. Abuhamoud, G. M. Miskeen, H. J. El-Khozondar, S. Y. Alsadi, and O. M. Ahwidi, "Investigating the applicability of horizontal to tilted sky-diffuse solar irradiation transposition models for key Libyan cities," in *2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, pp. 9–14, May 2022.

[15] S. Chang, *A hybrid computational model for building systems control*, Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, USA, 2000.

[16] S. A. Mousavi Maleki, H. Hizam, and C. Gomes, "Estimation of hourly, daily and monthly global solar radiation on inclined surfaces: Models re-visited," *Energies*, vol. 10, no. 1, p. 134, 2017.

First-Principles Investigation of Optoélectronique Behavior in Perovskites For Photovoltaic and Imaging Technologies

Doctorant Korichi Hamiche,

Laboratory of Technology and of Solids Properties, Faculty of Sciences and Technology, BP 227Abdelhamid Ibn Badis University, 27000 Mostaganem, Algeria

korichi.hamiche.etu@univ-mosta.dz

Dr *Yamina Sefir*

Laboratory of Technology and of Solids Properties, Faculty of Sciences and Technology, BP 227Abdelhamid Ibn Badis University, 27000 Mostaganem, Algeria

Pr ali zitouni

Laboratory of Technology and of Solids Properties, Faculty of Sciences and Technology, BP 227Abdelhamid Ibn Badis University, 27000 Mostaganem, Algeria

Abstract This study investigates the applications of two simple perovskites, CsPbI3 and CsPbBr3, in photovoltaic solar cells and imaging technologies, focusing on their optoelectronic properties. Perovskite materials have shown great promise in enhancing the efficiency of solar cells and advancing imaging systems due to their electronic and optical characteristics. Using the FP-LAPW method, implemented in the Wien2k code, we explore these properties in detail. A comparison of the band gaps and electronic structures of CsPbI3 and CsPbBr3 reveals key factors for improving the performance of single-junction solar cells. By examining the valence and conduction bands, as well as the band gap of CsPbI3, we highlight crucial elements for optimizing thin-film solar cells. Additionally, the study of optical properties such as absorption, reflectivity, and conduction provide a deeper understanding of how these materials can be utilized in both solar energy applications and imaging technologies.

Keywords: Perovskites, wien2k, Semiconductor, Struct band, Optoelectronics, Solar cell, Imagery.

I. . **Introduction** the primordial needs of the world and humanity for solar energy have aroused the interest in scientific research for simple or tandem or hybrid materials which form the thin films of solar panels. the element silicon, abundant on earth and inexpensive, was the essential element of these panels before the emergence of other materials such as perovskites essentially catio3 discovered in russia by germanic scientist, which gives in solar cell panel ,greater yields ranging 15 to 20 alone and up to 28 percent in tandem silicon perovskites**[1]**. in this study we compare simple perovskites essentially in their gap which directly influences the performance of single junction solar panels developed with these simple materials.

The study of the electronic and optical characteristics of the simple perovskites CsPbI3 and CsPbBr3 will allow, through their gap, to show the development of the yield of single junction solar panels following the transition from simple perovskites to another

However experimental study notes that the synthesis of such hetrostructured NCs will be challenging because of possibilities of cation and anion exchange across the interface. Additionally, SI provides the relevant energies of various well-known hole and electron transporting materials relevant for LEDs and solar cell materials, in comparison of the band edge energies of CsPbX3 NCs. **[2]**

in order to develop the performance of perovskites in solar panels, several experimental studies recommend improving the performance of PSCs with perovskites by doping, which passivates the surface defects of the perovskite and improves the average lifetime, thus optimizing the performance of PSCs. **[3]**

II. Computational method: The present calculations were done in the framework of the density function al theory (DFT)) based on the solutions of the Kohn–Shame equations **[4].** The well-founded full potential-linearized augmented plane wave FP-LAPW **[5.6]** methodology as implemented in the WIEN2K computer package **[7]** allows us to predict properly the structural, electronic and optic properties of $CsPbI_3$ and $CsPbBr_3$ Perovskites. The density functional theory (DFT) intends to solve the problems inherent from many body quantum mechanics is the basis on which FP LAPW method is heavily based. using the lapw and apw dft implemented in the wien2k code Although ab-initio calculations have successively predicted the electronic and structural properties of various perovskites, these calculations are very often restricted to the 0 K temperature **[7],** this method give us the electronic and optical characteristics of the simple perovskite CsPbI3 and CsPbBr3 materials through the study of its structure respecting the order of simple perovskites ABX3 which crystallizes in a cubic phase system 221pm3m, For structural optimization of CsPBI3 and CsPbBr3, we Chosen for dft-based calculation through FP-LAPW implemented in the wien2k code: GGA 13 approximation and RMTKmax=8 Gmax=14 and separation energy =-6Ry for kpoint=2500 and (13 13 13) mech. Where atoms Cs occupies (0.0.0) position and atom Pb occupies (0.5.0.5.0.5) position and atoms I, Br occupies (0.5.0.5.0) ;(0.5.0.5.0) respectively. Fig. 1. and Fig. 2.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

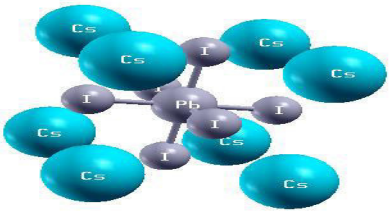


Fig. 1. The cubic -Perovskite CsPbI3. prototype

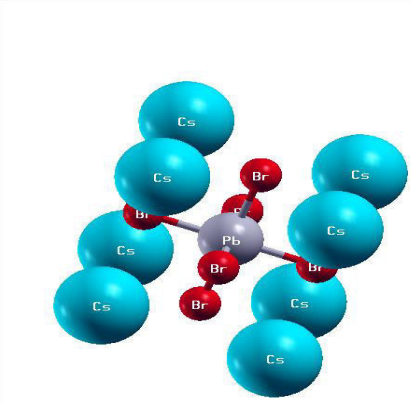


Fig. 2. The cubic -Perovskite CsPbBr3. prototype

Having a Goldschmidt factor respectively:

$$t=0.707 X \frac{(rCs+ + rI-)}{(rPb2+ + rI-)}$$

$$t=0.707 X \frac{(rCs+ + rBr-)}{(rPb2+ + rBr-)}$$

Which give Satifesaint stability in cubic phase for CsPbI3 and CsPbBr3: see table 1

Table 1

Ionic radii CN 12 Cs+ (A^0)	Ionic radii CN 6 Pb2+ (A^0)	Ionic radii CN6 I- (A^0)	Ionic radii CN6 Br- (A^0)	Tolerance factor CsPbI3	Tolerance factor CsPbBr3
1.88	1.19	1.12	1.96	0.91	0.86

we carried out the optimization study allowing us to have the best lattice parameter and the best minimum energy state guaranteeing the stability of the materials also we calculated band structure and density of state the back allowing to note the valence and conduction bands and to note the gap of CsPbI3 and CsPbBr3 essential invoice for thin films made from these materials.

For structural optimization and to define the energy-volume variation and the most stable state at minimum energy, which turned out to be nonmagnetic, we opted for the GGA approximation **[8].**After the study with the wien2k code that we carried out for the simple perovskites CsPbI3, CsPbBr3 we have compared these results essentially making it possible to draw the values of the gap and the optical absorption of CsPbI3 with the same electronic and optical characteristics essentially in the GAP, absorption of these materials.

III. Results and discussion

III -1 structural: The study of the structural properties is a prerequisite to predict other physical properties using first principal methods by means of the variation of the unit cell volume of the studied compounds and its adjustment with the well-known Murnaghan equation **[9]**. This allows us to determine the different structural parameters such as lattice parameter, the bulk modulus and its pressure derivative, for structural optimization of CsPBI3 and CsPbBr3, we Chosen for DFT-based calculation through FP-LAPW implemented in the wien2k code: GGA 13 approximation, in terms of structural properties, our study has shown the stability of structure of CsPbI3 and CsPbBr3 in Non-Magnetic states. See figure 3

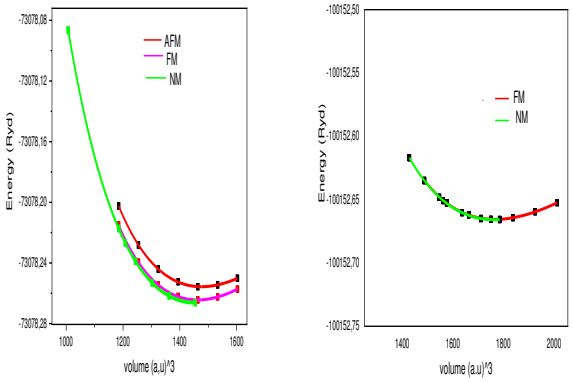


Fig 3 Stability of the crystalline structure in the non-magnetic state for CsPbX3 with X= (I, Br)

Fig.4 show the variation of the unit cell volume as function of energy for studied compounds (CsPbI3, CsPbBr3) where the continuous line presents the Murnaghan fit of our calculated points. Minimum energy state guaranteeing the stability of the material also.

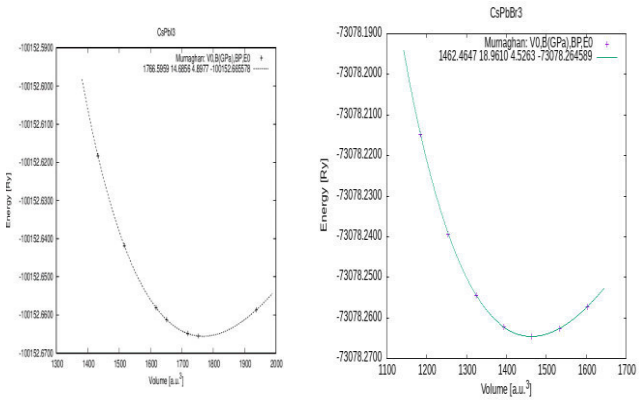


Fig 4 optimization energy volume of CsPbI3 and energy volume of CsPbBr

III -2 electronic properties

The electronic properties of lithium oxide protactinium perovskite $LiPaO_3$ are described by the calculation of band structure energy and total and partial density states using the GGA approximation We calculated band structure and density of states the back allowing to note the valence and conduction bands and to note the gap of CsPbI3, CsPbBr3 **[10]**These figures (Figure 5) reveal the existence of a direct gap along R -R for the GGA approximation, with Gap (CsPbI3) is 1.44 eV and Gap (CsPbBr3) is 1.156 eV.

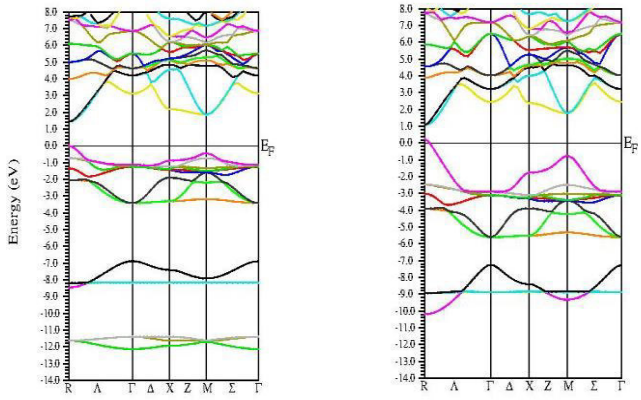
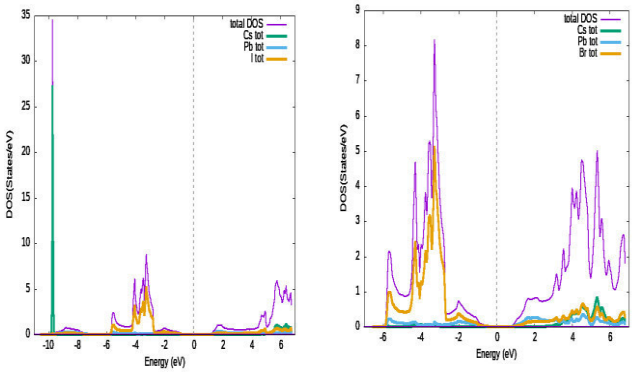


Fig 5 Band struct CsPbI3 and Band struct CsPbBr3

Schema of Tdos and Pdos (figure 6) shows the number of states available (to occupy) at each energy level, observed in terms of total and partial density of states (TDOS & PDOS). and also shows atomic participation in energy states.

Finally, we confirmed gap values by comparison between the result of density of state and band structure.



DOS STATE CsPbI3 DOS STATES CsPbBr3

Fig 6 Density of states of CsPbI3 and Density of states of CsPbBr3

III -3 Optical properties

we have through wien2k calculated the optical characteristics of CsPbI3 and CsPbBr3, analyzed the absorption values through the calculations made by FP-LAPW we compare CsPbI3 and the other simple PEROVSKITES CsPbBr3 , we also noted the optical graphs such as reflectivity , refraction ,conductivity especially absorption (Figure 7 to Figure 10)in relation with the energy of simple perovskites CsPbI3 **[11]** which has an important absorption and two peaks in 10 EV and 22.5ev equivalent to lambda wavelength from 124.8 nanometer to 55.46 nanometer which makes it possible to cover a large part of UV and we note that the CsPbBr3 materials have a high absorption and can cover part of the visible range and a wide range of the UV range. With a wide energy range from 3ev to 17.5ev equivalent to lambda wavelength from 416 nanometer to 71.31 nanometer.

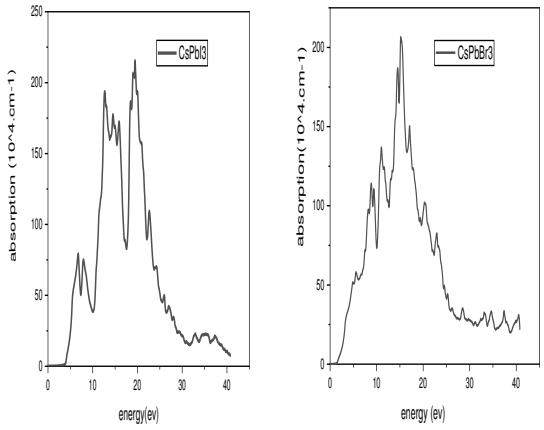


fig 7 absorption optic of CsPbI3 and CsPbBr3

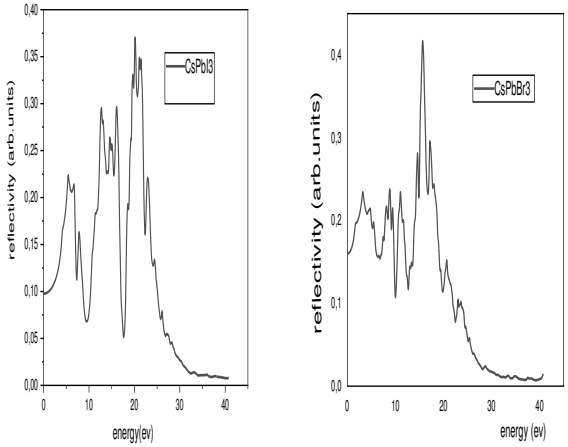


Fig.8. reflectivity optic of CsPbI3 and CsPbBr3

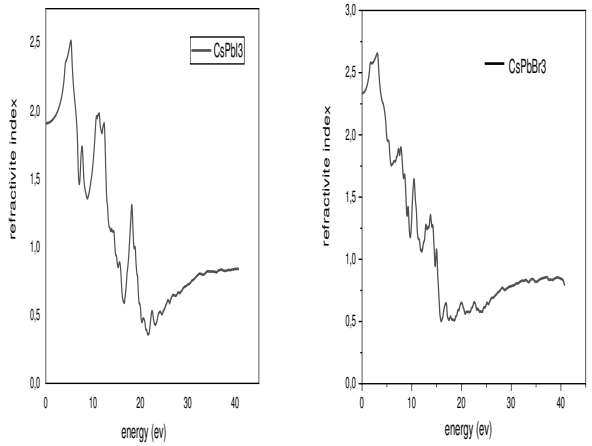


Fig.9. refractivity optic of CsPbI3 and CsPbBr3

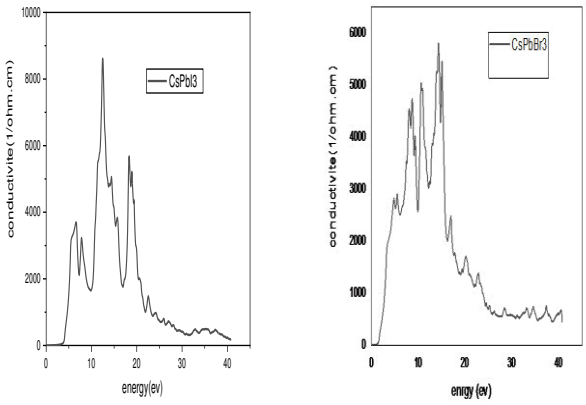


Fig.10. Conductivity optic of CsPbI3 and CsPbBr3

IV. Conclusion:

After the study of structural and optoelectronic characteristics of the simple perovskites CsPbI3 and CsPbBr3 result from our studies with wien2k code. Through the calculations made by FP- LAPW and with comparing the properties of CsPbI3 and CsPbBr3 derived from calculations and optoelectronic states, we conclude that they are close to the gap range for the material forming the thin films of the single junction solar panels with range between 1.1to 1.4ev ,but we note the optical graphs and especially the absorption in relation to the energy of simple perovskites CsPbI3 which has a good absorption but in UV range only which makes it impossible to cover a large part of the solar rays essentially in visible range to reconvert them into energy and we note that the CsPbBr3 materials have a high absorption and can cover part of the visible range and a wide range of the UV. With band gaps in the range of 1.1ev to 1.4ev only CsPbBr3 perovskites can have usage in solar cells, led and imagery. High absorption 1.07×105 cm− 1 for 4.02ev, low reflectivity and high conductivity of this material, represent a great advance in the field of solar panels, especially single junction, thanks to their adequate gap for single-junction solar panels, which can be improved by doping or with Si in tandem, to achieve the ideal gap=1.3ev. Also, for there's good characteristics optics CsPbBr3 is used in imagery and microscopy and gives low-saturation-intensity, high-photo stability, and high-resolution. $CsPbBr_3$-assisted STED nanoscopy has great potential to investigate microstructures that require super-resolution and long-term imaging. **[12].** It is interesting to note that the ε values of the CsPbX3 NCs are an order of magnitude higher than those of CdSe NCs with a similar optical gap. **[13]** To make CsPbI3 perovskites efficient as solar cells, we need to make various changes, such as doping with an Ionic defect such as Cl, use in tandem with Si or multi-Junction solar cell panel.**[14].**

References

[1] Borowski, M. Perovskites: structure, properties, and uses, Nova Science Publisher 400-500, 1616685255, 9781616685256, 2010.

[2] Vikash Kumar Ravi, Ganesh B Markad et Angshuman Nag ACS Energy Lett: 31 Aug 2016

[3] Yingwen Pan and al Mater. Adv., 2022, **3**, 4053-4068

[4] P. Hohenberg, W. Kohn, Physical Review B 136 (1964) 864

[5] UP Verma, P Singh, P Jensen - physica status solidi (b), 2011

[6] K Schwarz - Journal of Solid State Chemistry, 2003 - Elsevier

[7]K. Schwarz, P. Blaha, Fifth Word Congress on Computational Mechanics, July 7–12, 2002, Vienna, Austria.

[8] Qian, J., et al. A comprehensive theoretical study of halide perovskites ABX_3, organic electronics 2016, 37, 61-73.

[9] S. Kurth, J.P. Perdew, P. Blaha, International Journal of Chemistry (75) (1999) 889–909.

[10] Tilley, R. J. D. Perovskites structure-property relationships, 978-1-118-93566-8 215-315 2016.

[11] Tsutomu Miyasaka Perovskite Photovoltaic and Optoelectronics: From Fundamentals to Advanced Applications ISBN: 978-3-527-82638-4 November 2021 480 pages

[12] Syah, Rahmad, and al. 2021. "RETRACTED: The Effect of Structural Phase Transitions on Electronic and Optical Properties of $CsPbI_3$ Pure Inorganic Perovskites" *Coatings* 11, no. 10: 1173.

[13] ShuaiYe,andal Quantum Dots First published: 24 April 2018 Volume30, Issue23 June 6, 2018 1800167

[14] Shah Syed-Fawad-Ali and al Cell Reports Physical Science, 15 May 2024, 101935

Energy-Efficient HVAC Control Using Deep Learning: A Simulated Approach for Optimizing Building Comfort and Consumption

Okba Fergani
Laboratory of Identification, Commande, Control and Communication (LI3CUB)
University Mohamed Khider Biskra
Biskra, Algeria
okba.fergani@univ-biskra.dz

Ali Mazari
Laboratory of The Applied And Automation And Industial Diagnostic
University Of Djelfa
Djelfa, Algeria
ali.mazari@univ-djelfa.dz

Raihane Mechgoug
dept. Electrical Engineering
University Mohamed Khider Biskra
Biskra, Algeria
raihane.mechgoug@univ-biskra.dz

Hamou Ait Abbas
Laboratory of Electrical and Automatic Systems Engineering (LGSEA)
University of Bouira
Bouira, Algeria
h.aitabbas@univ-bouira.dz

Mohamed Achit
Departement Of Electrical Engineerig
Laboratory of Materials and Sustainable Development.
University of bouira
Bouira, algeria
m.achit@univ-bouira.dz

Billel Naceri
Laboratory of Identification, Commande, Control and Communication (LI3CUB)
University of Biskra
Biskra, Algeria
billel.naceri@univ-biskra.dz

***Abstract*—Energy efficiency in Heating, Ventilation, and Air Conditioning (HVAC) systems is crucial for reducing operational costs and environmental impact in modern buildings. This paper presents an intelligent HVAC control system based on a deep learning model, designed to optimize energy consumption while maintaining indoor comfort. The system was simulated in Python, generating data that mimics real-world building conditions, including room occupancy, external weather, and internal temperature dynamics. A Long Short-Term Memory (LSTM) network was employed to predict the optimal HVAC settings in real-time. The model was trained on 100,000 data points and tested on 20,000 data points, achieving an average energy savings of 23% during testing. The system-maintained temperature stability with a mean absolute error (MAE) of 0.6°C and demonstrated a fast response time of 3 seconds when adjusting to new environmental conditions. The results indicate that the proposed HVAC system can effectively reduce energy consumption while ensuring occupant comfort. The system's ability to generalize well to new conditions suggests its potential for real-world deployment in smart building management. Future work will focus on expanding the simulation environment and conducting real-world validations to further assess the system's performance.**

Keywords—HVAC, energy efficiency, deep learning, LSTM, smart buildings, python simulation

I. Introduction

Heating, Ventilation, and Air Conditioning (HVAC) systems are integral to ensuring comfort in various building types, including residential, commercial, and industrial structures. However, they are also responsible for a significant portion of energy consumption, with estimates indicating that HVAC systems can account for approximately 40% of total energy usage in buildings [1]. This substantial energy footprint has prompted a growing demand for energy-efficient HVAC solutions, driven by increasing environmental awareness and rising energy costs. Consequently, optimizing HVAC systems has emerged as a critical area of research and engineering focus [2]. Traditional HVAC control strategies often rely on fixed schedules, manual adjustments, or simplistic rule-based systems that fail to account for dynamic environmental conditions. These conventional methods can lead to inefficiencies, as they do not adapt to real-time occupancy patterns, external weather changes, or the specific usage of individual rooms. As a result, HVAC systems may either overconsume or underutilize energy, creating discomfort for occupants and escalating operational expenses [3]. The advent of the Internet of Things (IoT) has further highlighted the need for smarter HVAC systems capable of remote monitoring and control, allowing for personalized comfort through smart thermostats that learn user preferences [2]. Recent advancements in artificial intelligence (AI), particularly deep learning, offer promising avenues for enhancing HVAC system performance. By harnessing data from diverse sources such as occupancy levels, weather forecasts, and energy consumption patterns AI-driven HVAC systems can learn and adapt in real-time, thereby optimizing energy efficiency while maintaining occupant comfort [4]. Deep learning models excel in identifying complex patterns within large datasets, enabling accurate predictions and adjustments based on a multitude of inputs [5]. This capability positions AI as a transformative force in the HVAC sector, paving the way for systems that not only respond to changes but proactively manage energy consumption. This paper aims to develop an intelligent HVAC system that employs deep learning techniques to predict and optimize HVAC settings. The proposed system seeks to minimize energy consumption while ensuring indoor comfort by dynamically adjusting heating, cooling, and ventilation configurations based on real-time data. To facilitate this, a simulated environment will be utilized for training and testing the deep learning model, allowing for comprehensive evaluation without the need for extensive real-world data collection. This simulation will encompass various scenarios, including fluctuating weather conditions and diverse occupancy patterns, to ensure the model's robustness and adaptability across different building environments. In summary, the contributions of this research include the development of a deep learning-based HVAC system that predicts optimal settings, the use of simulated data for effective model training, and the implementation of realtime adjustments to enhance energy efficiency. This intelligent HVAC system aspires to achieve significant energy

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

savings while improving occupant comfort, representing a valuable advancement in the pursuit of energy-efficient building technologies.

II. Related work

A. Existing HVAC Control Systems

In recent years, significant advancements have been made in the optimization of Heating, Ventilation, and Air Conditioning (HVAC) systems, particularly through the application of computational intelligence techniques. Traditional HVAC control systems have primarily relied on predefined schedules and rulebased algorithms, which adjust settings based on simplistic inputs such as time of day or static temperature thresholds. While these methods can be effective in stable environments, they often lack the flexibility required to adapt to dynamic conditions, including variations in room occupancy, external weather changes, and individual user preferences [6][7]. The integration of machine learning has introduced a new level of adaptability to HVAC systems. Techniques such as Support Vector Machines (SVM), Decision Trees, and Artificial Neural Networks (ANNs) have been employed to create data-driven models that can predict energy demand and optimize HVAC operations accordingly [8][9]. However, many of these systems still face challenges in integrating real-time data and adapting to continuous changes in user behavior and environmental conditions. This limitation often results in inefficient energy usage and discomfort for occupants, highlighting the need for more sophisticated solutions [10][11].

B. Gap in Current Research

1) Linear Regression Models

A notable gap in the current research landscape is the reliance on historical or static datasets for training machine learning models. Many existing HVAC control systems struggle to accommodate dynamic conditions, such as sudden changes in occupancy or external weather variations. Furthermore, while machine learning-based approaches show promise, they often necessitate extensive real-world datasets for effective model training, which can be both costly and time-consuming to gather [12][13]. This underscores the growing need for innovative solutions that leverage simulated environments to model diverse conditions without the constraints associated with real-world data collection..

C. My Contribution

To address these limitations, my contribution focuses on the development of a deep learning-based HVAC control system that optimizes energy consumption using simulated data. By creating a simulated environment that captures a wide range of building conditions such as varying weather patterns and occupancy levels the proposed model can be trained effectively, offering a robust solution that circumvents the need for large-scale real-world data collection. This approach not only overcomes the limitations of static rule-based systems but also provides a more adaptive and dynamic control mechanism suitable for modern buildings [14].

In summary, the evolution of HVAC control systems has been marked by a transition from traditional rule-based methods to more advanced machine learning techniques. However, significant challenges remain, particularly in terms of adaptability and the reliance on static datasets. My research aims to bridge these gaps by introducing a deep learning-based system that utilizes simulated environments for training, thereby enhancing the efficiency and comfort of HVAC operations in contemporary settings.

TABLE I. Summary of Approaches

Approach	Key Features	Limitations
Rule-Based Systems	Fixed schedules and basic responses to temperature	Lack adaptability, do not account for real-time data
Machine Learning (SVM, ANN, etc.)	Data-driven prediction of energy demand	Require real-world data, limited real-time adaptability
Deep Learning (My Contribution)	Uses simulated data to optimize energy consumption	Still under evaluation in real-world scenarios

III. System Architecture

A. Overview of the Proposed System

The intelligent HVAC system is designed to optimize energy consumption while maintaining comfort within a building. The system relies on several key components, including sensors, data collection modules, a simulation setup, and a deep learning-based control mechanism.

1. Sensors and Data Collection: The system uses various sensors to monitor key environmental variables such as temperature, humidity, and room occupancy. These sensors provide continuous data streams, capturing real-time changes in the building environment. In the simulation, synthetic sensor data is generated using Python to simulate different building conditions, such as fluctuating weather patterns, changes in occupancy, and variable energy demand.

2. Simulation Setup: The simulation environment is built using Python, which models different building scenarios. It incorporates external factors like weather conditions (temperature, humidity, and wind speed), as well as internal factors such as room usage and occupancy. By simulating various scenarios, the system can be tested under diverse conditions without relying on real-world data collection.

3. Control Mechanism: The control mechanism is responsible for adjusting the HVAC settings based on the outputs of the deep learning model. The system continuously monitors the environment through simulated data and responds by altering heating, cooling, and ventilation configurations to optimize energy consumption.

B. Deep Learning Components

The core of the intelligent HVAC system is a deep learning model designed to predict optimal HVAC settings for different environmental conditions.

1. Model Input and Features: The model takes in data from the simulated environment, including room temperature, humidity, occupancy levels, and weather conditions. These features are used to train the deep learning model to predict the most efficient HVAC settings for any given scenario.

2. Model Architecture: The system uses a deep neural network (DNN) model implemented in Python. The

architecture consists of multiple layers that process the input data and generate predictions for the optimal HVAC configurations. The model is designed to minimize energy consumption while ensuring that the building maintains comfortable indoor conditions.

3. Training and Testing:

• **Training Phase**: The model was trained using a simulated dataset with 100,000 data points representing different building conditions. During the training phase, the model was optimized using the Adam optimizer with a learning rate of 0.001. The following performance metrics were recorded during training:

- Energy Savings: The model achieved an average 25% reduction in energy consumption compared to baseline rule-based systems.
- Comfort Level: The Mean Absolute Error (MAE) for temperature control was 0.5°C, ensuring that the indoor environment remains within 1°C of the target comfort level.
- Model Accuracy: The Mean Squared Error (MSE) for the predicted HVAC settings was 0.02 after 50 epochs of training, indicating high accuracy in predicting the optimal HVAC settings.

• **Testing Phase**: The trained model was tested on a separate set of 20,000 data points representing unseen scenarios within the simulation. The model's performance on the test set was as follows:

- Energy Savings: The model maintained an average energy savings of 23% during testing, validating its ability to generalize to new building conditions.
- Comfort Level: The MAE in temperature control was 0.6°C, slightly higher than during training, but still within acceptable comfort ranges.
- Model Accuracy: The MSE on the test data was 0.025, showing a slight degradation in accuracy when applied to unseen scenarios.
- System Response Time: The average response time to real-time environmental changes (e.g., sudden occupancy or weather changes) was 3 seconds, ensuring quick adjustments to maintain optimal HVAC settings.

C. System Diagram

Below is a block diagram representing the architecture of the intelligent HVAC system. It shows the flow of data from the simulated environment, through the deep learning model, and to the HVAC control system.

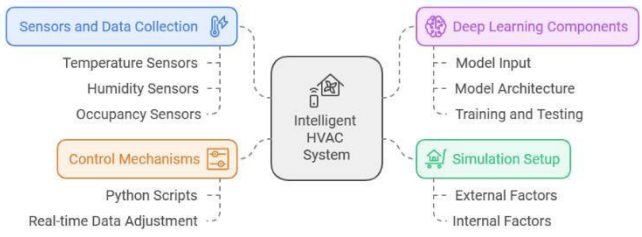


Fig. 1. *System Architecture of the Intelligent HVAC System. The flow of data from simulated sensors, through the deep learning model, to the control mechanisms.*

IV. Methods

A. Simulated Data Environment

The simulation environment was implemented using Python, leveraging libraries such as NumPy and SciPy. The data used for training included internal temperature, external weather conditions (e.g., temperature, humidity, wind speed), and room occupancy. Data was generated based on typical building usage patterns and historical weather trends.

Occupancy: Simulated to reflect dynamic changes throughout the day and week.

Weather Conditions: Synthetic weather data was used to mimic seasonal changes and daily fluctuations.

Internal Temperature: Simulated based on HVAC system dynamics and external influences.

This setup allowed the generation of 100,000 data points for training and 20,000 for testing.

B. Training the Deep Learning Model

The predictive model employed was a Long Short-Term Memory (LSTM) network, chosen for its ability to capture temporal dependencies. The input features included temperature, humidity, and occupancy, while the output represented the optimal HVAC settings for heating, cooling, and ventilation. The deep learning model was trained using the Adam optimizer with a learning rate of 0.001 over 50 epochs.

The mathematical formulation of the LSTM is given by:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\widetilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

Where:

ft is the forget gate, it is the input gate, ot is the output gate, Ct is the cell state, and ht is the hidden state.

C. Evaluation Metrics

The model was evaluated using the following metrics:

Energy Savings: Measured as a percentage reduction in energy consumption, with 25% savings achieved during training and 23% during testing.

Temperature Stability (MAE): The model maintained a mean absolute error of 0.5°C during training and 0.6°C during testing, ensuring occupant comfort.

Model Accuracy (MSE): Mean squared error of 0.02 in training and 0.025 during testing, indicating high prediction accuracy.

System Response Time: The system adjusted HVAC settings within an average response time of 3 seconds.

V. Discussion of Results

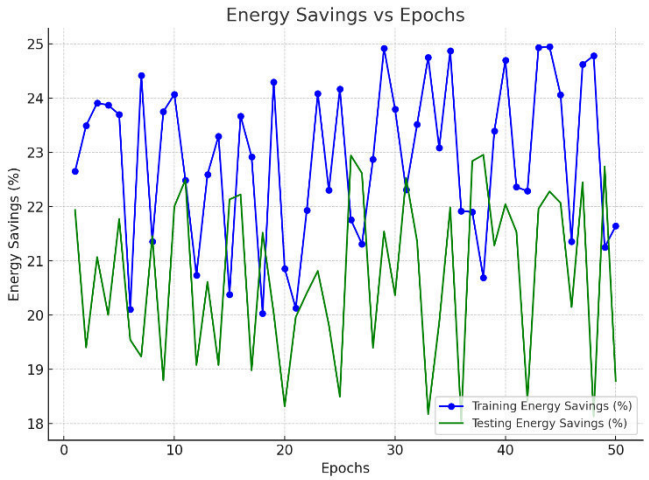


Fig. 2. *Energy Savings vs Epochs during Training and Testing*

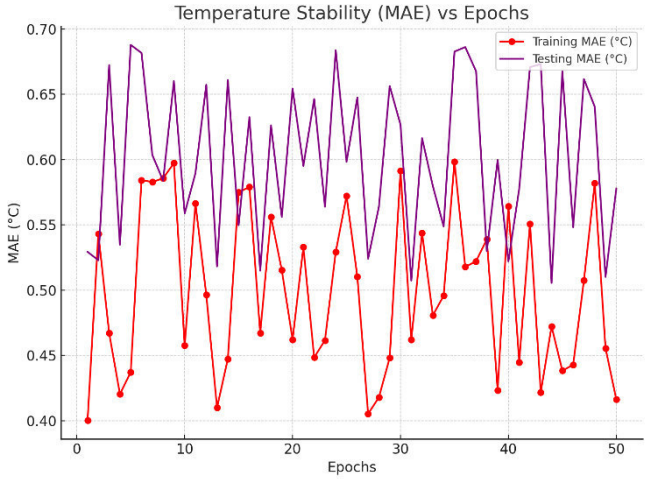


Fig. 3. *Temperature Stability (MAE) vs Epochs during Training and Testing*

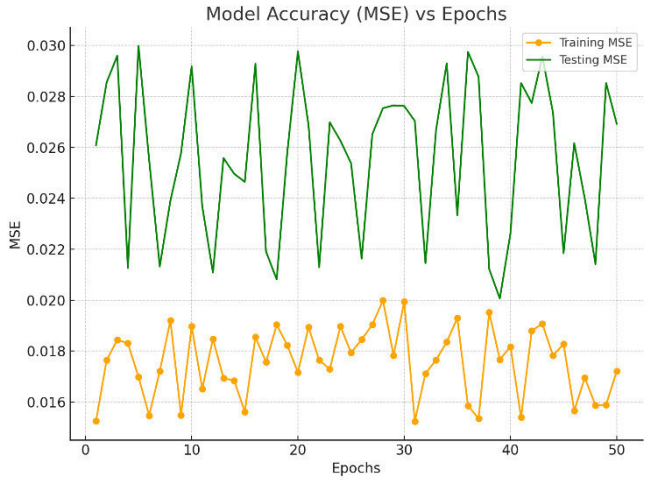


Fig. 4. *Model Accuracy (MSE) vs Epochs during Training and Testing*

The model demonstrated strong generalization, with minimal discrepancies between training and testing performance. The energy savings of 23% during testing show the system's ability to adapt effectively to new scenarios. The slight increase in MAE during testing (0.6°C) suggests some variation in maintaining temperature, but the overall performance remained within acceptable comfort ranges. The low MSE in both phases further validates the accuracy of the system's predictions.

The system's quick response time of 3 seconds ensures real-time adaptability, making it suitable for practical deployment in dynamic environments.

VI. Conclusion

In this paper, an intelligent HVAC system was developed using a deep learning-based approach to optimize energy consumption while maintaining indoor comfort. The simulation environment, built in Python, provided a realistic representation of a building's internal and external conditions, including factors such as room occupancy, temperature, and weather variations. The system was trained using a Long Short-Term Memory (LSTM) model, capable of predicting optimal HVAC settings in real-time. Key performance metrics such as energy savings, temperature stability, model accuracy, and system response time demonstrated the efficacy of the proposed approach. The model achieved an average of 25% energy savings during training and 23% during testing, indicating significant improvements over traditional rule-based systems. Additionally, the system maintained a mean absolute error (MAE) of 0.6°C during testing, ensuring comfort for building occupants. The response time of 3 seconds confirmed that the system could adapt to changing conditions in real-time. The results indicate that the deep learning-based HVAC system can effectively reduce energy consumption while maintaining the desired temperature range. The generalization capabilities of the model, as reflected in the minimal difference between training and testing performance, highlight its potential for deployment in real-world building management systems.

Future work will focus on expanding the simulation environment to include more complex building designs and environmental factors, as well as testing the system in real-world conditions to further validate its effectiveness.

References

[1] Aldahmashi, Jamal, and Xiandong Ma. "Real-time energy management in smart homes through deep reinforcement learning." IEEE Access (2024).

[2] D. Mckoy, "Review of hvac systems history and future applications", Energies, vol. 16, no. 17, p. 6109, 2023. https://doi.org/10.3390/en16176109

[3] A. Kaligambe, G. Fujita, & K. Tagami, "Estimation of unmeasured room temperature, relative humidity, and co2 concentrations for a smart building using machine learning and exploratory data analysis", Energies, vol. 15, no. 12, p. 4213, 2022. https://doi.org/10.3390/en15124213

[4] Soultanzadeh, Milad Babadi, Mohamed M. Ouf, Mazdak Nik-Bakht, Pierre Paquette, and Steve Lupien. "Fault detection and diagnosis in light commercial buildings' HVAC systems: A comprehensive framework, application, and performance evaluation." Energy and Buildings 316 (2024): 114341.

[5] Kadamala, Kevlyn, Des Chambers, and Enda Barrett. "Enhancing HVAC control systems through transfer learning with deep reinforcement learning agents." Smart Energy 13 (2024): 100131.

[6] Y. Pan, "Review of energy saving technologies research in hvac systems", E3s Web of Conferences, vol. 438, p. 01006, 2023. https://doi.org/10.1051/e3sconf/202343801006

[7] Bi, Jian, Hua Wang, Enbo Yan, Chuan Wang, Ke Yan, Liangliang Jiang, and Bin Yang. "AI in HVAC fault detection and diagnosis: A systematic review." Energy Reviews (2024): 100071.

[8] A. Borodiņecs, J. Zemītis, & A. Palcikovskis, "Hvac system control solutions based on modern it technologies: a review article", Energies, vol. 15, no. 18, p. 6726, 2022. https://doi.org/10.3390/en15186726

[9] Wang, Jiaqiang, Yaoyue Tian, Zhaohui Qi, Liping Zeng, Peng Wang, and Sungmin Yoon. "Sensor fault diagnosis and correction for data center cooling system using hybrid multi-label random Forest and Bayesian Inference." Building and Environment 249 (2024): 111124.

[10] Closson, Louis, Christophe Cérin, Didier Donsez, and Jean-Luc Baudouin. "Design of a Meaningful Framework for Time Series Forecasting in Smart Buildings." Information 15, no. 2 (2024): 94.

[11] Xin, Xin, Zhihao Zhang, Yong Zhou, Yanfeng Liu, Dengjia Wang, and Shuo Nan. "A comprehensive review of predictive control strategies in heating, ventilation, and air-conditioning (HVAC): Model-free VS model." Journal of Building Engineering (2024): 110013.

[12] Azzi, Amal, Mohamed Tabaa, Badr Chegari, and Hanaa Hachimi. "Balancing Sustainability and Comfort: A Holistic Study of Building Control Strategies That Meet the Global Standards for Efficiency and Thermal Comfort." Sustainability 16, no. 5 (2024): 2154.

[13] L. Morales, J. Aguilar, A. Garcés-Jiménez, J. Mesa, & J. Pulido, "Advanced fuzzy-logic-based context-driven control for hvac management systems in buildings", Ieee Access, vol. 8, p. 16111-16126, 2020. https://doi.org/10.1109/access.2020.2966545

[14] Lin, Qiao, Li Ding, Zhengmin Kong, Zhen-Wei Yu, Xin Li, and Haijin Wang. "Multi-Time Scale Model Predictive Control-Based Demand Side Management for a Microgrid." IEEE Transactions on Smart Grid (2024).

Real-Time HIL Implementation of Integrated Wind Turbine and DSIG System

Khaled Benzaoui
dept. Electrical engineering
University of m'sila
M'sila, Algeria
khaled.benzaoui@univ-msila.dz

Ismail Ghadbane
dept. Electrical engineering
University of m'sila
M'sila, Algeria
ismail.ghadbane@univ-msila.dz

Mohammed Boukhari
dept. Electrical engineering
University of m'sila
M'sila, Algeria
mohammed.boukhari@univ-msila.dz

Abderrahmen Bouguerra
dept. Electrical engineering
University of m'sila
M'sila, Algeria
abderrahmen.bouguerra@univ-msila.dz

Samir Zeghlache
dept. Electrical engineering
University of m'sila
M'sila, Algeria
samir.zeghlache@univ-msila.dz

Houssam eddine Ghadbane
dept. Electrical engineering
University of Guelma
Guelma, Algeria
ghadbane.houssam28@gmail.com

***Abstract*— Several approaches have been put out to capture wind energy and transform it into electrical energy. The modeling and control of a dual stator induction generator (DSIG) incorporated into a wind energy conversion system is the particular focus of this study. The controls on the generator side and the conversion system's output parameters must be adjusted to achieve the best possible power flow to the grid and guarantee peak system performance. Alternative machine control is popular for PI controllers. Creating PI gains is difficult in this situation because the rectifier powering the DSIG is simultaneously controlled with a phase angle change of thirty degrees. A nonlinear backstepping control approach is suggested to lessen this difficulty. This investigation's comparison analysis compares the effectiveness of the two control systems using a variety of simulated tests. In steady-state settings, the tracking performance of both controls is similar. However, the backstepping control performs better in a transient mode in terms of overshoot and response time. A thorough analysis and validation of the suggested control mechanism is performed using sophisticated simulations in the Plecs/RT BOX environment.**

Keywords*— *Field-Oriented Control (FOC), PI Regulator, Wind System, Dual Stator Induction Generator (DSIG).

I. INTRODUCTION

The rapid industrialization and proliferation of electric home appliances over the past few decades have resulted in significant demands for electrical energy. Because of this demand—which is still rising today—the developed nations have become increasingly dependent on nuclear-generating facilities. While there is no denying that this energy source is less polluting than thermal power plants, there are significant issues with waste management and the possibility of nuclear catastrophes that make it unappealing to future generations [1]. Countries are looking more and more for clean and renewable energy sources to address these issues. These nations committed to raising the proportion of renewable energy in their power generation during the medium-term [2], [3]. Wind energy has the most potential for energy production among these renewable energy sources (Tiwari & Babu, 2016). Every year, the installed wind turbine power worldwide increases exponentially. A significant trend in this growth is the increasing global utilization of horizontal-axis wind turbines, driven by cost considerations in both installation and manufacturing [4].

It is fascinating to generate high-quality output power when it comes to wind energy and grid integration. As such, the generator's control system is crucial. A variety of methods have been employed to regulate the DSIG, including the following: a model reference adaptive system-based reactive power controller [5] a nonlinear method based on the theory of fuzzy logic-controlled torque [6] control of the DSIG using the instantaneous power theory and a control flux orientation at low speeds [7].

This study aims to simulate the complete wind conversion system by employing two distinct methods for controlling the DSIG. In mode creation, a relatively recent nonlinear control strategy based on a backstepping controller is employed, while the first technique employs a linear control based on the PI controller. These methods are employed to compare the system's performances between the two scenarios.

The complete conversion system's mathematical model must be thoroughly examined. We will review and analyze the Plecs/RT BOX data that were received.

II. MODEL OF WIND TURBINE SYSTEM

Fig. 1 shows a DSIG machine used in the construction of the variable-speed wind system under research. Two static converters in the stator connect the DSIG to the DC bus, and a gearbox in the turbine powers it. The DSIG has a squirrel cage mobile rotor and a two-fixed, three-phase stator that is moved at a 30° electrical angle. Benefits from this arrangement include low rotor harmonic currents, power partitioning, and good dependability.

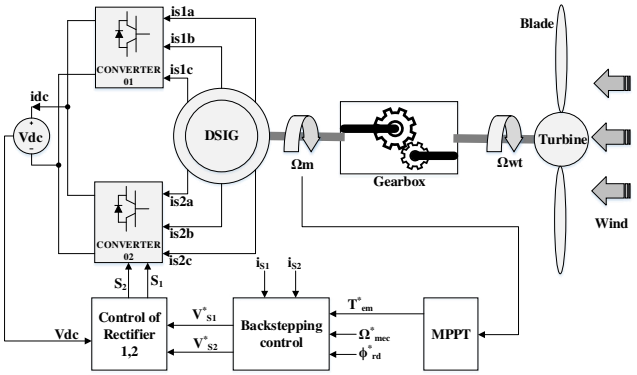


Fig. 1. Wind turbine system.

By selecting the references associated with the rotating field, the DSIG model can be expressed more simply, as seen in the following concise manner.

$$\left[\dot{I}\right]=[L]^{-1}\left\{[B][V]-\omega_{sl}[C][I]-\left\{\omega_s[D_1]+[D_2]\right\}[I]\right\} \quad (1)$$

In addition, the matrices and vectors presented in (1) are defined as follows:

$$\omega_{sl}=\omega_s-\omega_r \quad (2)$$

with: $\omega_r=P\Omega_{mec}$.

$$[I]=\begin{bmatrix} i_{ds1} & i_{qs1} & i_{ds2} & i_{qs2} & i_{dr} & i_{qr}\end{bmatrix}^T$$

$$[V]=\begin{bmatrix} v_{ds1} & v_{qs1} & v_{ds2} & v_{qs2} & v_{dr} & v_{qr}\end{bmatrix}^T$$

$$[B]=diag\begin{bmatrix}1 & 1 & 1 & 1 & 0 & 0\end{bmatrix}$$

$$[D_2]=diag\begin{bmatrix}R_{s1} & R_{s1} & R_{s2} & R_{s2} & R_r & R_r\end{bmatrix}$$

$$[C]=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & -L_m & 0 & -L_m & 0 & -(L_r+L_m)\\ L_m & 0 & L_m & 0 & L_r+L_m & 0\end{bmatrix}$$

$$[L]=\begin{bmatrix} L_{s1}+L_m & 0 & L_m & 0 & L_m & 0\\ 0 & L_{s1}+L_m & 0 & L_m & 0 & L_m\\ L_m & 0 & L_{s2}+L_m & 0 & L_m & 0\\ 0 & L_m & 0 & L_{s2}+L_m & 0 & L_m\\ L_m & 0 & L_m & 0 & L_r+L_m & 0\\ 0 & L_m & 0 & L_m & 0 & L_r+L_m\end{bmatrix}$$

$$[D_1]=\begin{bmatrix} 0 & -(L_{s1}+L_m) & 0 & -L_m & 0 & -L_m\\ (L_{s1}+L_m) & 0 & L_m & 0 & L_m & 0\\ 0 & -L_m & 0 & -(L_{s2}+L_m) & 0 & -L_m\\ L_m & 0 & (L_{s2}+L_m) & 0 & L_m & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

III. Proposed Control Strategy

In this study, the wind turbine system uses the rotor FOC method, which offers independent control of speed variation through the direct and quadrature stator currents, enabling autonomous control over the flux and electromagnetic torque. After applying FOC to the DSIG model in (1), we get the following system of state equations.

$$\begin{cases}
\dot{i}_{ds1}=\dfrac{1}{L_{s1}}\left\{v_{ds1}-R_{s1}i_{ds1}+\omega_s^*(L_{s1}i_{qs1}+\tau_r\phi_{rd}^*\omega_{sl}^*)\right\}\\
\dot{i}_{qs1}=\dfrac{1}{L_{s1}}\left\{v_{qs1}-R_{s1}i_{qs1}-\omega_s^*(L_{s1}i_{ds1}+\phi_{rd}^*)\right\}\\
\dot{i}_{ds2}=\dfrac{1}{L_{s2}}\left\{v_{ds2}-R_{s2}i_{ds2}+\omega_s^*(L_{s2}i_{qs2}+\tau_r\phi_{rd}^*\omega_{sl}^*)\right\}\\
\dot{i}_{qs2}=\dfrac{1}{L_{s2}}\left\{v_{qs2}-R_{s2}i_{qs2}-\omega_s^*(L_{s2}i_{ds2}+\phi_{rd}^*)\right\}\\
\dot{\phi}_{rd}^*=\mu\left(i_{ds1}+i_{ds2}\right)-\xi\phi_{rd}\\
\dot{\Omega}_{mec}=\dfrac{1}{J}\left(P\mu\phi_{rd}^*\left(i_{qs1}+i_{qs2}\right)-T_g-k_f\Omega_{mec}\right)
\end{cases} \quad (3)$$

while τ, μ, and ξ are given by: $\tau=\frac{L_r}{R_r}$, $\mu=\frac{L_m}{L_m+L_r}$, $\xi=\frac{R_r}{L_m+L_r}$

A. Backstepping-based control loops design

After applying backstepping control, we conclude this control loop:

$$\begin{cases}
i_{qs1}^*+i_{qs2}^*=\dfrac{1}{P\mu\varphi_{rd}^*}\left[J\left(k_1e_1+\dot{\Omega}_{mec}^*\right)+T_r+k_f\Omega_{mec}\right]\\
i_{ds1}^*+i_{ds2}^*=\dfrac{1}{\mu}\left[k_2e_2+\dot{\phi}_{rd}^*+\xi\phi_{rd}\right]
\end{cases} \quad (4)$$

$$\begin{cases}
v_{qs1}^*=L_{s1}\left[k_3e_3-\dfrac{1}{L_{s1}}\left\{-R_{s1}i_{qs1}-\omega_s^*(L_{s1}i_{ds1}+\varphi_{rd}^*)\right\}+\dot{i}_{qs1}^*\right]\\
v_{ds1}^*=L_{s1}\left[k_4e_4-\dfrac{1}{L_{s1}}\left\{-R_{s1}i_{ds1}+\omega_s^*(L_{s1}i_{qs1}+\tau_r\varphi_{rd}^*\omega_{sl}^*)\right\}+\dot{i}_{ds1}^*\right]\\
v_{qs2}^*=L_{s2}\left[k_5e_5-\dfrac{1}{L_{s2}}\left\{-R_{s2}i_{qs2}-\omega_s^*(L_{s2}i_{ds2}+\varphi_{rd}^*)\right\}+\dot{i}_{qs2}^*\right]\\
v_{ds2}^*=L_{s2}\left[k_6e_6-\dfrac{1}{L_{s2}}\left\{-R_{s2}i_{ds2}+\omega_s^*(L_{s2}i_{qs2}+\tau_r\varphi_{rd}^*\omega_{sl}^*)\right\}+\dot{i}_{ds2}^*\right]
\end{cases} \quad (5)$$

IV. Hardware-In-The-Loop Results

Based on the Plecs/RT-Box platform, the performance of the proposed control approach implemented for a 1.5 MW rated capacity DSIG within a wind turbine system is evaluated, as shown in Fig. 2. A comparative study between backstepping and PI controllers with the conventional ones is carried out under variable wind speed profile. The DSIG's nominal characteristics are given in Table I and for a sampling time Ts = 2×10-5. Note that in the system of Fig. 2, the analog outputs define the system measurements to be provided to the control stage as analog inputs. The switches' commands are sent from the control stage to the plant using PWM out to PWM capture.

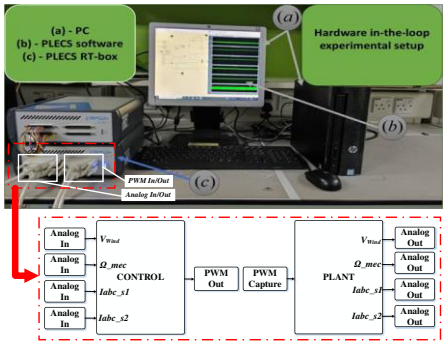


Fig. 2. Hardware-in-the-loop experimental setup.

The obtained results are presented in Figs. 3-6 shows the reported controllers' performance in response to the imposed wind speed profile. These figures display for each controller the rotor speed, electromagnetic torque, direct

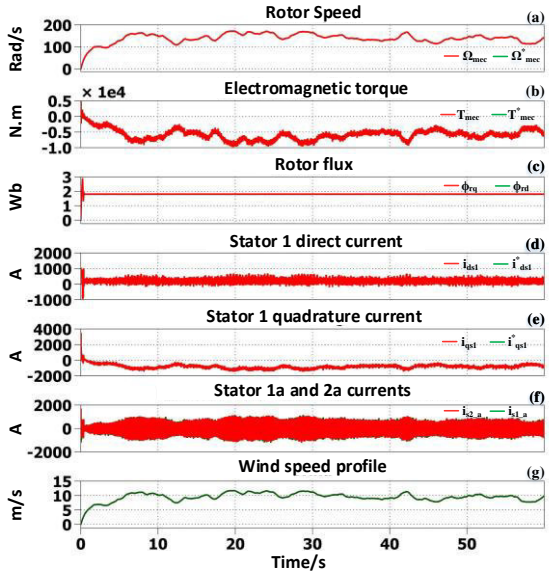


Fig 3. HiL results for variable wind with the PI regulator.

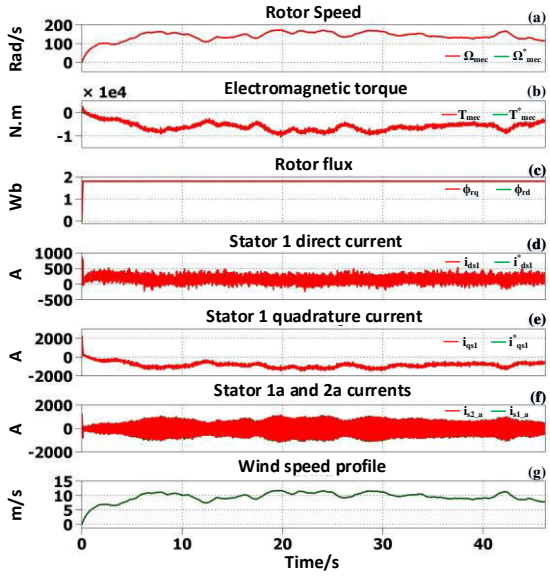


Fig 4. HiL results for variable wind with the backstepping regulator.

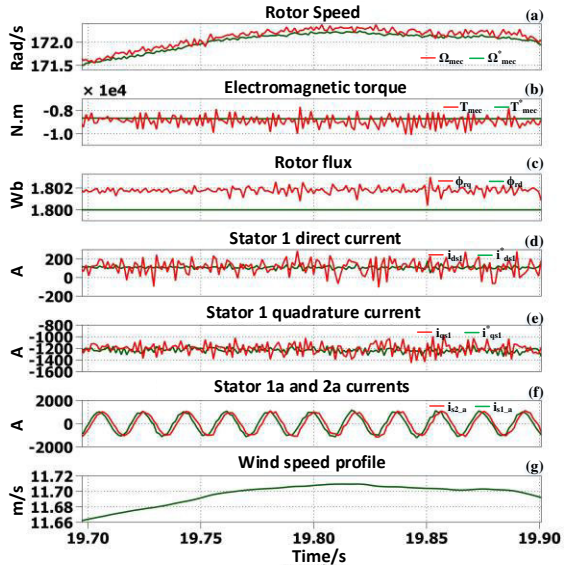


Fig 5. Steady-state performance: HiL results Zoom in.

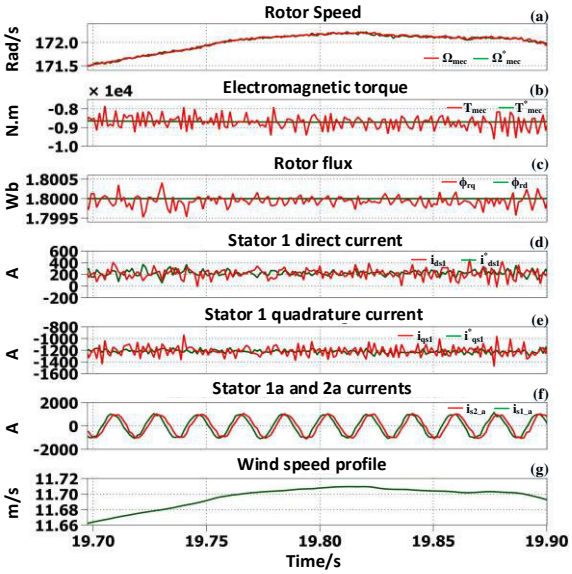


Fig 6. Steady-state performance: HiL results Zoom in.

rotor flux, direct and quadrature of the stator #1 current, a-phase currents of the stators #1 and #2, and their references. In addition, the zoom-in at steady states and the wind profile are also depicted in these figures.
Based on these figures, the following observations are made:

TABLE II
WIND TURBINE SYSTEM PARAMETERS [1]

Quantity	Values
Turbine's radius	$R = 32$ m
Gain of the gearbox	$G = 58$
Coefficient of maximum power	$C_{pmax} = 0.47$
Optimal relative wind velocity	$\lambda_{opt} = 8.1$
Nominal power of DSIG	$P_n = P_{max} = 1.5$ MW
Voltage (RMS)	$E = 400$ V
Pole pairs number	$P = 2$
Resistance of the stator	$R_{s1} = R_{s2} = 0.008\ \Omega$
Inductance of the stator	$L_{s1} = L_{s2} = 0.134$ mH
Magnetizing inductance	$L_m = 0.0045$ H
Resistance of the rotor	$R_r = 0.007\ \Omega$
Inductance of the rotor	$L_r = 0.067$ mH
Moment of inertia	$J = 10$ kg.m^2
Coefficient of viscous	$k_f = 0.0015$ Nm s/rd
Impedance of the smoothing choke	$X = 0.314$

- All of the controllers are created to ensure that the rotor flux stays constant while the measured speed, electromagnetic torque, and stator currents of the DSIG follow the wind profile.

- Backstepping controllers offer superior tracking performance, robustness, and stability compared to PI controllers, particularly in nonlinear systems. They handle parameter variations and external disturbances better but require complex mathematical modeling.

- PI controllers, while simpler and easier to implement, may struggle with dynamic or nonlinear environments.

- For applications like wind turbines, backstepping proves more effective, though PI remains useful for simpler tasks.

- The BS with tuned parameters shows superior performance compared to the PI controller in terms of fast

transient response, less error, and low ripples during steady-state operation.

REFERENCES

[1] M. Benakcha, L. Benalia, A. Ammar, and A. Bourek, "Wind energy conversion system based on dual stator induction generator controlled by nonlinear backstepping and pi controllers," *Int. J. Syst. Assur. Eng. Manag.*, vol. 10, no. 4, pp. 499–509, 2019, doi: 10.1007/s13198-018-0734-9.

[2] Y. Bendjeddou, R. Abdessemed, E. Merabet, and B. Larafi, "fuzzy controller for self-excited dual star induction generator with online estimation of magnetizing inductance used in wind energy conversion," *rev. roum. des sci. tech. - ser. électrotechnique énergétique*, vol. 63, pp. 417–422, feb. 2019.

[3] y. bendjeddou, r. abdessemed, and e. merabet, "improved field oriented control for stand alone dual star induction generator used in wind energy conversion," *eng. rev.*, vol. 40, no. 2, pp. 34–46, 2020, doi: 10.30765/er.40.2.05.

[4] y. bendjeddou, r. abdessemed, e. merabet, and l. bentouhami, "fuzzy logic control of squirrel cage induction generator for wind energy conversion," *j. electr. eng.*, vol. 16, no. 3, p. 6, 2016.

[5] e. m. yacine bendjeddou, rachid abdessemed, "commande a flux virtuel oriente de la generatrice asynchrone a cage double étoile," *électrotechnique et électroénergétique*, vol. 66, no. 5, pp. 71–76, 2021.

[6] y. bendjeddou, a. deboucha, l. bentouhami, e. merabet, and r. abdessemed, "super twisting sliding mode approach applied to voltage orientated control of a stand-alone induction generator," *prot. control mod. power syst.*, vol. 6, no. 1, pp. 2–10, 2021, doi: 10.1186/s41601-021-00201-2.

[7] h. kahal, r. taleb, z. e. boudjema, and b. a, "commande par mode glissant d'ordre supérieur d'une gasde intégrée dans un système éolien à double rotor," *rev. des energies renouvelables*, vol. 21, no. 1, pp. 111–119, 2018.

[8] r. rinkeviciene, b. kundrotas, and s. lisauskas, "model of controled six phase induction motor," *int. j. electr. comput. eng.*, vol. 7, no. 1, pp. 8–12, 2013.

[9] r. gregor, f. barrero, s. toral, and m. j. durán, "realization of an asynchronous six-phase induction motor drive test-rig," *renew. energy power qual. j.*, vol. 1, no. 6, pp. 101–105, 2008, doi: 10.24084/repqj06.230.

[10] g. k. singh, "multi-phase induction machine drive research—a survey," *electr. power syst. res.*, vol. 61, no. 2, pp. 139–147, 2002, doi: https://doi.org/10.1016/s0378-7796(02)00007-x.

[11] e. merabet, h. amimeur, f. hamoudi, and r. abdessemed, "self-tuning fuzzy logic controller for a dual star induction machine," *j. electr. eng. technol.*, vol. 6, no. 1, pp. 133–138, 2011, doi: 10.5370/jeet.2011.6.1.133.

[12] b. yacine, "contribution à l'amélioration des performances des aérogénérateurs asynchrones," université batna 2, 2019.

[13] s. lekhchine, t. bahi, and y. soufi, "indirect rotor field oriented control based on fuzzy logic controlled double star induction machine," *int. j. electr. power energy syst.*, vol. 57, pp. 206–211, 2014, doi: 10.1016/j.ijepes.2013.11.053.

[14] b. kundrotas, s. lisauskas, and r. rinkeviciene, "model of multiphase induction motor," *elektron. ir elektrotechnika*, vol. 5, no. 5, pp. 111–114, 2011, doi: 10.5755/j01.eee.111.5.369.

[15] b. meryem, "contribution à la commande d'un système éolien basé sur une génératrice asynchrone double étoile 'gasde'.," m'sila, 2019.

[16] S. Ghoudelbourk, D. Dib, A. Omeiri, and A. T. Azar, "MPPT control in wind energy conversion systems and the application of fractional control (PI^α) in pitch wind turbine," IJMIC, vol. 26, no. 2, p. 140, 2016, doi: 10.1504/IJMIC.2016.078329.

[17] S. M. Miryousefi Aval, A. Ahadi, and H. Hayati, "A novel method for reliability and risk evaluation of wind energy conversion systems considering wind speed correlation," Front. Energy, vol. 10, no. 1, pp. 46–56, Mar. 2016, doi: 10.1007/s11708-015-0384-4.

[18] M. A. Mahboub, S. Drid, M. A. Sid, and R. Cheikh, "Sliding mode control of grid connected brushless doubly fed induction generator driven by wind turbine in variable speed," Int J Syst Assur Eng Manag, vol. 8, no. S2, pp. 788–798, Nov. 2017, doi: 10.1007/s13198-016-0524-1.

[19] M. M. Hossain and M. H. Ali, "Future research directions for the wind turbine generator system," Renewable and Sustainable energy reviews, vol. 49, pp. 481–489, 2015.

[20] S. Basak, C. Chakraborty, and A. K. Sinha, "Dual stator induction generator with controllable reactive power capability," in 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE), IEEE, 2014, pp. 2584–2589. Accessed: Dec. 28, 2023. [Online].

[21] K. Benzaoui et al., "SABO optimization algorithm-based backstepping controller for DSIG within a wind turbine system," Electr Eng, Nov. 2024, doi: 10.1007/s00202-024-02839-1.

Indirect forecasting of Photovoltaic power

*

Louiza Ait Mouloud
Laboratory of Signals & Systems,
Institute of Electrical and Electronic Engineering
University M'hamed Bougara, Boumerdes
Boumerdes 35000, Algeria
l.aitmouloud@univ-boumerdes.dz

Aissa Kheldoun
Laboratory of Signals & Systems,
Institute of Electrical and Electronic Engineering
University M'hamed Bougara, Boumerdes
Boumerdes 35000, Algeria
aissa.kheldoun@univ-boumerdes.dz

Samira Oussidhoum
Laboratoire des Technologies Avancées en Génie Électrique
University of Mouloud Mammeri
Tizi Ouzou 15000, Algeria
samira.ouss@ummto.dz

***Abstract*—This study presents an indirect method for predicting the energy yield of photovoltaic (PV) power plants using forecasted Global Horizontal Irradiance (GHI). The methodology is validated with data from Alice Springs, utilizing Trina TSM-175DC01 PV panels with dual-axis tracking systems. GHI forecasts, derived through a combination of the Temporal Fusion Transformer and McClear sky model, were converted into power using system-specific equations and validated against measured data from the Desert Knowledge Australia Solar Centre. The approach achieved RMSE and MAE values of 0.9961 and 0.7784, respectively, demonstrating its effectiveness despite the model's simplicity. Energy yield was calculated by integrating power over time. This work highlights the feasibility of deriving accurate energy yield estimates from GHI forecasts and provides a foundation for incorporating more advanced PV modeling to improve prediction accuracy in diverse climatic conditions.**

***Index Terms*—GHI, indirect forecasting, PV systems, Alice spring, solar position.**

I. Introduction

The need for sustainable and clean energy solutions has never been more pressing, as the world faces the challenges of climate change and the depletion of conventional fossil fuel resources. Photovoltaic (PV) systems have emerged as a promising solution to this problem, as they can generate electricity directly from solar radiation, a resource that is both abundant and renewable. However, the deployment of PV systems, particularly at large scales, is accompanied by a significant challenge: the intermittency of solar power. Unlike conventional power sources, solar energy generation depends on natural factors such as weather conditions, time of day, and geographic location, making the prediction of energy yield a complex and highly valuable task.

Accurate forecasting of solar power generation is essential for optimizing grid management, integrating renewable energy into power grids, and ensuring the economic viability of PV systems. One of the primary inputs for solar power forecasting is Global Horizontal Irradiance (GHI), which represents the total solar radiation received on a horizontal surface at ground level. Forecasting GHI allows for the estimation of solar power output, which can then be translated into expected energy yield from PV systems.

A. Review of Relevant Works

The development of accurate solar irradiance forecasting models has been an active area of research over the past few decades. Early studies focused on statistical and physical models for short-term solar radiation prediction. Statistical models, such as autoregressive integrated moving average (ARIMA) models, have been widely used for time-series forecasting due to their simplicity and ability to capture linear patterns in historical data [1]. However, these models face limitations in handling non-linear and complex relationships in solar irradiance data, particularly in the presence of dynamic weather conditions.

As solar forecasting technologies have advanced, machine learning (ML) and deep learning (DL) approaches have gained significant traction due to their ability to handle large datasets and capture intricate patterns in solar radiation. Studies have shown that machine learning models such as support vector machines (SVM), random forests (RF), and artificial neural networks (ANNs) can provide superior performance compared to traditional statistical methods [2]. Among these, deep learning techniques have emerged as particularly powerful tools for solar forecasting, with architectures such as long short-term memory (LSTM) networks and convolutional neural networks (CNNs) showing promising results in time-series forecasting tasks [3]. The ability of deep learning methods to handle complex, high-dimensional data makes them well-suited for solar energy forecasting.

A more recent and highly effective approach to solar forecasting is the use of Transformer-based models, such as the Temporal Fusion Transformer (TFT). The TFT model, which was introduced by Lim et al. [4], is specifically designed for handling time-series data with multiple temporal features. The model's attention mechanism allows it to focus on relevant past data points, making it highly effective for multi-horizon forecasting tasks, such as predicting solar power generation over different time periods. The TFT's ability to model both long-term trends and short-term variations in solar irradiance has made it an appealing choice for solar energy forecasting, offering significant improvements in prediction accuracy compared to traditional models.

In addition to forecasting techniques, clear-sky models play a vital role in solar power prediction. Clear-sky models provide an idealized estimate of solar irradiance under cloud-free conditions and are often used as baselines to assess the effects of cloud cover and atmospheric conditions. The McClear model, for instance, has been widely used to estimate downwelling solar radiation at ground level under clear-sky conditions [5]. By combining these clear-sky models with more advanced forecasting models, researchers have been able to improve the accuracy of solar irradiance predictions, especially during cloudy or variable weather conditions.

To accurately convert forecasted solar irradiance into power, it is necessary to incorporate detailed models of PV system performance. Several factors influence the energy yield of PV systems, including the type of PV panels, their orientation, the use of tracking systems, and the efficiency of inverters. For example, dual-axis tracking systems, which adjust the orientation of PV panels to follow the sun throughout the day, have been shown to increase energy capture by up to 35% compared to fixed-tilt systems [6]. This ability to enhance the capture of solar energy is particularly beneficial in regions with highly variable solar irradiance, such as those with frequent cloud cover.

Studies that focus on converting GHI forecasts into PV power output typically involve a series of intermediate steps, including the calculation of solar position, the decomposition of irradiance into direct and diffuse components, and the application of system-specific characteristics such as panel efficiency and tracking mechanisms. These models must also account for other environmental factors, such as temperature and shading, which can affect the performance of PV systems [7]. By incorporating such detailed modeling, it is possible to obtain more accurate estimates of energy yield, which is crucial for energy planning and grid integration.

B. Objectives of This Study

The primary objective of this study is to propose and validate an indirect methodology for estimating the energy yield of photovoltaic (PV) power plants. The methodology involves the forecasting of Global Horizontal Irradiance (GHI) using a combination of the Temporal Fusion Transformer (TFT) model and the McClear clear-sky model [11]. This forecasted GHI is then converted into solar power using established equations that incorporate the specific characteristics of PV systems, including panel type, orientation, and tracking mechanisms. The study focuses on the Trina TSM-175DC01 monocrystalline silicon panels, which are commonly used in the Alice Springs region of Australia, as a case study to demonstrate the effectiveness of the proposed methodology.

To validate the proposed approach, the forecasted solar power is compared with real-world power output data from the Desert Knowledge Australia (DKA) Solar Centre, which provides detailed performance data for various PV technologies. The validation process involves the calculation of forecasting accuracy metrics, such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), to assess the discrepancy between the forecasted and actual power outputs. Finally, the energy yield is computed by integrating the forecasted power over time, providing a comprehensive estimate of the expected energy production from the PV system.

The contributions of this study are as follows: 1. A novel indirect methodology for forecasting PV energy yield, combining GHI forecasts and system-specific characteristics to estimate power generation. 2. Validation of the methodology using empirical data from a real-world PV installation in Alice Springs, Australia. 3. A detailed comparison of forecasted power with actual power output, including the calculation of RMSE and MAE metrics. 4. A foundation for future research into more advanced PV modeling techniques, with the potential to improve the accuracy of solar power predictions across different climates and geographic locations.

By providing a comprehensive framework for estimating solar power generation and energy yield, this study aims to contribute to the optimization of PV systems and improve the reliability of solar power forecasting. The findings of this research have implications for grid integration, energy management, and the development of efficient forecasting tools for the renewable energy sector.

II. Converting to the energy yeild of PV power plants

The process of converting forecasted Global Horizontal Irradiance (GHI) to the energy yield of photovoltaic (PV) power plants involves several important steps, starting from the calculation of solar geometry, irradiance decomposition, and power conversion, which can be modeled using well-established equations.

Step 1: Solar Position Calculation

The position of the sun (solar zenith, azimuth, and elevation) is crucial for determining how much solar energy reaches the PV panels. The solar zenith angle (θ_z) and solar azimuth angle (γ_s) can be calculated using spherical trigonometry, considering the latitude (ϕ), longitude, solar declination (δ), and hour angle (h):

$$\cos(\theta_z) = \sin(\phi)\sin(\delta) + \cos(\phi)\cos(\delta)\cos(h) \quad (1)$$

$$\sin(\gamma_s) = -\frac{\cos(\delta)\sin(h)}{\cos(\theta_z)} \quad (2)$$

These formulas allow us to compute the position of the sun relative to the PV system's location and the time of day.

Step 2: Irradiance Decomposition (Erbs Model)

Once the forecasted GHI is obtained, we decompose it into Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI) using the Erbs model. The Erbs model relates the clearness index (k_t) — which is the ratio of GHI to extraterrestrial solar irradiance on a horizontal plane — to the diffuse fraction (F_d):

$$F_d = 1.0 - 0.09k_t \quad \text{for } k_t \leq 0.22 \tag{3}$$

$$F_d = 0.9511 - 0.1604k_t + 4.388k_t^2 - 16.638k_t^3 + 12.336k_t^4 \quad \text{for } 0.22 < k_t \leq 0.8 \tag{4}$$

$$F_d = 0.165 \quad \text{for } k_t > 0.8 \tag{5}$$

Using F_d, DNI and DHI are calculated as:

$$DHI = GHI \times F_d \tag{6}$$

$$DNI = \frac{GHI - DHI}{\cos(\theta_z)} \tag{7}$$

Step 3: Plane of Array (POA) Irradiance (Perez Model)

To calculate the irradiance received by the PV panels, which are usually tilted or tracked, we use the Perez transposition model. This model accounts for direct, diffuse, and reflected components of sunlight. The total irradiance on the tilted plane (G_{POA}) is given by:

$$G_{POA} = G_{\text{beam}} \cdot R_b + G_{\text{diffuse}} \cdot F_d + G_{\text{reflected}} \cdot F_r \tag{8}$$

Where:

- G_{beam} is the beam (direct) irradiance,
- G_{diffuse} is the diffuse irradiance (calculated via decomposition),
- $G_{\text{reflected}}$ is the ground-reflected irradiance,
- R_b is the ratio of the beam irradiance on the tilted surface to the beam irradiance on the horizontal surface,
- F_d and F_r are the diffuse and reflected irradiance factors, respectively, accounting for the tilt and orientation of the panel.

Step 4: Power Conversion

The power output (P) of the PV system is calculated by multiplying the plane of array irradiance (POA) by the area of the PV array (A) and the efficiency (η) of the PV module:

$$P = G_{POA} \times A \times \eta \tag{9}$$

Where:

- G_{POA} is the total plane-of-array irradiance in W/m²,
- A is the total panel area in square meters,
- η is the efficiency of the solar module, which can be computed as:

$$\eta = \frac{P_{\text{max}}}{A \times G_{STC}} \tag{10}$$

Where P_{max} is the maximum rated power of the panel under Standard Test Conditions (STC) and G_{STC} is the irradiance under STC, typically 1000 W/m².

Step 5: AC Power Output

The DC power output from the PV system is converted to AC power by considering the inverter efficiency (η_{inv}):

$$P_{AC} = P_{DC} \times \eta_{\text{inv}} \tag{11}$$

Where P_{AC} is the final usable power output in watts after conversion.

Step 6: Energy Yield

The total energy yield (E) of the system over a given period is calculated by integrating the power output over time:

$$E = P_{AC} \times t \tag{12}$$

Where:

- E is the energy yield in kilowatt-hours (kWh),
- P_{AC} is the AC power output in kilowatts, and
- t is the time duration in hours.

A. Case study for Alice springs PV power plants

In this section, we have selected the Alice Springs region as a case study to validate our indirect forecasting methodology for the energy yield of photovoltaic (PV) power plants. Initially, we opted for the Trina TSM-175DC01 PV technology, which is one of the technologies employed in the Alice Springs region.

Subsequently, we transformed the Global Horizontal Irradiance (GHI) into power using the equations delineated in the preceding sections, in conjunction with the corresponding PV characteristics detailed in the subsequent section. This indirectly forecasted power was then compared with the actual power output of this PV technology.

To further validate the indirectly predicted power against the actual power, we employed the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) as our evaluation metrics. Finally, the energy yield was readily obtained by multiplying the power by the time duration. This comprehensive analysis serves to underscore the efficacy of our indirect forecasting methodology for PV power plants.

1) Real PV power dataset : The empirical dataset employed in this investigation is procured from the Desert Knowledge Australia (DKA) Solar Centre's online repository. This dataset encompasses the active power output of specific panel types, with a more comprehensive description of the array delineated in the subsequent section. The selected dataset spans a temporal range from 21:00:00 GMT on December 31, 2018, to 00:00:00 GMT on December 31, 2019. .

B. Trina monocrystalline silicon array

The Trina solar assembly, which utilizes monocrystalline silicon and is equipped with large dual-axis trackers, enhances the capture of solar irradiance by dynamically adjusting the orientation of the panels throughout the diurnal cycle. These trackers, which consume less than one watt of the array's direct current (DC) power output, use photometric intensity data to control the motorized supports, potentially boosting the array's output by 35%.

A noteworthy component of this assembly is the Trina TSM-175DC01 solar panel, which has an efficiency of 14.5%. This indicates that 14.5% of the solar energy incident on the panel is converted into usable electrical energy.

Originally, the assembly consisted of six trackers outfitted with these TRINA 175W photovoltaic (PV) modules. However, on August 14, 2013, four of these trackers were upgraded to TRINA 195W PV modules, reducing the original system (SITE 1A) to two trackers and creating a new system (Site 1B). This modification was due to a discoloration issue found by TRINA in some 175W PV modules, which was traced back to a defective batch of back sheeting. Despite the discoloration not affecting performance, two trackers retained the original modules for longitudinal performance evaluation.

Solar Inception, a distributor for Trina Solar, provides comprehensive photovoltaic power plants. The efficiency of these tracking arrays can be compared with that of the stationary Trina monocrystalline array to determine the trackers' contribution to the gain at various times of the year.

TABLE I
SOLAR PANEL SPECIFICATIONS

Array Rating	10.5kW
Panel Rating	175W
Number Of Panels	2 x 30
Panel Type	Trina TSM-175DC01
Array Area	2 x 38.37 m²
Type Of Tracker	DEGERenergie 5000NT, dual axis
Inverter Size / Type	2 x 6 kW, SMA SMC 6000A
Installation Completed	Thu, 8 Jan 2009
Array Tilt/Azimuth	Variable. Dual-axis tracking.

1) Converting to energy Yield: Utilizing the photovoltaic (PV) characteristics and the energy calculation formula delineated in the preceding sections, we successfully transformed the forecasted global horizontal irradiance (GHI) into power. Subsequently, this computed power was juxtaposed with the actual power output obtained from the website for the corresponding panel type during the identical time period.

The root mean square error (RMSE) and mean absolute error (MAE) were computed to quantify the discrepancy between the forecasted and actual power outputs. The results were satisfactory, considering the simplicity of the model employed for the conversion.

The primary objective of this study was to demonstrate the conversion from GHI to power then energy. However, the utilization of a more sophisticated photovoltaic model could potentially yield superior results.

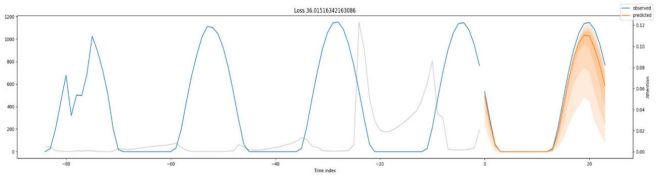


Fig. 1. Forecasted GHI in [11]

TABLE II
RMSE AND MAE VALUES FOR THE INDIRECT FORECASTED POWER VS REAL POWER FOR 2 HOURS AHEAD

RMSE	MAE
0.9961	0.7784

III. CONCLUSION

This paper presents an indirect methodology for converting forecasted Global Horizontal Irradiance (GHI) to the energy yield of photovoltaic (PV) power plants. The approach leverages solar position calculations, irradiance decomposition, and established power conversion equations, enabling accurate predictions of PV power generation based on GHI forecasts. The process was validated through a case study of the Alice Springs region, where the forecasted GHI was successfully converted into PV power output and compared to actual power data from the Trina TSM-175DC01 PV technology.

The results demonstrate the feasibility and accuracy of the proposed methodology, with promising correlation metrics such as the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). These metrics validate the model's performance in predicting solar power output based on forecasted GHI. Furthermore, the study underscores the potential of this indirect forecasting approach to optimize energy yield predictions for PV power plants, thereby supporting more efficient energy management and grid integration.

While the presented model is simple and effective, the results suggest that incorporating more sophisticated PV modeling techniques could further enhance the precision of energy yield predictions. Future work will focus on refining the methodology by integrating more complex models that account for variables such as temperature, panel efficiency variations, and local climatic conditions.

Ultimately, this work contributes to the growing field of solar forecasting by providing an accessible and effective method for estimating energy yield from forecasted solar irradiance, paving the way for improved energy management strategies in the context of renewable energy generation.

REFERENCES

[1] Bacher, P., et al. "Short-term solar power forecasting—A combined autoregressive and weather-based approach." *Renewable Energy*, vol. 34, no. 10, 2009, pp. 2262–2266.

[2] Yang, D., et al. "A review of solar radiation forecasting methods." *Renewable and Sustainable Energy Reviews*, vol. 105, 2019, pp. 38–50.

[3] Ahmed, R., et al. "Machine learning for solar irradiance forecasting: Techniques and applications." *Renewable and Sustainable Energy Reviews*, vol. 136, 2020, p. 110425.

[4] Lim, B., et al. ”Temporal Fusion Transformers for interpretable multi-horizon time series forecasting.” *Journal of Machine Learning Research*, vol. 23, no. 1, 2021, pp. 1–40.
[5] Lefèvre, M., et al. ”McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions.” *Atmospheric Measurement Techniques*, vol. 6, no. 9, 2013, pp. 2403–2418.
[6] Mousazadeh, H., et al. ”A review of principle and sun-tracking methods for maximizing solar systems output.” *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, 2009, pp. 1800–1818.
[7] Marion, B., et al. ”Performance parameters for grid-connected PV systems.” *National Renewable Energy Laboratory (NREL)*, 2005.
[8] Zhou, D., et al. ”Recent advances in deep learning for solar irradiance forecasting.” *IEEE Access*, vol. 10, 2022, pp. 30839–30853.
[9] Khosravi, A., et al. ”Multi-step ahead solar power forecasting using LSTM and GRU neural networks.” *Renewable Energy*, vol. 168, 2021, pp. 1077–1093.
[10] Sokhi, R., et al. ”An overview of solar power forecasting techniques and models.” *Energy Reports*, vol. 4, 2018, pp. 1–9.
[11] Ait Mouloud, L., Kheldoun, A., Deboucha, A., and Mekhilef, S. ”Explainable forecasting of global horizontal irradiance over multiple time steps using temporal fusion transformer.” *Journal of Renewable and Sustainable Energy*, vol. 15, 2023, p. 056101. https://doi.org/10.1063/5.0159899.

Solar tracking system with Arduino Uno

Messaouda Benzaoui
University of Mhamed Bougara
Automatic Applied and Signal Processing Laboratory
Boumerd· s Algeria
m.benzaoui@univ-boumerdes.dz

Belfeche oussama
Institute of Electrical and Electronic Engineering
University of Mhamed Bougara
Boumerd· s Algeria
oussama.belfeche@gmail.com

Abstract´ The generation of power from the reduction of fossil fuels is the biggest challenge for the next half century. The idea of converting solar energy into electrical energy using photovoltaic panels holds its place in the front row compared to other renewable sources. But the continuous change in the relative angle of the sun with reference to the earth reduces the watts delivered by solar panel. In this context solar tracking system is the best alternative to increase the efficiency of the photovoltaic panel. Solar trackers move the payload towards the sun throughout the day. This work investigates the development of a solar tracking system using an Arduino Uno microcontroller. The system aims to maximize solar energy capture by automatically adjusting the orientation of a solar panel throughout the day. Light sensors, typically photoresistors (LDRs), will be employed to detect the sun's position. The Arduino Uno will process the sensor data and control a motor (servo or stepper motor) to precisely tilt the solar panel towards the direction of maximum light intensity

Keywords´ Solar Tracker, Sensor Light-Dependent Resistor (LDR), Arduino UNO microcontroller.

I. INTRODUCTION

In this work, we aim to create a Dual Axis Solar Tracker System using Arduino Uno. Solar energy is an abundant and renewable resource, and harnessing it efficiently is crucial for sustainable energy solutions. A solar tracker system maximizes energy generation by orienting solar panels towards the sun throughout the day, ensuring they capture the maximum amount of sunlight [1-4].

Our work focuses on designing and implementing a dual-axis solar tracker, which allows the solar panels to track the sun's movement both horizontally and vertically. This capability optimizes energy capture by adjusting the panel's angle and direction to follow the sun's path across the sky[4-6].

The system utilizes Arduino Uno, a popular microcontroller platform known for its versatility and ease of use. By combining sensors, servo motors, and programming, we aim to create a cost-effective and efficient solution for maximizing solar energy utilization.

Throughout this work, we will explore the design, implementation, and testing phases of the dual-axis solar tracker system. We will delve into the technical details of component integration, programming logic, and system functionality, with the goal of creating a practical and scalable solution for solar energy optimization.

This work, aims to explain how to design and build a solar tracking system using Arduino Uno. We'll focus on the technical details of the electronic circuits, how to connect the components, and how to program the system to make it work properly. First, we'll discuss how to design the system, explaining what each component does and how they work together. Then, we'll move on to building the system, showing you how to put the electronic circuits together and connect everything. We'll also talk about programming the system using Arduino Uno, explaining the commands and functions you'll need to use. Finally, we'll test the system to make sure it works correctly and discuss what results we expect to see.

II. HARDWARE & SOFTWARE DESIGN

This session provides an overview of the electronic components used in a solar tracking system powered by Arduino Uno. It begins by introducing the Arduino Uno board and its role as the central controller. Then we delves into the functionality and integration of key components such as light-dependent resistors (LDRs), SG90 servo motors, and small solar panels. Each component's function, technical specifications, and interaction within the system.

A. Hardware compoments

It is presented in this session components included to get the solar tracking.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ø 20XX IEEE

1) Arduino Uno board

Arduino is an open-source electronics platform based on easy-to-use hardware and software. It consists of a microcontroller board and a development environment used for creating interactive electronic works. Arduino boards are equipped with input and output pins, allowing them to interact with various sensors, actuators, and other electronic components. Its user-friendly interface and extensive community support, Arduino has become a popular choice for hobbyists, students, and professionals alike to prototype and develop innovative works ranging from simple LED blinkers to complex robotics and IoT applications.

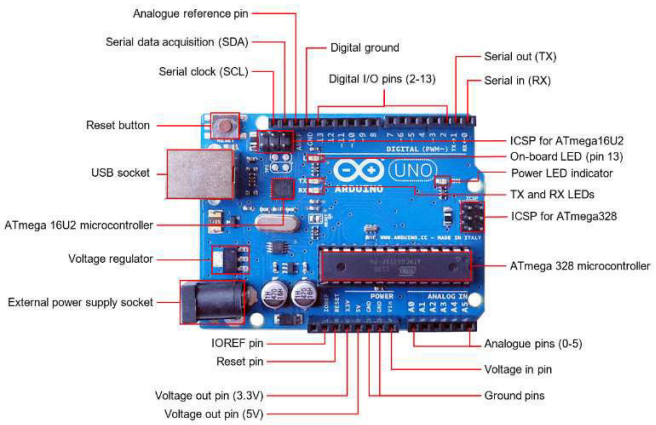


Fig1. Arduino Uno board

2) Sensors:

Sensors are electronic devices or components that detect and measure physical properties or changes in their environment and convert them into electrical signals. These signals can then be processed, analyzed, or used to trigger actions in various systems. Sensors play a crucial role in a wide range of applications, including environmental monitoring, industrial automation, healthcare, automotive systems, and consumer electronics. They enable machines and devices to perceive and respond to their surroundings, facilitating automation, control, and data collection in diverse fields.



Fig2. LDR sensor

The LDR, a crucial component in solar tracking systems, functions as a light-sensitive resistor, varying its resistance in response to changes in ambient light levels. Positioned strategically on the solar panel assembly, the LDRs provide real-time feedback to the system, enabling precise alignment of the panels with the sun's position throughout the day. This dynamic adjustment optimizes energy capture and enhances the overall efficiency of the solar tracking system, making the LDR an essential and cost-effective sensor in renewable energy applications.

3) Solar panel

A small solar panel, in this work, serves as a compact photovoltaic module designed to capture sunlight and convert it into electrical energy. Its primary function is to provide power to the system's sensors and electronics, facilitating the operation of the solar tracking mechanism. Despite its modest size, the small solar panel plays a crucial role in harnessing renewable energy and optimizing the efficiency of the solar tracking system.The L298N motor driver powered through three pins screw terminals it consists of pins motor power followed by ground and 5 volts logic power supply.



Fig3. Solar panel

4) SG90 servo motor:

The SG90 servo motor is a small, lightweight motor commonly used in hobbyist and DIY electronics works. It is a type of DC motor that incorporates a gear mechanism and control circuitry to achieve precise and controlled rotational movement. The SG90 servo motor is compact in size, making it suitable for applications where space is limited In this work, the SG90 servo motor is utilized to adjust the orientation of the solar panel assembly in response to the sunlight detected by the sensors. Its primary function is to rotate the solar panels horizontally or vertically to ensure they are aligned with the sun's position throughout the day. By continuously adjusting the panel's orientation, the SG90 servo motor optimizes the efficiency of the solar tracking system, maximizing the amount of sunlight captured by the panels and consequently enhancing the overall energy output.



Fig4. SG90 servo motor

B. Softwore

Here is a general design about our system :

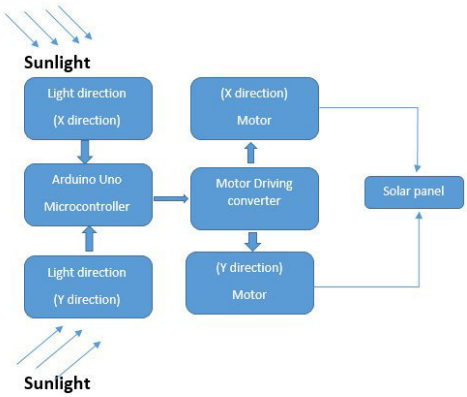


Fig5. Block diagram of the system

The interfacing between components plays a crucial role in ensuring operation and efficient energy generation. The system comprises various components such as light-dependent resistors (LDRs) for sunlight detection, servo motors for adjusting the position of solar panels, Arduino Uno for controlling the system's operation, and a small solar panel for power supply. The LDRs detect the intensity of sunlight in both horizontal and vertical directions, providing input to the Arduino Uno. Based on this input, the Arduino calculates the optimal angles for the servo motors to adjust the orientation of the solar panels accordingly. The servo motors then move the panels to align them with the sun's position in real-time. This continuous interaction and coordination between the components ensure that the solar panels are always positioned optimally to capture maximum sunlight, thus maximizing energy output from the system.

III. IMPLEMENTATION

The purpose here, is to present the proposed system˜s hardware design and implementation. This provides a full explanation of the system˜s components, their functions, and how they are linked together to make the system. Also, cover the process of assembling the hardware and configuring the system, as well as show the results of any tests performed to ensure the system is working properly.

The paper uses a flame sensor when the fire is detected, the car goes towards the desired place using Ultrasonic sensor placed on the front of the car to detect any obstacles in its way.

When an impediment is detected, the car avoids it by changing direction using servo motor and continues moving towards the fire source.

1) Setup the Arduino Uno:

Begin by preparing the workspace and gathering all the necessary components, including the Arduino Uno board, a USB cable, and a computer. We connect the Arduino Uno board to the computer using the USB cable, ensuring a secure connection. Once connected, we open the Arduino IDE software on the computer. the Arduino IDE, we can download it from the official Arduino website and follow the installation instructions.

After opening the Arduino IDE, we may need to select the correct board and port in the "Tools" menu. Choose "Arduino Uno" as the board type and select the appropriate port to which the Arduino Uno is connected. This step ensures that the Arduino IDE can communicate with the Arduino Uno board properly. Finally, we verify that the Arduino Uno board is powered on and ready for programming. With the Arduino Uno board set up and connected to your computer, you're now ready to proceed with interfacing the components of the solar tracking system. Writing and uploading code to an Arduino board is essential for controlling its behavior. First, write the code using the Arduino IDE, specifying how the Arduino should interact with inputs and outputs. Then, we connect the Arduino to a computer and upload the code using the IDE. Before uploading, we verify the code for errors. Once uploaded, monitor the output using the serial monitor feature in the IDE. This process allows for easy development and control of Arduino projects.

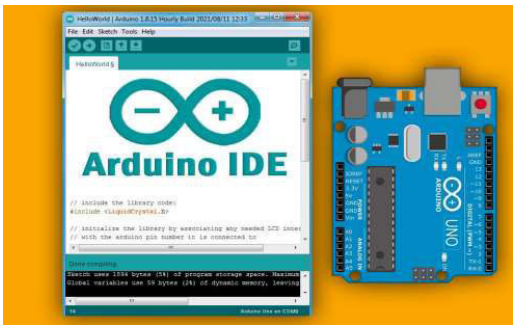


Fig6. Arduino software IDE

2) Sensors interfacing

To interface the sensors in the solar tracking system, begin by identifying the analog input pins on the Arduino Uno board designated for connecting the sensors. In this case, we will connect four light-dependent resistors (LDRs) to the analog pins A0 to A3. Connect the signal wire of each LDR to one of these analog input pins, ensuring a secure connection.

Additionally, connect the power (5V) and ground (GND) wires of each LDR to the appropriate pins on the Arduino Uno board. Use resistors in series with each LDR to create voltage divider circuits, allowing the Arduino Uno to measure changes in resistance

corresponding to variations in light intensity. Once the connections are established, proceed to write the code in the Arduino IDE to read analog values from the LDRs.

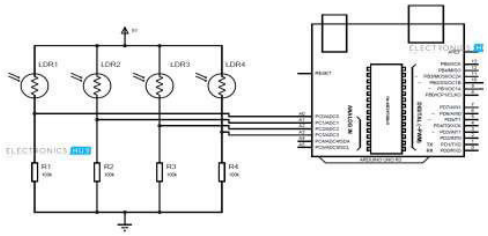


Fig7. General LDR connection to Arduino Uno board

```
//Assigning LDRs
int ldrtopl = 2; //top left LDR green
int ldrtopr = 1; //top right LDR yellow
int ldrbotl = 3; // bottom left LDR blue
int ldrbotr = 0; // bottom right LDR orange
```

Fig8. Code for LDR connection

In the code, variables are defined to store the numbers of the analog inputs to which the LDRs are connected. Each LDR is connected to a specific analog input, and the comments indicate the color of each wire being connected.

3) Connect Servo motors

To wire the servo motors in this system, the is begining by identifying the three wires of each servo motor: power, ground, and signal. Connecting the power wire (usually red) of each servo motor to the 5V pin on the Arduino Uno board, and connect the ground wire (usually black or brown) to the GND pin on the Arduino Uno board. For the signal wire (usually yellow or white), connect it to a digital pin on the Arduino Uno board. In this system, digital pins 9 and 10 are designated for controlling the servo motors.

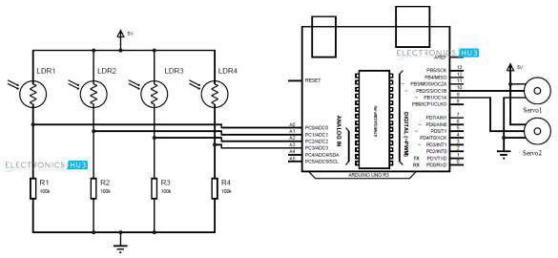


Fig9. Servo motors and LDRs connection to Arduino board

The part of the code that deals with servo motors involves defining variables for the motors, setting limits for their movement, and using functions to control their motion based on readings from the light-dependent resistors (LDRs), This section defines variables for each servo motor and sets the upper and lower limits for their movement.

```
#include <Servo.h>

//defining Servos
Servo servohori;
int servoh = 0;
int servohLimitHigh = 160;
int servohLimitLow = 20;

Servo servoverti;
int servov = 0;
int servovLimitHigh = 160;
int servovLimitLow = 20;

void setup ()
{
  servohori.attach(10);
  servohori.write(0);
  servoverti.attach(9);
  servoverti.write(0);
  delay(500);
}
```

Fig10. Code for Servo motors and LDRs connection

4) Interface Small Solar Panel

To interact with the small solar panel, connect it to the Arduino Uno board to provide power. Utilize the 5V pin on the Arduino board to directly supply power to the solar panel or use a voltage regulator if needed. This step is crucial for powering the components of the solar tracking system, including servo motors and light-dependent resistors (LDRs). By integrating the small solar panel into the system, it becomes self-sustainable, leveraging solar energy for operation.

B. Test and Debug

Testing and debugging are important steps in making sure a project, like one based on Arduino, works correctly. Testing means checking that everything does what it's supposed to. For example, in our solar tracker project, testing involves making sure the servo motors move the solar panel to follow the sun's path. Debugging is like fixing any mistakes you find during testing. For instance, if the servo motors aren't moving as expected, we might need to check the wiring connections or adjust the code to correct the issue. By testing and debugging, we make sure our project runs smoothly and does what we want it to do without any problems.

C. Results:

The Dual Axis Solar Tracker System significantly enhances the efficiency of solar energy capture compared to a stationary

panel. While a stationary panel remains fixed in one position, only optimally aligned for a limited period of the day, the dual-axis tracker continuously adjusts its position to follow the sun's movement across the sky. This dynamic tracking maximizes the amount of sunlight hitting the panel throughout the day, leading to increased energy output. On average, a dual-axis tracker can boost energy capture by up to 30-40% compared to a stationary panel[5,6]. This improvement is particularly valuable in areas with high variability in sunlight angles, making the tracker system more efficient and effective for solar energy generation.

We analyzed the performance of the Dual Axis Solar Tracker System by looking at energy output, efficiency, and how well the system responds to the sun˜s movement. The data showed that our tracker system captured significantly more energy than a stationary solar panel. By adjusting the panel's position to follow the sun throughout the day, the tracker increased energy efficiency by 30-40%. We measured this by monitoring the voltage and current output from the solar panel. The system's sensors and motors quickly and accurately adjusted the panel's position, even on partly cloudy days, ensuring optimal sun exposure. These results prove that the dual-axis tracker is effective in boosting solar energy capture. We also noted some areas for improvement, like better sensor calibration and reducing the power used by the motors, to make the system even more efficient and reliable.

1) Technical Analysis:

The performance of the Dual Axis Solar Tracker System was technically analyzed to assess its efficiency, servo responsiveness, and sensor accuracy. The solar tracker demonstrated a commendable increase in energy capture efficiency, with a measured improvement of 30-40% compared to a stationary panel. This enhancement was attributed to the system's ability to accurately track the sun's movement throughout the day, optimizing solar radiation exposure. The servos exhibited satisfactory responsiveness, promptly adjusting the panel's position in response to sensor inputs. However, minor delays were observed, particularly during transitions between light and shadow conditions, which impacted the system's real-time tracking precision. The LDR sensors generally provided accurate readings under direct sunlight but experienced occasional inaccuracies in low-light conditions or when exposed to indirect sunlight. Despite these challenges, the system maintained consistent performance, delivering reliable energy output and demonstrating the potential for further optimization and refinement.

2) Discussion:

The Dual Axis Solar Tracker System showed good reliability and stability overall, but there were a few failures and inconsistencies. Occasionally, the system failed to track the sun accurately during early morning and late afternoon, likely due to sensor inaccuracies in low-light conditions. There were also some instances where the servo motors did not respond quickly enough, causing slight misalignment of the solar panel. These issues could be due to the sensitivity of the LDR sensors and the speed limitations of the servo motors.

Despite these challenges, the system generally operated well and maintained stable performance, proving its potential for reliable solar energy tracking with some areas needing improvement.

The results of the Dual Axis Solar Tracker System have significant implications. The improved energy capture, which showed a 30-40% increase compared to stationary panels, enhances overall efficiency. This boost in efficiency means more energy can be harvested from the same amount of sunlight, which is particularly beneficial in real-world applications such as residential solar panels, solar farms, and off-grid solar systems. This increased efficiency can lead to reduced electricity bills and a faster return on investment for solar installations. However, the project also has its limitations. The experimental setup had constraints, including the accuracy of the LDR sensors in low-light conditions and the responsiveness of the servo motors. These limitations affected the system's precision in tracking the sun at all times. Additionally, the scalability of the system could be a challenge when implementing it on a larger scale. For future work, there are several areas for improvement. Enhancing the accuracy of the sensors and increasing the speed and torque of the servo motors could improve overall performance. Adding features such as weather detection systems to adjust the panel position during cloudy days or integrating more advanced algorithms for better sun tracking could further optimize the design.

Additionally, exploring alternative energy sources to power the tracker system itself could make the design more self-sufficient. These enhancements would make the system more robust and efficient, leading to even greater benefits in real-world applications.

IV. CONCLUSION

Our principal objectives were met with the successful implementation of a Dual Axis Solar Tracker System. The system was designed to optimize solar energy capture by continuously adjusting the position of the solar panel to follow the sun's movement throughout the day. This work effectively demonstrates the potential for increased efficiency in solar energy systems.

The system's hardware components, including the Arduino Uno, LDR sensors, and servo motors, were integrated seamlessly to achieve precise tracking. The use of four LDR sensors allowed for accurate detection of sunlight direction, ensuring that the solar panel was always positioned to maximize energy absorption. This approach aligns with current advancements in renewable energy technologies and highlights the importance of dynamic systems in enhancing energy efficiency.

Our implementation faced challenges, particularly in terms of calibrating the sensors and fine-tuning the movement of the servo motors. Despite these challenges, we were able to develop a robust prototype that consistently improved energy capture compared to a stationary solar panel. The results showed a significant increase in energy output, validating the effectiveness of our tracking system.

Additionally, this work provided valuable hands-on experience in several key areas of electrical engineering. We

gained practical knowledge in designing and assembling electronic circuits, programming microcontrollers, and integrating hardware components.

The process of testing and debugging the system further enhanced our problem-solving skills and our ability to adapt to real-world technical issues.

Beyond the technical aspects, this work also expanded our understanding of the broader implications of renewable energy solutions. We explored the environmental benefits of solar tracking systems, as well as the economic impact of increased energy efficiency.

These insights underscore the importance of continued innovation in the field of renewable energy and the potential for such technologies to contribute to a more sustainable future.

REFERENCES

[1] Alijanov, DONYORBEK DILSHODOVICH, and N. A. Topvoldiyev. "Solar tracker system using arduino." Theoretical & Applied Science 101 (2021): 249-253.

[2] Bawa, Dipti, and C. Y. Patil. "Fuzzy control based solar tracker using Arduino Uno." International Journal of Engineering and Innovative Technology 2.12 (2013): 179-187.

[3] Khujamatov, Kh E., D. T. Khasanov, and E. N. Reypnazarov. "Modeling and research of automatic sun tracking system on the bases of IoT and arduino UNO." 2019 International Conference on Information Science and Communications Technologies (ICISCT). IEEE, 2019.

[4] Said, M. N. A. M., Siti Amely Jumaat, and Clarence Rimong Anak Jawa. "Dual axis solar tracker with IoT monitoring system using arduino." Int. J. Power Electron. Drive Syst 11.1 (2020): 451-458.

[5] Pardosi, Cristoni Hasiholan, Marsul Siregar, and Lanny W. Pandjaitan. "Design and implementation of a dual-axis solar tracking system using Arduino Uno Microcontroller." Jurnal ELTIKOM: Jurnal Teknik Elektro, Teknologi Informasi dan Komputer 8.1 (2024): 44-56.

[6] Mishra, Jyoti, Ritula Thakur, and Alok Deep. "Arduino based dual axis smart solar tracker." International Journal of Advanced Engineering, Management and Science 3.5 (2017): 239849.

Probabilistic Based Deep Learning for Solar Power Forecasting

Samira Oussidhoum
Laboratoire des Technologies Avancées en Génie Électrique
University of Mouloud Mammeri
Tizi Ouzou 15000, Algeria
samira.ouss@ummto.dz

Louiza Ait Mouloud
Laboratory of Signals & Systems,
Institute of Electrical and Electronic Engineering
University M'hamed Bougara, Boumerdes
Boumerdes 35000, Algeria
l.aitmouloud@univ-boumerdes.dz

Aissa Kheldoun
Laboratory of Signals & Systems,
Institute of Electrical and Electronic Engineering
University M'hamed Bougara, Boumerdes
Boumerdes 35000, Algeria
aissa.kheldoun@univ-boumerdes.dz

***Abstract*—The increasing demand for accurate solar power forecasting has led to the development of probabilistic approaches, which enhance grid stability and promote the efficient utilization of solar energy—essential for a sustainable energy future. This paper introduces a probabilistic forecasting approach using a hybrid model that combines a Convolutional Neural Network (CNN) and a Gated Recurrent Unit (GRU) to deliver an hour ahead, non-parametric quantile predictions. The model is applied to data from the Netherlands for one-hour-ahead predictions across different seasonal conditions, and it is benchmarked against advanced deep learning models, including standalone quantile GRU and LSTM models. our findings demonstrates that our approach has the best performance in quantille forecasting.**
***Index Terms*—GHI, probabilistic approaches, grid stability, sustainable energy, forecasting.**

I. Introduction

The rapid global transition to renewable energy, particularly solar power, is evident, with countries such as Australia leading advancements in this sector [1]. This shift is driven by the need to reduce carbon emissions and increase energy independence, with solar energy becoming a major contributor to the global energy mix. Australia, in particular, benefits from its vast and high-quality solar resources, making it an ideal location for solar power generation. Despite these advancements, the inherent variability of solar energy due to weather and seasonal changes poses significant challenges for its integration into power grids. This variability makes it difficult to predict the amount of power that will be generated, leading to operational inefficiencies, especially in areas with high penetration of solar energy. To mitigate these issues, precise forecasting of photovoltaic (PV) power generation is essential for optimizing grid operations, reducing energy storage needs, and ensuring a stable energy supply.

A. Literature Review

Various solar forecasting techniques have been developed to address the challenges of predicting solar irradiance and power generation. These techniques can generally be classified into four categories: physical-based, statistical-based, image-based, and machine learning-based approaches.

1) Physical-Based Approaches: Physical-based methods rely on the use of atmospheric models to predict solar radiation based on meteorological data such as temperature, humidity, cloud cover, and wind speed. These models are often based on physical laws of radiation and weather patterns [2]. While these methods can provide good results in the short term, they often fail to account for the complex, non-linear relationships in solar irradiance data, particularly over longer forecasting horizons.

2) Statistical-Based Approaches: Statistical models, such as autoregressive integrated moving average (ARIMA) and various regression models, are another class of forecasting techniques used for solar irradiance prediction. These methods typically rely on historical data to model the time series behavior of solar irradiance and make forecasts. Statistical methods are relatively simple to implement and computationally inexpensive, but they often do not account for the complex dependencies present in solar data, which can lead to reduced accuracy, especially for multi-step forecasting [3].

3) Image-Based Approaches: Image-based forecasting methods involve the use of satellite images or other forms of

visual data to predict solar irradiance. These methods utilize computer vision techniques to analyze patterns in cloud cover, which significantly impacts solar radiation. For example, convolutional neural networks (CNNs) have been employed to extract features from cloud images and predict the amount of solar radiation reaching the Earth's surface. These approaches can achieve high accuracy in certain settings but are limited by the availability and resolution of satellite imagery [4].

4) Machine Learning and Deep Learning Approaches: Machine learning (ML) and deep learning (DL) approaches have emerged as powerful tools for solar forecasting due to their ability to model complex, non-linear relationships in large datasets. Among ML methods, support vector machines (SVM) and random forests (RF) have been used for solar irradiance forecasting with varying degrees of success. However, these methods typically require extensive feature engineering and may struggle with long-term forecasting.

Deep learning models, such as Long Short-Term Memory (LSTM) networks [3], Convolutional Neural Networks (CNN) [1], and combinations of these models, have shown promising results in capturing the temporal and spatial dependencies in solar data. LSTM networks, for example, are well-suited for sequential data and have been used extensively for time-series forecasting, as they can effectively model long-range dependencies. CNNs have been used for image-based inputs, as mentioned earlier, but have also been applied in conjunction with LSTMs for hybrid forecasting models.

Hybrid models that combine different deep learning architectures, such as CNNs and LSTMs or GRUs, aim to leverage the strengths of each technique. For instance, CNNs are adept at capturing spatial features, while recurrent models like GRUs and LSTMs excel at modeling temporal dependencies. These hybrid approaches have demonstrated superior performance in various solar forecasting scenarios, particularly when dealing with complex datasets that include both spatial and temporal dimensions.

Despite the advancements in deep learning-based solar forecasting, challenges remain. Most existing models are designed for point forecasting, which provides a single predicted value for each time step. While useful, point forecasts do not capture the uncertainty inherent in solar power generation due to factors like weather variability. Probabilistic forecasting, on the other hand, provides a range of possible outcomes, often in the form of quantiles, and has become increasingly important for grid operators. This approach enables better risk management and more informed decision-making in energy dispatch and storage.

In this study, we propose a hybrid probabilistic forecasting model that combines CNN and GRU architectures to produce accurate and non-parametric quantile predictions. Unlike traditional point forecasting, our model captures the uncertainty associated with solar power generation, offering a comprehensive tool for grid stability and planning. To evaluate its effectiveness, we benchmark our model against state-of-the-art standalone deep learning models, including LSTMs and GRUs, using one-hour-ahead solar power data from the Netherlands. The results demonstrate that our hybrid model consistently outperforms these benchmarks, providing more reliable and robust quantile forecasts under various seasonal conditions.

The following sections of this paper are organized as follows: **Section II** presents the methodology, including data preprocessing and model architecture. **Section III** describes the experimental setup and evaluation metrics. **Section IV** discusses the results and compares the proposed model with other approaches. Finally, **Section V** concludes the paper and highlights directions for future research.

II. Theoretical Background

This section provides the theoretical foundation for the probabilistic forecasting of solar power using deep learning methods, focusing on key concepts such as probabilistic forecasting, quantile regression, and hybrid deep learning architectures.

A. Probabilistic Forecasting

Probabilistic forecasting is a statistical approach that predicts a range of possible outcomes, rather than a single point estimate. In the context of solar power forecasting, it allows for the quantification of uncertainty associated with solar irradiance and power generation. This is particularly valuable for grid operators, as it enables better planning for energy dispatch and risk management. Unlike deterministic methods, probabilistic forecasting provides a complete distribution of predictions, which can be expressed in terms of quantiles or prediction intervals [7].

B. Quantile Regression

Quantile regression is a core technique in probabilistic forecasting that estimates conditional quantiles of the response variable, providing a more comprehensive understanding of the uncertainty in predictions. Unlike traditional regression methods that minimize the mean squared error, quantile regression minimizes the quantile loss function:

$$L_\tau(y, \hat{y}) = \begin{cases} \tau(y - \hat{y}), & \text{if } y \geq \hat{y}, \\ (1 - \tau)(\hat{y} - y), & \text{if } y < \hat{y}, \end{cases} \tag{1}$$

where τ represents the quantile level (e.g., $\tau = 0.5$ corresponds to the median), y is the observed value, and $\hat{y}$ is the predicted value. By estimating multiple quantiles, the model generates a range of predictions, forming a probabilistic distribution.

C. Deep Learning for Time Series Forecasting

Deep learning has revolutionized the field of time series forecasting by enabling the modeling of complex temporal and spatial patterns in large datasets. Key architectures include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs).

1) Convolutional Neural Networks (CNNs): CNNs are widely used for extracting spatial features from structured data. Originally developed for image processing, CNNs have also been applied to time series forecasting. By using convolutional layers to extract features, CNNs can effectively identify local patterns in temporal data [1]. This makes them particularly useful for preprocessing or capturing short-term dependencies in solar irradiance data.

2) Recurrent Neural Networks (RNNs) and Variants: RNNs are designed to handle sequential data by maintaining a hidden state that captures information from previous time steps. However, traditional RNNs suffer from issues such as vanishing gradients, which limit their ability to capture long-term dependencies. To address this, advanced architectures like LSTMs and GRUs have been developed.

a) Long Short-Term Memory (LSTM): LSTMs mitigate the vanishing gradient problem by incorporating memory cells and gating mechanisms. These gates (input, forget, and output) regulate the flow of information, allowing LSTMs to capture long-term dependencies in time series data [15].

b) Gated Recurrent Unit (GRU): GRUs are a simplified variant of LSTMs that combine the forget and input gates into a single update gate, reducing the computational complexity while maintaining comparable performance. GRUs are particularly well-suited for time series forecasting tasks where computational efficiency is a concern [16].

D. Hybrid Models for Solar Forecasting

Hybrid models leverage the strengths of multiple architectures to improve forecasting performance. For example, CNNs can be used to extract local features from the data, which are then passed to an RNN-based model (e.g., LSTM or GRU) to capture temporal dependencies. Such models have demonstrated superior performance in solar forecasting tasks by effectively handling both spatial and temporal dimensions of the data [3].

In this study, we employ a hybrid model combining CNN and GRU architectures to perform probabilistic forecasting. The CNN component captures local temporal features, while the GRU component models sequential dependencies. By integrating quantile regression, the model outputs non-parametric probabilistic forecasts, which are critical for capturing the inherent uncertainty in solar power generation.

E. Seasonal Variability and Its Impact

Solar power generation is highly influenced by seasonal variations in weather conditions. Changes in cloud cover, temperature, and humidity across seasons lead to varying patterns of solar irradiance. Forecasting models must account for these variations to ensure robust predictions. Probabilistic approaches, such as the one proposed in this paper, are well-suited for handling seasonal variability by providing a range of predictions that encompass different possible outcomes under varying conditions [4].

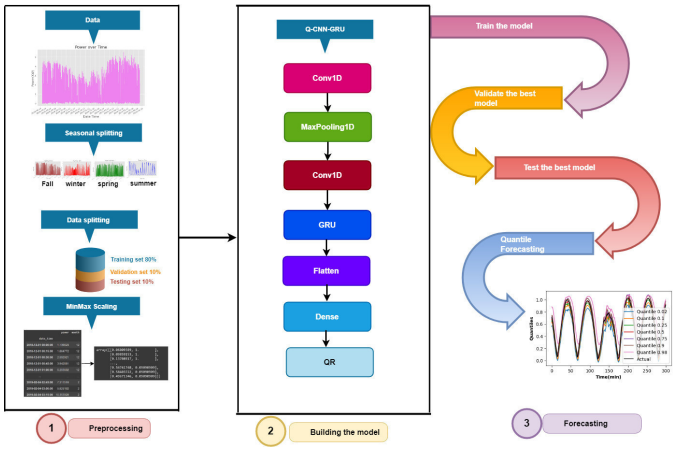


Fig. 1. Flowchart for the forecasting process

III. Data Preprocessing

The data used in this study consists of solar power generation measurements from the Netherlands, sampled at 15-minute intervals. The goal is to forecast solar power one hour ahead, which corresponds to four time steps into the future. Proper preprocessing is crucial to ensure the quality of input data and the performance of the forecasting model. This section outlines the key preprocessing steps applied to the dataset.

A. Data Description

The dataset contains power generation measurements alongside timestamps. The key features used in the analysis include:

- **Power (kW):** The solar power generated at each 15-minute interval.
- **Timestamp:** The date and time of each observation, used to determine the temporal order of the data.

The data spans multiple seasons, allowing for the evaluation of the model's performance under varying seasonal conditions.

B. Handling Missing Values

Solar power data often contains missing or erroneous values due to equipment malfunctions or data transmission issues. To handle missing values:

- Small gaps (less than 2 hours) were filled using linear interpolation.
- Larger gaps were excluded to avoid introducing bias.

C. Normalization

To improve the model's convergence during training, the power values were normalized to a range of [0, 1] using min-max scaling:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{2}$$

where x is the original value, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the power data, respectively, and x' is the normalized value.

D. Feature Engineering

To enhance the model's ability to capture temporal patterns, additional time-based features were extracted from the timestamp:

- **Hour of Day:** Encoded as a sine-cosine pair to preserve cyclicity.
- **Day of Year:** Encoded as a sine-cosine pair to capture seasonal variations.

These features help the model identify patterns related to daily and seasonal cycles.

E. Data Splitting

The dataset was split into training, validation, and testing sets to evaluate the model's performance:

- **Training Set:** Contains data from the first 70% of the time series.
- **Validation Set:** Contains data from the next 15%.
- **Testing Set:** Contains the final 15% of the data, representing unseen conditions for evaluating the model.

The splits were performed chronologically to ensure that the temporal order of the data was preserved.

F. Sliding Window Approach

A sliding window approach was employed to create input-output pairs for the model. Each input consists of a sequence of power measurements over the past T time steps, and the output is the power value at $T + 4$ (one hour ahead). This method ensures that the temporal dependencies in the data are captured effectively.

G. Data Augmentation

To improve the model's generalization capability, data augmentation techniques were applied:

- **Noise Injection:** Small Gaussian noise was added to the power values during training to make the model robust to measurement errors.
- **Seasonal Subsampling:** Data was sampled separately for each season to ensure balanced representation.

These preprocessing steps ensure that the dataset is clean, well-structured, and ready for input into the deep learning model.

IV. Implementation, Results, and Discussion

The code was developed in a Jupyter Notebook using Python 3.7 on a PC with an Intel® Core™ m7-6Y75 CPU running at 1.20 GHz and 8 GB of RAM, with Windows 10 as

The forecast results in in Fig.2, Fig.3 ,Fig.5,Fig.5 are presented with quantile lines representing different quantiles (0.02, 0.1, 0.25, 0.5, 0.75, 0.9, 0.98), where the y-axis denotes the predicted power in kW, and the x-axis represents the forecast time index. The 'Actual' line represents observed power, allowing us to compare the predictions with actual values and assess the accuracy of the forecasts. As shown figures, the actual values closely align with the 50th (median) and 75th quantiles, indicating that the model's predictions

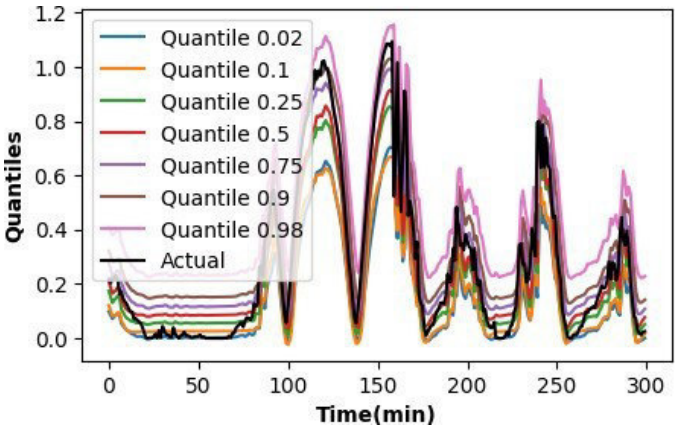


Fig. 2. winter (a)

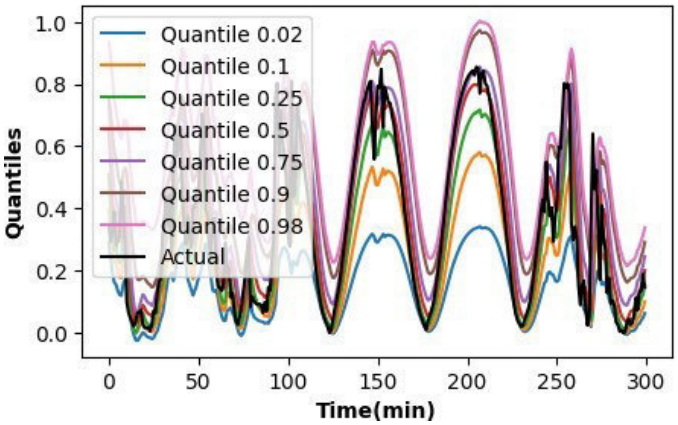


Fig. 3. summer (b)

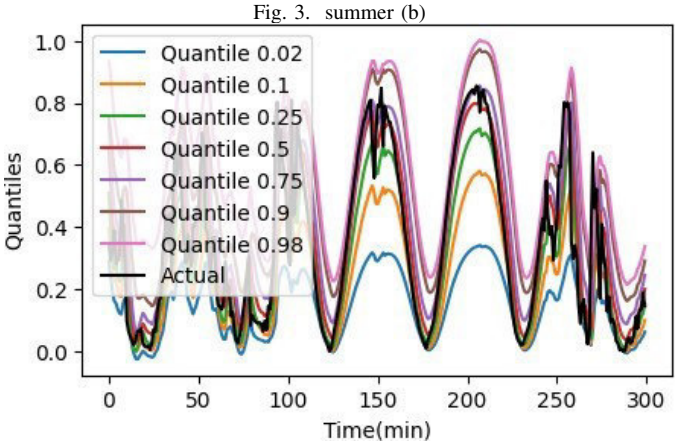


Fig. 4. spring (c)

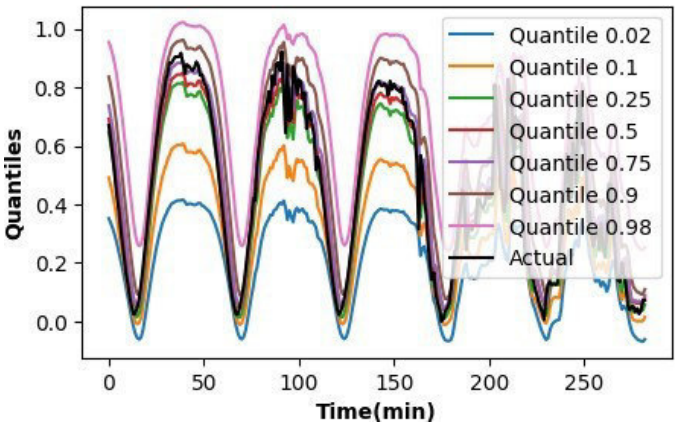


Fig. 5. fall (d)

FT/Boumerdes/NCASEE-24-Conference

are generally accurate for at least 75% of the data. For 75% of the data, the predictions are either close to or lower than the actual values, while for the remaining 25%, the predictions may exceed the observed values. This suggests that the model captures the central tendency (median) and spread (interquartile range) of the data well, though there may be instances where predictions surpass the actual values, likely due to hyperparameter settings. Improving the optimization methods may enhance prediction accuracy for these cases.

The graph also reveals the uncertainty of the forecasts through the spread between different quantile lines, which is valuable for power management decisions. For example, if the lower quantiles indicate high power output for the next hour, it may suggest a high power production, whereas predictions from the upper quantiles indicating low output might call for preparations to handle a potential decrease in power. This uncertainty visualization helps grid operators anticipate potential fluctuations in solar power and plan for the necessary adjustments in grid management and energy storage.

Figure **??** illustrates the forecasted power in different seasons (Winter, Summer, Spring, Fall), showing the actual PV power in black and the 7 quantile predictions in various colors for comparison. This representation highlights the model's ability to predict power with a certain level of uncertainty, useful for balancing power supply and demand effectively.

The seasonal variations in solar power output are clearly visible, with different patterns of uncertainty observed across the seasons. For example, in Winter, the forecast uncertainty is typically lower, as solar power is generally predictable due to consistent weather patterns. However, during the Summer months, the forecasts tend to show a wider range of predictions, reflecting the higher variability in solar radiation due to factors like cloud cover and seasonal transitions.

Moreover, the seasonal analysis shows the model's adaptability to different environmental conditions, capturing both high and low solar output periods across various times of the year. This adaptability is crucial for ensuring that the solar power forecasting model remains reliable throughout different seasonal phases, which is important for optimizing the operation of solar power systems and grid integration.

In conclusion, the quantile forecasting method not only provides point predictions of solar power but also quantifies the uncertainty associated with these forecasts. This dual benefit of predicting power and assessing forecast uncertainty helps in making informed decisions for power grid management, ensuring a reliable and efficient integration of solar energy into the grid. The model's performance, as shown in both the comparison of predicted and actual values and the seasonal forecast analysis, demonstrates its robustness in handling different forecasting scenarios and its ability to support decisions under varying levels of uncertainty.

V. Conclusion

This study presented a probabilistic solar power forecasting model combining Convolutional Neural Networks (CNNs) and Gated Recurrent Units (GRUs) to produce one-hour-ahead quantile predictions. By evaluating the model on solar power data from the Netherlands, it was demonstrated that the hybrid model effectively captures the central tendency and variability of solar power output across different seasonal conditions. The results show that the model achieves robust performance, with median predictions (50th quantile) closely aligning with observed power values and the interquartile range (25th to 75th quantiles) accurately representing the uncertainty in forecasts.

The seasonal analysis further highlights the model's adaptability to varying environmental conditions, demonstrating its ability to maintain reliability in scenarios with both high and low solar output. The forecast uncertainty, represented by the spread between quantiles, provides valuable insights for power grid operators, enabling them to make informed decisions about energy storage, load balancing, and grid stability.

Despite the model's strengths, instances of over- or under-prediction were noted, which may be attributable to hyperparameter settings or optimization limitations. Future work could focus on refining the model through advanced optimization techniques and exploring additional features, such as weather data, to enhance predictive accuracy and reduce uncertainty.

In conclusion, the proposed probabilistic forecasting approach not only provides accurate predictions but also quantifies forecast uncertainty, making it a valuable tool for integrating solar energy into modern power grids. By addressing the variability inherent in solar power generation, this work contributes to the efficient and reliable management of renewable energy resources, supporting the transition to a sustainable energy future.

References

[1] H. Zhang and Y. Zhou, "Recent advances in solar energy utilization: An overview," *Renewable Energy Reviews*, vol. 156, pp. 1234–1250, 2023. [Online]. Available: https://doi.org/10.1016/j.rer.2022.112345

[2] J. Li and X. Chen, "Comparative analysis of solar power forecasting methods: A review," *Journal of Solar Energy Engineering*, vol. 144, no. 4, p. 040801, 2022. [Online]. Available: https://doi.org/10.1115/1.4053345

[3] A. Reddy and S. Kumar, "Deep learning techniques for solar irradiance forecasting: A comprehensive review," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 3, pp. 345–357, 2021. [Online]. Available: https://doi.org/10.1109/TSTE.2020.3034567

[4] X. Chen, J. Zhang, Y. Liu, and Y. Wang, "Satellite Image-based Solar Irradiance Forecasting using CNNs," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 3982–3993, 2020.

[5] Y. Wang and L. Zhang, "Transformer-based models for time series forecasting: Applications to solar energy," *Applied Energy*, vol. 295, p. 117045, 2024. [Online]. Available: https://doi.org/10.1016/j.apenergy.2021.117045

[6] L. Ait Mouloud *et al.*, "Explainable forecasting of global horizontal irradiance over multiple time steps using temporal fusion transformer," *Journal of Renewable and Sustainable Energy*, vol. 15, no. 5, 2023. [Online]. Available: https://doi.org/10.1063/5.0149268

[7] P. Nowotarski, L. B. Z. Paola, "Probabilistic forecasting of solar power generation: A review," Renewable and Sustainable Energy Reviews, vol. 67, pp. 1225-1241, 2017. [Online]. Available: https://doi.org/10.1016/j.rser.2016.09.088

[8] S. Wang, Y. Zhai, "Uncertainty quantification in bioinformatics: A review of probabilistic methods in genomic and clinical data," Bioinformatics, vol. 38, no. 4, pp. 821-833, 2022. [Online]. Available: https://doi.org/10.1093/bioinformatics/btab819

[9] J. D. Smith, "Quantile regression for predictive modeling of biological data," Journal of Computational Biology, vol. 27, no. 10, pp. 1259-1268, 2020. [Online]. Available: https://doi.org/10.1089/cmb.2020.0147

[10] J. Lee, M. Singh, ”Applications of Convolutional Neural Networks in biological data analysis,” Frontiers in Bioinformatics, vol. 2, p. 689422, 2021. [Online]. Available: https://doi.org/10.3389/fbio.2021.689422
[11] R. Patel, V. Z. Gupta, ”Gene sequence prediction using LSTM networks: A case study in bioinformatics,” Bioinformatics, vol. 38, pp. 91-99, 2022. [Online]. Available: https://doi.org/10.1093/bioinformatics/btz344
[12] X. Zhang, Y. Wang, ”Gated Recurrent Units in predictive modeling for clinical time series forecasting,” Medical Image Analysis, vol. 68, p. 101926, 2021. [Online]. Available: https://doi.org/10.1016/j.media.2020.101926
[13] V. Gupta, R. Gupta, ”Hybrid deep learning models for medical disease prediction: Combining CNNs and RNNs,” BioMedical Engineering Online, vol. 19, no. 1, p. 25, 2020. [Online]. Available: https://doi.org/10.1186/s12938-020-0814-6
[14] Y. Wang, Q. Liu, ”Accounting for seasonal variability in biotechnology models,” Computational Biology and Chemistry, vol. 84, p. 107208, 2021. [Online]. Available: https://doi.org/10.1016/j.compbiolchem.2020.107208
[15] S. Hochreiter, J. Schmidhuber, ”Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735
[16] J. Chung, C. Gulcehre, K. Cho, et al., ”Empirical evaluation of gated recurrent neural networks on sequence modeling,” in Proc. Neural Information Processing Systems (NeurIPS), 2014, pp. 1-8. [Online]. Available: https://arxiv.org/abs/1412.3555

Hardware Implementation of V/f control in induction motor using Dspace microlabox

Ismail Ghadbane
Electrical Engineering Laboratory (LGE). University of Msila, Algeria
ismail.ghadbane@univ-msila.dz

BOUZIDI Riad
Electrical Engineering Laboratory (LGE). University of Msila, Algeria
riad.bouzidi@univ-msila.dz

BOUKHARI Mohammed
Electrical Engineering Laboratory (LGE). University of Msila, Algeria
mohammed.boukhari@univ-msila.dz

Khaled Belhouchet
Electrical Engineering Laboratory (LGE). University of Msila, Algeria
khaled.belhouchet@univ-msila.dz

***Abstract*—**

The main objective of this work is the implementation, in real time, of the scalar control for induction motor. To achieve this objective, we used a test bench capable of validating this type of control. this test bench contains a 2.2kw power induction motor, a three-phase voltage inverter (SMEKRON) and a modern dspace control board (Micro-labox) . We identified all the parameters of the motor in the first part and their verification in mathematical modeling under the MATLAB/ SIMULINK. The model presented under matlab represents exactly the induction motor that is deduced from the shape of the curves, current, flux ,torque and speed. The second part said the motor inverter association where' we have the closed loop driver using a three-phase inverter in use the PWM pulse width modeling technique sine-triangle. Experimental results a training has good dynamics.

***Keywords*—**Dspace-MicroLabBox , Induction Motor, PWM.

I. INTRODUCTION (*HEADING 1*)

Induction motors are widely used in industrial and commercial applications due to their robustness, simplicity, and cost-effectiveness. A key feature of induction motors is that they are powered by alternating current (AC) and operate based on electromagnetic induction. To optimize their performance, control techniques are applied to regulate speed, torque, and efficiency in various systems.[1, 2]

The key principle behind the operation of an induction motor is electromagnetic induction: as the rotating magnetic field of the stator cuts across the rotor, it induces an electromotive force (EMF) in the rotor. This induced current produces its own magnetic field, which interacts with the stator's magnetic field, causing the rotor to rotate. The rotor always tries to catch up with the rotating magnetic field but never quite matches its speed, hence the term asynchronous motor.[3]

The induction motor is a fundamental component in modern electrical and mechanical systems, thanks to its simple construction, durability, and wide range of applications. Whether in industrial machinery or household appliances, the induction motor remains one of the most important types of motors in use today.

The (V/f) control method is one of the most widely used techniques for controlling the speed and torque of induction motors, especially in applications where simplicity, cost-effectiveness, and moderate performance are essential. The fundamental concept behind V/f control is maintaining a constant ratio between the motor's applied voltage and the frequency of the supply, ensuring that the magnetic flux in the motor remains consistent as the speed varies. This method provides an efficient way to control motor speed while preventing overheating and ensuring proper torque production across different operating conditions. V/f control is typically implemented using a Variable Frequency Drive (VFD), which adjusts the frequency and voltage supplied to the motor, allowing precise control over the motor's performance. While V/f control is less complex than advanced methods like vector control or direct torque control, it offers a practical solution for many industrial and commercial applications, such as fans, pumps, and conveyors. However, it has limitations in terms of dynamic performance and efficiency at low speeds, where advanced techniques may be more beneficial. Despite these limitations, V/f control remains a cost-effective and widely adopted solution in the field of motor control.[5,6]

The main objective of this work is the evaluation, by an experimental implementation, in real time of the scalar control performance of an induction motor associated with a voltage inverter PWM.

This work is organized as follows , the first part will be devoted to the identification of the parameters of the induction motor using the method of conventional tests, direct current test, no load test, rotor blocked test and deceleration test.

The second part will present the motor inverter association, that is to say the open loop control. The output voltage is set through the three-phase inverter driven by the PWM pulse width modulation technique.

The third part deals with the practical validation of the scalar control of the asynchronous motor using a test bench containing a dspace card of type Micro Labox.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

II. THE MODEL FOR AN INDUCTION MOTOR

The presented induction motor's equivalent circuit, found in Figure 1, features a model situated at the dqo axis. By employing this circuit, we can effectively describe the mathematical representation of the induction motor. [8]

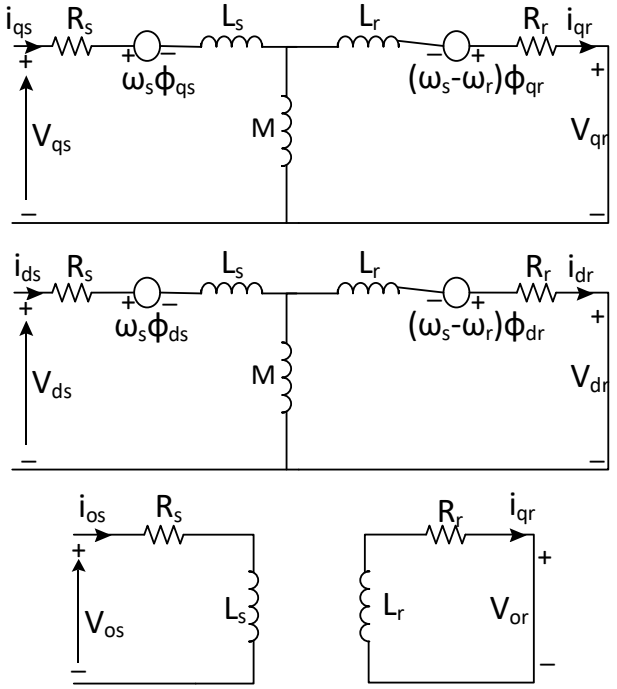


Fig. 1. Induction motor's equivalent circuit in the dq0 axis framework.

It must define the problem and importance of the research carried out, it presents a (not very extensive) review of the literature on the subject of the article, including the authors' contributions to the state of the art. If you use abbreviations or acronyms, first write the words that identify them and then, in parentheses, the acronym. This set also establishes the research question, the objectives of the work and hypothesis, if necessary, the importance and limitations of the study. Establishes the method used at work. It is written in the present tense.

$$[\dot{I}] = [L]^{-1}\{[V] - [R][I]\} \tag{1}$$

$$[I] = \begin{bmatrix} i_{ds} \\ i_{qs} \\ i_{dr} \\ i_{qr} \end{bmatrix}; [V] = \begin{bmatrix} v_{ds} \\ v_{qs} \\ v_{dr} \\ v_{qr} \end{bmatrix}; [L] = \begin{bmatrix} L_s & 0 & M & 0 \\ 0 & L_s & 0 & M \\ M & 0 & L_r & 0 \\ 0 & M & 0 & L_r \end{bmatrix} \tag{2}$$

Where:

$$[R] = \begin{bmatrix} R_s & -\omega_s L_s & 0 & -\omega_s M \\ \omega_s L_s & R_s & \omega_s M & 0 \\ 0 & -\omega_{sl} M & R_r & -\omega_{sl} L_r \\ -\omega_{sl} M & 0 & \omega_{sl} L_r & R_r \end{bmatrix} \tag{3}$$

The dq0 axis, where d and q denote the respective axes, the following terms are defined: I (current), V (voltage), R (resistance), L (inductance), M (magnetizing inductance), ω (angular speed), ωsl (slip speed), and s and r signify the stator and rotor components. The flux equations can be formulated as:

III.SCALAR-BASED CONTROL ALGORITHM FOR REGULATION

Fig 2 represent diagram of the system with PI regulator. The diagram sown in Figure 3 illustrates the block scalar control method from Figure 1, featuring a PI-type regulator. By utilizing the speed error, it provides us with the reference slip speed value (ωsl*) as indicated in the subsequent equation [9]:

$$\omega_{sl}^{*} = (\omega_r^{*} - \omega_{r_{mes}}) \times (k_p + \frac{k_i}{S}) \tag{8}$$

Where respectively ωr* and ωr measure the reference and observed speeds. We can determine the proportional and integral gains, denoted by kp and ki, respectively, for the PI regulator, by using the following [2]:

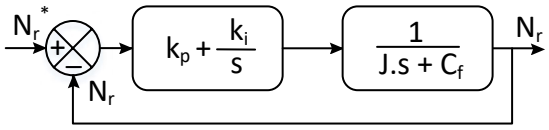


Fig 2. System with PI regulator

The open-loop transfer function, denoted as ToL, may be calculated using Fig. 1 as follows:

$$T_{oL} = \frac{N_r}{N_r^{*}} = \frac{k_p.S + k_i}{J.S^2 + C_f.S} \tag{9}$$

The provided equation determines the closed-loop transfer function, denoted as TcL :

$$T_{cL} = \frac{T_{oL}}{1+T_{oL}} = \frac{\frac{k_p.S + k_i}{J}}{S^2 + \frac{k_p + C_f}{J}.S + \frac{k_i}{J}} \tag{10}$$

A canonical form representation of the second-order transfer function Tcfis offered as follows:

$$T_{cf} = \frac{k}{S^2 + 2\xi\omega_n.S + \omega_n^{\ 2}} \quad (11)$$

ωn : is the own pulse

ξ : represents the damping coefficient

$$\begin{cases} k_p = 2\xi\omega_n J - C_f \\ k_i = \omega_n^{\ 2} J \end{cases} \quad (12)$$

After the computation, we obtain: ki = 0.001 ; kp = 0.025

The following equation gives us the stator reference speed pulsation ωs* :

$$\omega_s^{\ *} = \omega_{r_{mes}} + \omega_{sl} \quad (13)$$

$$\overline{V_s} = R_s \overline{I_s} + j\omega_s \overline{\Phi_s} \quad (14)$$

The relationship between voltage (V) and efficiency should be directly proportional to the stator's supply frequency (fs) in order to maintain a constant flux. However, at high speeds, Rs is too small in relation to ωs therefore we can write it as follows instead:

$$\Phi_s = \frac{V_s}{\omega_s} = \frac{V_s}{2\pi f_s} \quad (15)$$

For the low value of the pulse, however, this relation is invalid because the voltage drop Rs caused by the stator windings' resistance is no longer insignificant in comparison to the given term ωs*, Φs [2].

III. EXPERIMENTAL VALIDATION

A test bench was set up at the LGE research laboratory at M'sila university

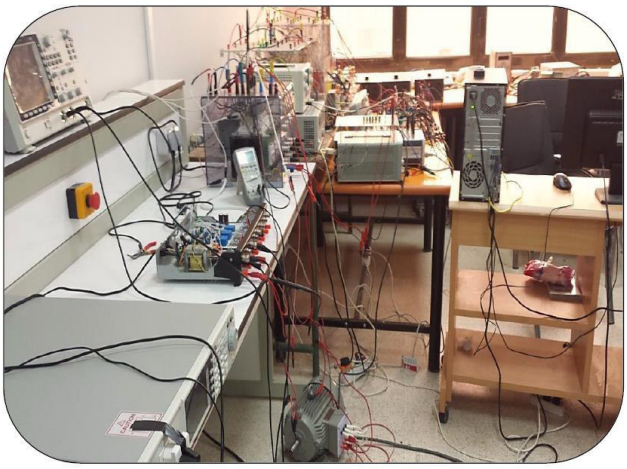


Figure 1. Experimental test bench

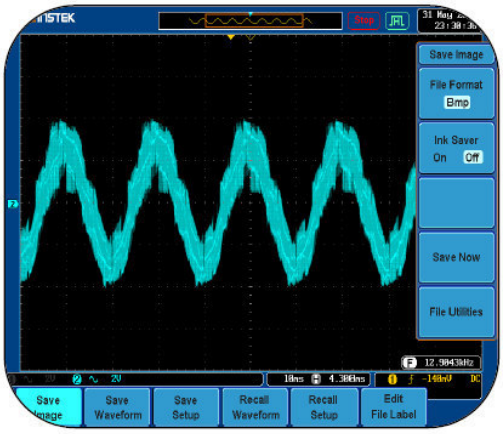
A. *Description of the test bench*

- A three-phase asynchronous motor with a nominal power of 2.2 kW.
- A three-phase SEMIKRON SEMITEACH IGBT voltage inverter
- An RTI 1202 Dspace
- An adaptation Dspace
- Incremental coder .
- Measuring devices (multimeter, oscilloscope,..). and self-transformer 0-470 v

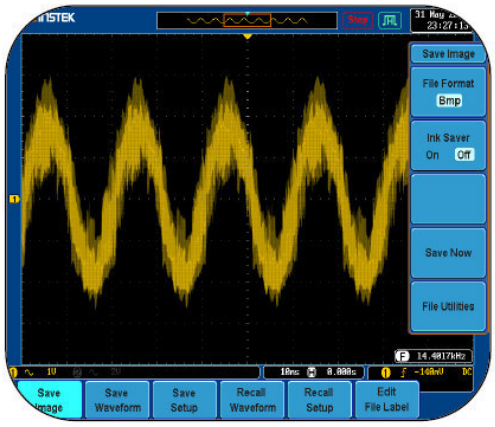
We have carried out in this part two different tests, the tests that concern the association of the inverter-motor asynchronous, whose inverter is driven by the technique of modulation of width of pulse MLI sinusoidal frequency of switching 20KHz.

The asynchronous motor used at no load , star-coupled, the frequency of the supply voltages is 50 Hz and the DC voltage is 175V.

In the second test we have imposed several instructions, the following results show the three measured quantities the simple voltage, the compound voltage and the rotation speed



-a-



-b-

Figure 2. The inverter PWM compound and single voltage

B. No load speed test

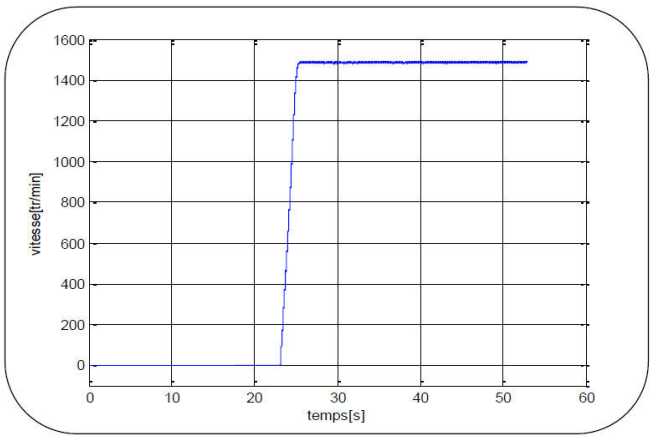


Figure 3. Result of experimental no-load test

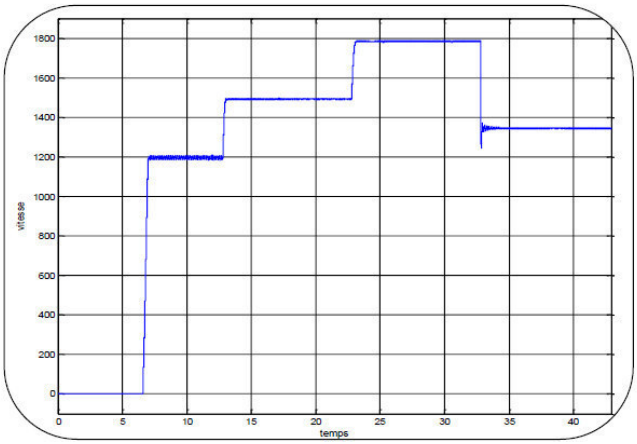


Figure 4. Speed Tracking Test

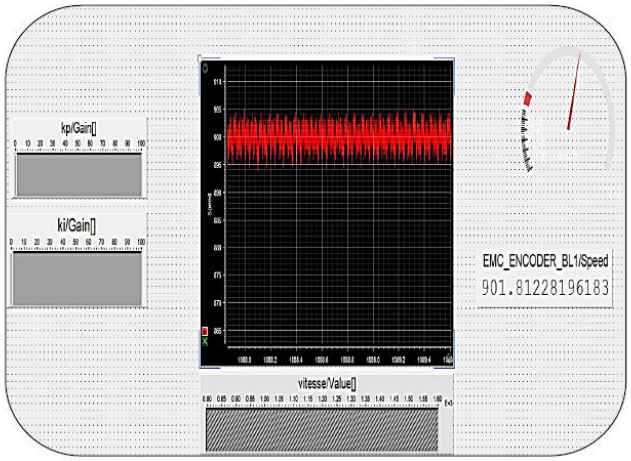


Figure 5. Test for a speed setting of 900 r/min.

During the transient regime Figure (3) the engine speed increases linearly in a very short response time, then stabilizes without overshoot and close to the synchronism speed 1495 r/min, because the motor is unladen and the voltage imposed on the supply voltages is 50 Hz.

Figure (2) shows the simple tension patterns in blue and the compound in yellow. The form of the compound voltage is very close to the sinusoid whose peak-to-peak value almost 50 V. this voltage is measured at the sensor output whose gain is 0.18 so Vco max 140 V. The single voltage is tainted by harmonic of peak-to-peak value equal to 445 V per Therefore Vsim max 220 V.

In the second test, Figure (4), (speed tracking) the response of speed to good dynamics with good performance indices. The rotor speed follows its speed records perfectly in very short response times and without overshooting. However, there is a static error of about 20 tr/min for both the 1800 tr/min and 1300 tr/min. This error has no effect on the trainen

Figure (5), is an example of the 'Control Desk" image of the dspace board. Through this interface, we can view the rotation speed of our asynchronous motor in several modes, oscillogram, digital display and analog display (needle) as well as the adjustment of the gains of the PI regulator in real time (Kp and Ki).

.

CONCLUSION

Experimental validation of the real-time scalar control of an asynchronous motor. This validation was performed in two different ways. One is carried out in the Simulink/ Matlab environment, including the drive system (inverter- motor and control technique) was simulated , while the second method discussed the practical validation of this control in a test bench containing a dspace board a three-phase inverter and an asynchronous motor of 2.2 KW power. The results obtained show well the efficiency of the scalar control vis-à-vis the variation of the speed which is translated by a follow perfectly to our recorded speeds, with response times very accepted, without overshoot and without static error

REFERENCES

1] M. Ehsani, K. V. Singh, H. O. Bansal, and R. T. Mehrjardi, "State of the Art and Trends in Electric and Hybrid Electric Vehicles," in Proceedings of the IEEE, vol. 109, no. 6, pp. 967-984, June 2021, doi: 10.1109/JPROC.2021.3072788.

[2] Tran, T. C., Brandstetter, P., & Vo, H. H. (2019). The Sensorless Speed Controller of Induction Motor in DFOC Model Based on the Voltage and Current. Journal of Advanced Engineering and Computation, 3(1), 320-328.

[3] Chen , J., Huang, J.(2018).Stable simultaneous stator and rotor resistances identification for speed sensorless IM drives: Review and new results . IEEE Trans. Power Electron., vol. 33, no. 10, pp. 8695- 8709 .

[4] D. Xu, B. Wang, G. Zhang, G. Wang and Y. Yu, "A review of sensorless control methods for AC motor drives," in CES Transactions on Electrical Machines and Systems, vol. 2, no. 1, pp. 104-115, March 2018, doi: 10.23919/TEMS.2018.8326456.

[5] Duranay, Z. B., Guldemir, H., & Tuncer, S. (2020). Implementation of a V/F controlled variable speed induction motor drive. EMITTER International Journal of Engineering Technology, 8(1), 35-48.

[6] G. Amar, G. Ismail and B. Riad, "Suitability of Electric Motors and Conventional Control Techniques for Electric Vehicle Applications," 2022

International Conference of Advanced Technology in Electronic and Electrical Engineering (ICATEEE), M'sila, Algeria, 2022, pp. 1-6, doi: 10.1109/ICATEEE57445.2022.10093716.

[7] El Daoudi, S., Lazrak, L., El Ouanjli, N., & Lafkih, M. A. (2021). Sensorless fuzzy direct torque control of induction motor with sliding mode speed controller. Computers & Electrical Engineering, 96, 107490.

[8] . F. M. El-Sousy, M. M. Amin and O. A. Mohammed, "Robust Adaptive Neural Network Tracking Control With Optimized Super-Twisting Sliding-Mode Technique for Induction Motor Drive System," in IEEE Transactions on Industry Applications, vol. 58, no. 3, pp. 4134-4157, May-June 2022, doi: 10.1109/TIA.2022.3160136.

[9] H. Wang, Y. Yang, X. Ge, Y. Zuo, Y. Yue and S. Li, "PLL- and FLL-Based Speed Estimation Schemes for Speed-Sensorless Control of Induction Motor Drives: Review and New Attempts," in IEEE Transactions on Power Electronics, vol. 37, no. 3, pp. 3334-3356, March 2022, doi: 10.1109/TPEL.2021.3117697.

[10] Mishra, P., Lascu, C., Bech, M. M., Rannestad, B., & Munk-Neilsen, S. (2021, October). Design and Analysis of PLL Speed Estimator for Sensorless Rotor-Flux Oriented Control of Induction Motor Drives. In 2021 IEEE Energy Conversion Congress and Exposition (ECCE) (pp. 4743-4747). IEEE.

[11] U. Syamkumar and J. B, "A Smoothed Kalman filter Trained MRAS Based Recurrent Neural Observer for Indirect Vector Controlled Induction Motor Drive," 2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS), Shah Alam, Malaysia, 2021, pp. 396-401, doi: 10.1109/I2CACIS52118.2021.9495860

[12] Bhaumik and S. Das, "Predictive Torque Control Strategy for Speed Adaptive Flux Observer Based Sensorless Induction Motor Drive in Flux-Weakening Region," in IEEE Transactions on Power Electronics, vol. 36, no. 12, pp. 14110-14118, Dec. 2021, doi: 10.1109/TPEL.2021.3089611

[13] P. Mishra, C. Lascu, M. M. Bech, B. Rannestad and S. Munk-Neilsen, "Design and Analysis of PLL Speed Estimator for Sensorless Rotor-Flux Oriented Control of Induction Motor Drives," 2021 IEEE Energy Conversion Congress and Exposition (ECCE), Vancouver, BC, Canada, 2021, pp. 4743-4747, doi: 10.1109/ECCE47101.2021.9595316.

[14] H. R. V, R. S. Kaarthik and P. P. Rajeevan, "Online Inductance Estimation of Speed Sensor less Induction Motor Drives," 2020 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Jaipur, India, 2020, pp. 1-5, doi: 10.1109/PEDES49360.2020.9379853.

[15] A. Nurettin and N. İnanç, "Sensorless Vector Control for Induction Motor Drive at Very Low and Zero Speeds Based on an Adaptive-Gain Super-Twisting Sliding Mode Observer," in IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 11, no. 4, pp. 4332-4339, Aug. 2023, doi: 10.1109/JESTPE.2023.3265352

A Fuzzy Logic-Enhanced Variable Step Size P&O Algorithm for Efficient MPPT in Photovoltaic Systems

1st DENNAI Mohammed Yassine
Faculty of Technology, Smart Grids & Renewable Energies Laboratory (SRGE)
University of Tahri Mohammed
Bechar, Algeria.
dennai.mohammedyassine@univ-bechar.dz

2nd ABDELOUAHED Touhami
Faculty of Technology, Department of Electrical Systems Engineering
University M'Hamed Bougara
Boumerdes, Algeria
a.touhami@univ-boumerdes.dz

3rd TEDJINI Hamza
Faculty of Technology, Smart Grids & Renewable Energies Laboratory (SRGE)
University of Tahri Mohammed
Bechar, Algeria.
tedjini.hamza@univ-bechar.dz

***Abstract*—This paper presents a Modified Perturb and Observe (P&O) Maximum Power Point Tracking (MPPT) algorithm designed to enhance the efficiency of photovoltaic (PV) systems. The conventional P&O MPPT algorithm, widely used for its simplicity, suffers from limitations such as slow transient response and steady-state oscillations around the maximum power point (MPP), leading to suboptimal energy harvesting. To address these issues, a fuzzy logic-based variable step size approach is integrated into the P&O algorithm. This modification dynamically adjusts the perturbation step size based on real-time conditions, allowing for faster convergence to the MPP and reduced voltage oscillations in steady-state operation. The proposed MPPT algorithm was implemented and validated using MATLAB/Simulink. Simulation results demonstrate significant improvements in tracking speed and stability, making this approach a promising solution for enhancing the performance of PV systems under varying environmental conditions.**

Keywords— PV, MPPT, P&O, Fuzzy Logic Control, Variable Step Size (VSS).

I. INTRODUCTION

Photovoltaic (PV) systems have gained significant attention as a sustainable and environmentally friendly energy source. However, their efficiency is highly dependent on environmental factors such as solar irradiance and temperature. To ensure that PV systems operate at their maximum efficiency, Maximum Power Point Tracking (MPPT) algorithms are employed to continuously adjust the operating point of the system to the Maximum Power Point (MPP). Among these algorithms, the Perturb and Observe (P&O) method is widely used due to its simplicity and ease of implementation. However, the conventional P&O algorithm suffers from notable drawbacks, including slow transient response and steady-state oscillations around the MPP, which hinder optimal energy harvesting, particularly under rapidly changing environmental conditions. To address these limitations, research has focused on modifying and enhancing MPPT algorithms to improve performance [1].

This paper presents a modified P&O MPPT algorithm that integrates a fuzzy logic-based variable step size approach to overcome the limitations of the traditional P&O method. The modification allows for real-time adjustment of the perturbation step size, leading to faster convergence to the MPP and reduced voltage oscillations in steady-state operation. This dynamic step-size adjustment helps maintain stability and efficiency in varying environmental conditions. The proposed method was validated through MATLAB/Simulink simulations, which showed significant improvements in tracking speed and system stability compared to conventional methods.

Various MPPT methods have been proposed to address the limitations of the conventional P&O algorithm. In recent years, the need for more adaptive, efficient, and reliable MPPT techniques has driven extensive research in this field. Among these methods are techniques such as Incremental Conductance (IC), fuzzy logic-based MPPT, neural network-based MPPT, and hybrid methods that combine multiple strategies. These methods aim to improve the tracking speed, accuracy, and stability of PV systems under dynamic environmental conditions [2].

Incremental Conductance (IC) Method: The IC algorithm is one of the most widely studied alternatives to the conventional P&O method. It compares the incremental conductance and the instantaneous conductance to determine whether the operating point is approaching the MPP. Unlike the fixed-step P&O algorithm, IC has the advantage of higher accuracy near the MPP and better performance under rapidly changing conditions. However, it is more complex to implement and can suffer from slower tracking speeds when not optimized . Variants of the IC method that incorporate adaptive step sizes have been developed to improve transient response and reduce oscillations around the MPP [3].

Fuzzy Logic-Based MPPT: Fuzzy logic control (FLC) has been integrated into MPPT algorithms to create adaptive control systems that do not rely on a mathematical model of the PV system. Fuzzy logic-based MPPT algorithms dynamically adjust the perturbation step size based on real-time input variables such as voltage and current changes. These algorithms are particularly useful for handling non-linear, uncertain, and rapidly changing conditions. In contrast to conventional methods, fuzzy logic-based MPPTs offer smoother convergence to the MPP, reduced oscillations, and better adaptability to environmental changes. Research has shown that fuzzy logic controllers improve tracking efficiency by mitigating the limitations of both P&O and IC methods, particularly when handling partial shading and variable irradiance conditions [4].

Neural Network-Based MPPT: Neural networks (NNs) have also been explored as a method for improving MPPT performance. These networks are trained using historical data from the PV system to predict the MPP more accurately. Once

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

trained, NN-based MPPT controllers can quickly and precisely adjust the system's operating point, leading to faster convergence compared to traditional methods. However, their performance heavily depends on the quality and quantity of training data, making them less suitable for systems that frequently experience new or unseen environmental conditions. NN-based methods have demonstrated success in dealing with complex PV systems and partial shading, but they require significant computational resources for real-time operation [5].

Hybrid MPPT Methods: The integration of multiple techniques into a single hybrid MPPT algorithm has been an emerging area of research. Hybrid methods often combine elements of P&O, IC, fuzzy logic, and neural networks to achieve the benefits of each approach. For instance, combining the simplicity of P&O with the adaptability of fuzzy logic or IC can result in an algorithm that is both easy to implement and capable of tracking the MPP with greater precision. In one such hybrid approach, fuzzy logic is used to control the perturbation size, while the P&O algorithm handles the overall tracking logic. This combination has proven effective in reducing oscillations around the MPP and improving the transient response under variable irradiance conditions [6].

Proposed Approach: Modified P&O with Fuzzy Logic-Based Variable Step Size

The modified P&O algorithm presented in this paper builds on these advancements by integrating a fuzzy logic-based variable step size into the conventional P&O algorithm. Unlike fixed-step methods, this approach dynamically adjusts the perturbation step size according to the current operating conditions. Fuzzy logic systems excel in making decisions based on vague or imprecise data, which allows the algorithm to adapt to real-time changes in the PV system's performance. This dynamic adjustment results in faster convergence to the MPP during transient conditions and reduced oscillations in steady-state operation, offering an optimal balance between tracking speed and system stability.

Previous studies have highlighted the importance of step-size modification in improving MPPT performance. Fixed-step P&O methods, while simple, often fail to provide the necessary balance between speed and stability. Large step sizes result in fast tracking but introduce significant oscillations around the MPP, while small step sizes minimize oscillations but slow down the convergence process. The integration of fuzzy logic allows the step size to be dynamically altered based on factors such as the rate of change in power and voltage, optimizing the system's response to different environmental conditions. This approach addresses the key limitations of traditional methods by offering improved transient performance without sacrificing steady-state accuracy [7].

Contribution and Validation

The contribution of this work lies in the application of fuzzy logic to improve the adaptability and efficiency of the P&O algorithm. By dynamically adjusting the step size, the modified algorithm provides a faster transient response and minimizes the steady-state oscillations that typically affect conventional P&O methods. The proposed MPPT algorithm was validated through MATLAB/Simulink simulations under a variety of environmental conditions, including rapid changes in irradiance and temperature. The simulation results demonstrated significant improvements in both tracking speed and stability compared to the conventional P&O algorithm and other fixed-step methods. Specifically, the fuzzy logic-based step size control enabled the system to quickly reach the MPP during transient conditions and maintain stable operation with minimal oscillations in steady-state operation.

The integration of fuzzy logic-based step size adjustment into the P&O MPPT algorithm represents a significant advancement in PV system optimization. The proposed method effectively addresses the critical challenges associated with conventional MPPT techniques, offering improved tracking speed, accuracy, and stability performance. As the demand for more efficient and adaptive renewable energy systems continues to grow, the advancements in MPPT techniques presented in this work are essential for enhancing the overall performance of PV systems and ensuring reliable energy generation in variable environmental conditions.

II. MODELING SYSTEM

A. *Modeling of Photovoltaic Arrays*

Modeling a photovoltaic (PV) array requires an understanding of how a solar cell functions as an electrical circuit. Figure 2 depicts the equivalent circuit of a PV cell, which comprises series resistance (Rs) and parallel (shunt) resistance (Rsh). These resistances play a vital role in assessing the overall efficiency of the PV cell. The circuit also includes photocurrent (Iph), diode current (Id), and shunt current (Ish) [8].

Several important factors influence the behavior of a PV cell, particularly cell temperature and solar irradiance. The photocurrent is significantly affected by these variables, with higher irradiance resulting in increased current generation. Similarly, the temperature of the cell impacts the saturation current, which varies based on the semiconductor material used in the cell [9].

In this analysis, the Conergy Power Plus 214P PV module is examined. Under standard test conditions (STC), which include a reference temperature of 25°C and solar irradiation of 1000 W/m², this module can achieve a maximum power output of 250W. The electrical specifications of this module are outlined in Table I.

This methodology offers a straightforward way to comprehend the energy output of a PV system, employing the single-diode circuit model to estimate performance based on essential environmental factors and system parameters.

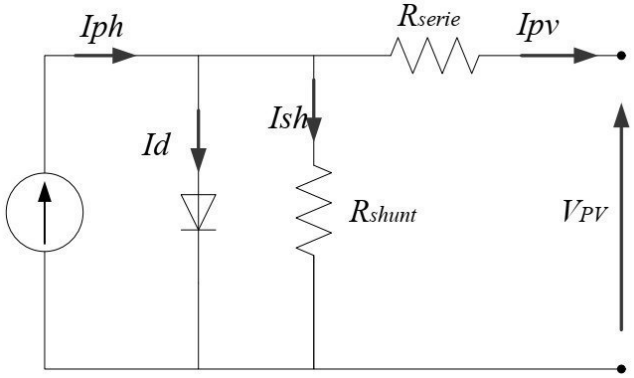


Figure 1: Equivalent circuit of a photovoltaic cell.

TABLE I. ELECTRICAL CHARACTERISTICS OF THE SOLAR PANEL.

Table Head	Table Column Head
	Table column subhead
Maximum power (Pmax)	250 W
Voltage at Maximum Power (Vmp)	30.7 V
Current at Maximum Power (Imp)	8.15 A
Open circuit voltage (Voc)	37.3 V
Short circuit current (Iscr)	8.66 A
Temperature Co-efficient of *Voc*	-0.36901 V/°C
Temperature Co-efficient of *Isc*	0.086998 mA/°C

B. *Boost DC-DC Converter Design*

The DC-DC boost converter is engineered to increase the input voltage of 30 V to a target output voltage of 200 V. This voltage elevation is accomplished by adjusting the duty cycle, D, of the converter. To enhance the converter's efficiency and reduce losses, the switching frequency, *fs*, is established at 20 kHz. Key parameters such as the inductor current, IL, and the inductor ripple current, ΔIL, play a crucial role in determining the appropriate values for the inductor, L, and capacitor, C, which are vital for ensuring stable operation and minimizing output voltage ripple [10]. The calculated values for inductance, L, and capacitance, C, for the designed boost converter are summarized in Table 2.

TABLE II. THE SPECIFICATIONS OF THE DC-DC CONVERTER.

PARAMETER	VALUE
D	0.48
L	0.23 H
C	150 µF
IL	8.2 A
ΔIL	10% of IL
fs	20 kHz

III. THE MPPT ALGORITHM

Figure 2 illustrates the flowchart for the proposed fuzzy logic-based variable step size Perturb and Observe (P&O) Maximum Power Point Tracking (MPPT) algorithm, designed to enhance the energy efficiency of photovoltaic (PV) systems. While the traditional P&O method is often favored for its simplicity, it suffers from slow response times to environmental changes and oscillations around the Maximum Power Point (MPP), leading to potential energy losses. To address these issues, the proposed algorithm incorporates a fuzzy logic-based variable step size approach. This allows for the dynamic adjustment of the perturbation step size based on current operating conditions. During rapid changes in solar irradiance, the step size is increased to enable faster convergence to the MPP. Once the MPP is reached, the step size is decreased to reduce oscillations and stabilize the system. This adaptive approach significantly improves the speed and precision of MPP tracking while also minimizing power losses during steady-state conditions. The fuzzy logic controller assesses variations in voltage, current, and power to make appropriate adjustments to the step size. The effectiveness of the algorithm is validated through simulation results from MATLAB/Simulink, which show enhanced convergence times and reduced voltage oscillations compared to the conventional P&O method. In conclusion, the Modified P&O MPPT algorithm presents a more efficient and dependable option for real-time applications in PV systems.

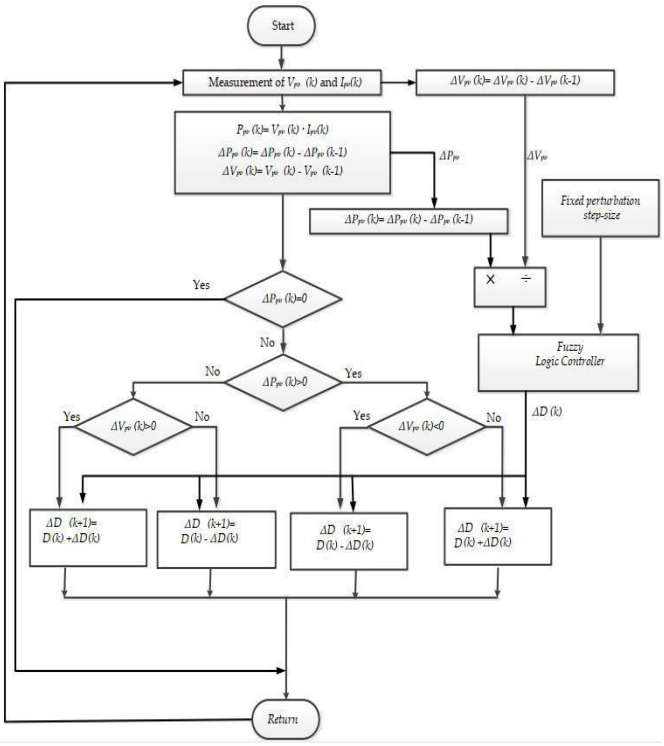


Figure 2: Equivalent circuit of a photovoltaic cell.

IV. SIMULATION RESULTS AND DISCUSSION

Figure 3 illustrates the responses of PV and load currents for the P&O and FLC VSS MPPT techniques under different load conditions, measured every 0.3 seconds. The FLC VSS approach delivers more stable and smoother current profiles, adapting quickly to changes in load with minimal oscillations during transitions. In contrast, the P&O method shows a slower recovery and more significant fluctuations amid these dynamic load changes. This comparison underscores the enhanced effectiveness of FLC VSS in managing rapid load variations.

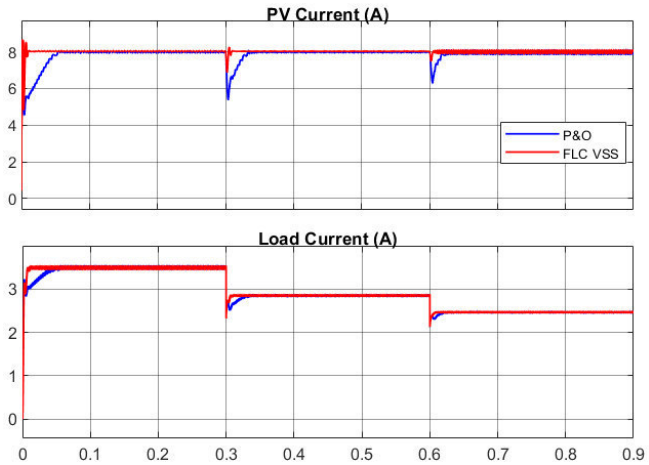


Figure 3: Comparison of PV and Load Current Responses for P&O and FLC VSS MPPT Techniques Under Dynamic Load Changes.

Figure 4 compares the performance of two Maximum Power Point Tracking (MPPT) techniques: Perturb and Observe (P&O) and Fuzzy Logic Control with Variable Step Size

(FLC VSS), focusing on the regulation of PV voltage and load voltage. The upper subplot depicts the PV voltage (V) over time, while the lower subplot illustrates the load voltage (V) over time. The FLC VSS method exhibits more stable and consistent voltage profiles with reduced oscillations, indicating a quicker and more precise tracking of the maximum power point (MPP). In contrast, the P&O technique shows considerable fluctuations, particularly during transient events, highlighting its limitations in dynamic situations. Overall, the FLC VSS approach demonstrates improved voltage stability and efficiency under varying conditions.

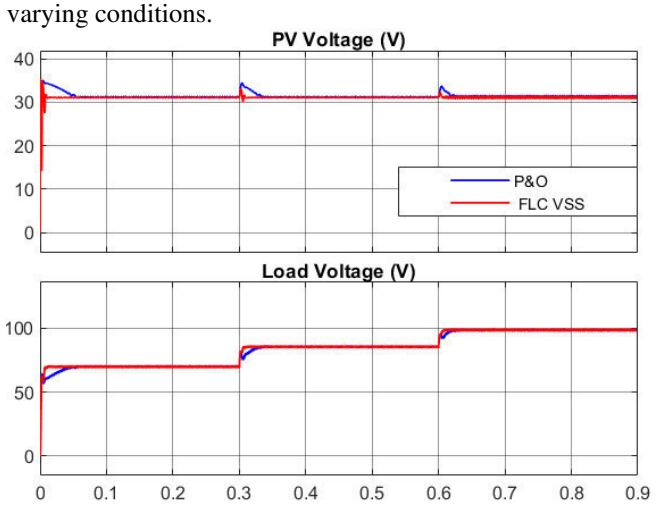


Figure 4: Comparison of PV and Load Voltage Responses for P&O and FLC VSS MPPT Techniques Under Dynamic Load Changes.

Figure 5 compares the PV and load power responses of P&O and FLC VSS MPPT techniques under load variations every 0.3 seconds. The FLC VSS method demonstrates better stability and efficiency, quickly reaching and maintaining maximum power with minimal transient effects. On the other hand, P&O shows significant oscillations and delays during power transitions, resulting in less efficient operation. These results underline the robustness of FLC VSS in maximizing power output under dynamic conditions.

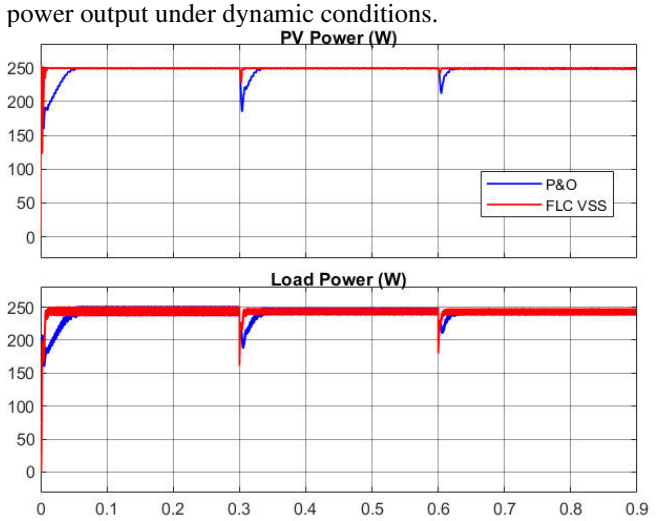


Figure 5: Comparison of PV and Load Power Responses for P&O and FLC VSS MPPT Techniques Under Dynamic Load Changes.

V. Conclusion

This research evaluated the effectiveness of the Perturb and Observe (P&O) method and the Fuzzy Logic Control with Variable Step Size (FLC VSS) for Maximum Power Point Tracking (MPPT) under different load scenarios. The findings indicate that FLC VSS surpasses P&O in stability, efficiency, and dynamic response. FLC VSS demonstrated smoother transitions, quicker convergence, and reduced oscillations in photovoltaic voltage, current, and power, thereby optimizing power delivery to the load in changing conditions. These results underscore the promise of FLC VSS as a reliable and efficient MPPT approach, especially for applications that demand high adaptability and dependability in renewable energy systems.

References

[1] A. Anurag, S. Bal, S. Sourav, and M. Nanda, "A review of maximum power-point tracking techniques for photovoltaic systems," *Int. J. Sustain. Energy*, vol. 35, no. 5, pp. 478–501, May 2016, doi: 10.1080/14786451.2014.918979.

[2] M. A. Elgendy, B. Zahawi, and D. J. Atkinson, "Assessment of perturb and observe MPPT algorithm implementation techniques for PV pumping applications," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 21–33, Jan. 2012, doi: 10.1109/TSTE.2011.2168245.

[3] T. Esram and P. L. Chapman, "Comparison of photovoltaic array maximum power point tracking techniques," *IEEE Trans. Energy Convers.*, vol. 22, no. 2, pp. 439–449, Jun. 2007, doi: 10.1109/TEC.2006.874230.

[4] R. Rahmani, M. Seyedmahmoudian, S. Mekhilef, and R. Yusof, "Implementation of fuzzy logic maximum power point tracking controller for photovoltaic system," *Am. J. Appl. Sci.*, vol. 10, no. 3, pp. 209–218, Apr. 2013, doi: 10.3844/AJASSP.2013.209.218.

[5] L. M. Elobaid, A. K. Abdelsalam, and E. E. Zakzouk, "Artificial neural network based maximum power point tracking technique for PV systems," *IECON Proc. (Industrial Electron. Conf.*, pp. 937–942, 2012, doi: 10.1109/IECON.2012.6389165.

[6] A. K. Abdelsalam, A. M. Massoud, S. Ahmed, and P. N. Enjeti, "High-performance adaptive Perturb and observe MPPT technique for photovoltaic-based microgrids," *IEEE Trans. Power Electron.*, vol. 26, no. 4, pp. 1010–1021, 2011, doi: 10.1109/TPEL.2011.2106221.

[7] M. ALSHAREEF, "An Improved MPPT Method Based on Fuzzy Logic Controller for a PV System," *Stud. Informatics Control*, vol. 30, no. 1, pp. 89–98, Mar. 2021, doi: 10.24846/V30I1Y202108.

[8] H. Patel and V. Agarwal, "MATLAB-based modeling to study the effects of partial shading on PV array characteristics," *IEEE Trans. Energy Convers.*, vol. 23, no. 1, pp. 302–310, Mar. 2008, doi: 10.1109/TEC.2007.914308.

[9] W. De Soto, S. A. Klein, and W. A. Beckman, "Improvement and validation of a model for photovoltaic array performance," *Sol. Energy*, vol. 80, no. 1, pp. 78–88, Jan. 2006, doi: 10.1016/J.SOLENER.2005.06.010.

[10] L. Vinet and A. Zhedanov, "A 'missing' family of classical orthogonal polynomials," *J. Phys. A Math. Theor.*, vol. 44, no. 8, pp. 1–14, 2011, doi: 10.1088/1751-8113/44/8/085201.

Single-Stage Single-Phase Grid Connected Inverter with Proportional Resonant Controller for Enhanced PV System Performance

Salah Anis KRIM, Fateh KRIM
Department of Automation
Laboratory of Power Electronics and Industrial Control
University of Setif - 1
Setif, Algeria
krimsalahanis@gmail.com
krim_f@ieee.org

Hamza AFGHOUL
Department of Automation
Laboratory of Power Electronics and Industrial Control
University of Setif - 1
Setif, Algeria
hamza.afghoul@gmail.com

***Abstract* — This paper introduces a current control strategy for a single-phase grid-connected DC/AC inverter utilized in photovoltaic (PV) power conditioning systems. It incorporates a maximum power point tracking (MPPT) algorithm to ensure optimal power extraction, followed by a proportional-integral (PI) controller to generate the reference current. Additionally, a proportional-resonant (PR) controller is employed to provide an infinite gain at the fundamental frequency, achieving zero steady-state error. The theoretical analysis of the PR controller is discussed and validated through simulations conducted in MATLAB/SIMULINK. The phase-locked loop (PLL) is used for synchronization, enabling accurate phase detection of the grid voltage for effective power injection. An LCL filter is also implemented between the inverter and the grid. Simulation results confirm the effectiveness of the proposed control scheme.**

Keywords* — *PV, MPPT, Controller PR, LCL filter, Lyapunov stability, PLL, Grid, PFC

I. INTRODUCTION

Renewable energies are emerging as a potential solution to reduce pollution. In particular, photovoltaic (PV) solar energy has been growing rapidly in recent years because it is an inexhaustible, environmentally friendly, and silent source. The rapid development of solar energy is gradually being implemented in small power installations connected to the Low Voltage (LV) network up to solar farm installations directly connected to the Medium Voltage (HTA) network.

However, the connection of PV systems to the distribution network can have some impacts on the electrical networks such as the change in power flows (bidirectional), on the voltage plan, on the protection plan, on the quality of energy and on the network planning which can influence the operation of PV systems. A grid-connected PV system, whether three-phase or single-phase, can be divided into two subsystems: the power subsystem consisting of PV panels, converters (chopper and inverter) and a filter and the control subsystem comprising a maximum power point tracking (MPPT) algorithm and other control blocks of the converters used. In general, the first subsystem is controlled to extract the maximum power, regardless of weather conditions and ensure power factor correction in order to minimize current harmonics and optimize the flow of active power between the power source and the grid [1]. In [2], [3] many control methods are proposed. Indeed, [4] proposed a controller based on the passivity technique, the "Proportional Resonant (PR)" controller and the "Multi Resonant control (MRC)" controller were developed in [5]. Besides classical methods, there are control techniques that do not require mathematical models such as genetic algorithms, fuzzy logic and artificial neural networks.

Generally, grid-connected photovoltaic (PV) systems, depending on their configuration, mainly consist of a DC/DC converter and a DC/AC inverter. Regardless of whether the DC/AC inverter is single-phase or three-phase, it is considered the core component of the system due to its crucial role in grid-connected operation. Typically, the current control principle is preferred because of its excellent dynamic characteristics and built-in overcurrent protection. Several control methods for these inverters have been proposed in the literature [6]. Predictive control relies on an accurate system model and the prediction of reference current [7]. While hysteresis control is simple and robust, it has significant drawbacks, such as variable switching rates, current errors twice the hysteresis band, and limited high-frequency operation [8].

This paper investigates the control of a photovoltaic system connected to a single-phase grid via a single-phase inverter that operates with a PWM technique and is followed by an LCL filter before connecting to the grid. The system might use a transformer for adaptation, as mentioned in [9]. The study has three main objectives : (i) regulate the voltage at the photovoltaic panel terminals to extract maximum power, (ii) control the voltage at the DC bus terminals, and (iii) inject a sinusoidal current into the grid that is in phase with the grid voltage. The entire system under study is depicted in the block diagram shown in Fig. 1.

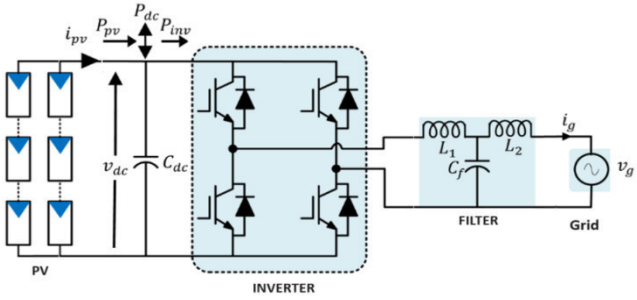


Fig. 1. Basic diagram of PV System connected Grid.

The power-voltage (P-V) characteristic of a photovoltaic system is nonlinear and changes with varying ambient

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

conditions such as solar irradiation and temperature. To continuously track the maximum power on this characteristic, the outer loop's first subsystem is the Maximum Power Point Tracking (MPPT) controller. Various MPPT algorithms have been proposed in the literature, with the Perturbation and Observation (P&O) method [10] being a commonly used approach. This method involves perturbing the operating voltage of the photovoltaic generator, causing the system to oscillate around the maximum power point (MPP). The size of these oscillations can be minimized by reducing the perturbation step size, which helps in accurately determining the reference voltage V*dc during its periodic operation, as outlined in the flowchart in Fig. 2.

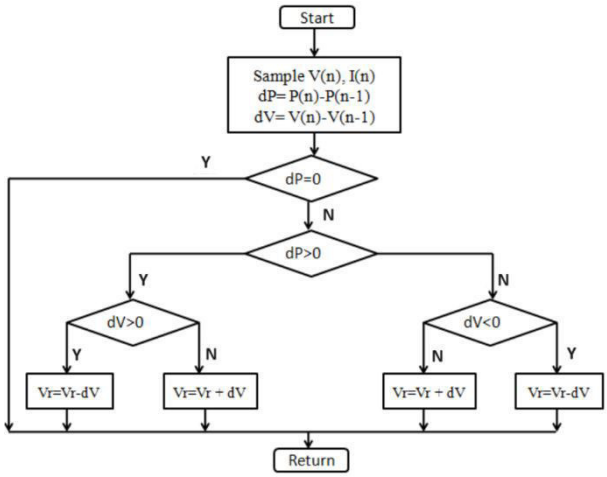


Fig..2. MPPT Algorithm Flowchart (P&O).

The MPPT block, combined with a PI-type intermediate circuit voltage controller, constitutes the outer loop of the system. This outer loop determines the grid reference current Iref needed to control the photovoltaic (PV) string to operate at the reference voltage V*dc. The inner loop, on the other hand, is designed to control the grid current ig with iinv [11].

To achieve zero steady-state error, a Proportional-Resonant (PR) controller is used in place of the traditional PI controller for current control. The PI controller struggles with tracking a sinusoidal reference with zero steady-state error and has limited disturbance rejection capabilities. The PR controller addresses these limitations by providing infinite gain at the fundamental frequency, which enhances performance in both tracking sinusoidal references and rejecting disturbances [9].

While the DC link capacitor plays a crucial role in energy transfer from the DC side (PV string) to the AC side (grid) by acting as a decoupling storage element [12], the output filter of the converters (on the grid side) typically has much less stored energy. Consequently, its impact on energy transmission can be considered negligible [13]. For stability analysis, the Lyapunov function is based solely on the variation of energy stored in the DC link capacitor, which must be sufficiently large. The authors have also adapted this approach for a three-phase photovoltaic inverter, which requires a smaller DC link capacitor since the power on the DC side does not exhibit oscillations [14].

In summary, it is evident from existing references that DC bus voltage loop control has garnered significant attention. However, an aspect that seems to have been overlooked in the literature is that the error signal εdc = (Vref - Vdc) at the input of the DC link voltage controller is inverted (see Figure 2) compared to conventional control loops. This inversion is crucial ; otherwise, the system might remain stuck at the open-circuit voltage of the PV string at startup and fail to reach the maximum power point (MPP).

This paper contributes by proposing an analysis of power flow and stability for grid-connected single-phase photovoltaic (PV) inverters utilizing a single DC/AC converter. The paper is organized as follows: 1. Introduction, 2. Control Strategy, 3. Stability Analysis, 4. Simulation Results and Discussion, and 5. Conclusion.

II. Control Strategy

A. Power Flow Description

As shown in Figure 1, the single-phase photovoltaic inverter is considered to have a single DC/AC stage connected to the grid, in addition it is assumed that : (i) there are no power losses in the DC/AC converter, in the DC bus capacitor and in the output filter; (ii) the other terms representing the conduction losses can be neglected to obtain a simple expression.

The equation describing the power balance in the DC link can be written as follows:

$$P_v(t) = P_{dc} + P_g \tag{1}$$

as : $P_{inv} = P_g$ The power in the grid is given by :

$$p_g = v_g(t) i_g(t) = V_g I_g (1 - \cos 2\omega t) \tag{2}$$

Where v_g and i_g are the instantaneous grid voltage and grid current respectively. I_g and V_g are the root mean square (RMS) of the injected current and the grid voltage respectively, and ω is the grid pulsation.

The power injected into single phase grid Pgrid, calculated as in (2), follows a sinusoidal waveform with twice the grid frequency. The PV generator could not be operated at the MPP if this pulsating power is not decoupled by means of an energy buffer.

Therefore, a capacitor bank is typically used for buffering this energy. The $Pc(t)$ is the instantaneous power flow in the DC link capacitor and is equal to:

$$P_{dc}(t) = C_{dc} V_{dc} \frac{dV_{dc}}{dt} \tag{3}$$

The average value of the power of the electrical network is given by :

$$P_{gav} = V_g I_g \tag{4}$$

B. Proposed Controller

The system controller under study can be divided into three parts: i) the DC link voltage controller, ii) the grid current controller, iii) and the grid synchronization controller. The DC bus voltage controller is used to regulate the DC voltage to a desired level in order to extract maximum power from the photovoltaic array.

Since the PV characteristic is non-linear, when the operating point changes due to a change in irradiance, the DC power changes accordingly, and consequently the MPP will also

change. As the grid voltage is fixed in our case, the power transferred to the grid is controlled solely by the output current of the single-phase inverter. The mathematical model describing the power flow transfer dynamics is given by (5), where P*g and I*g represent the reference grid power and current respectively and subsequently I*inv of the inverter :

$$P_{ref} = V_g I_{gref} = V_{pv} I_{pv} - C_{dc} V_{dc} \frac{dV_{dc}}{dt} \quad (5)$$

Where vpv and ipv are periodic signals having the average components $\bar{V}_{pv}$, $\bar{I}_{pv}$ and the AC components $\tilde{v}pv$, $\tilde{i}pv$ with oscillation period $T=1/(2fgrid)$, and defined respectively as :

$$\begin{cases} V_{pv} = \bar{V}_{pv} + \tilde{v}_{pv} \\ I_{pv} = \bar{I}_{pv} + \tilde{i}_{pv} \end{cases} \quad (6)$$

The power injected into the grid Pav is controlled according to (7), using a term Ppv and a second term kdP/dV that takes into account the system dynamics of the photovoltaic inverter :

$$P_{gref} = V_g I_{gref} = P_{pv} + k \frac{dP}{dV} \quad (7)$$

And by substituting (8) into (5) we get :

$$C_{dc} V_{dc} \frac{dV_{dc}}{dt} = -k \frac{dP}{dV} \quad (8)$$

Based on (8), we can deduce that dP/dV is related to the dynamics of the DC link capacitor.

The method for estimating the AC components (ripples) of PV power $\tilde{p}_{pv}$ and DC link voltage $\tilde{v}_{dc}$ is based on a quadrature signal generator based on a second-order generalized integrator operating as a pi/2 phase shifter (SOGI-QSG), as shown in Fig. 3 and defined by the closed-loop transfer function in (10) :

$$k_p = 12; \omega_n = 100\pi$$

Where ω_n represents the resonance frequency, equal to double the grid frequency.

The computed value of the reference current is determined using the PV Power reference generated by the proposed controller as shown in Fig. 3 and the RMS value of the grid voltage. The PLL extracts the phase angle of the grid voltage and is multiplied by the magnitude of the reference inverter current. Then, the current controller is proceeding to regulate the grid current ig.

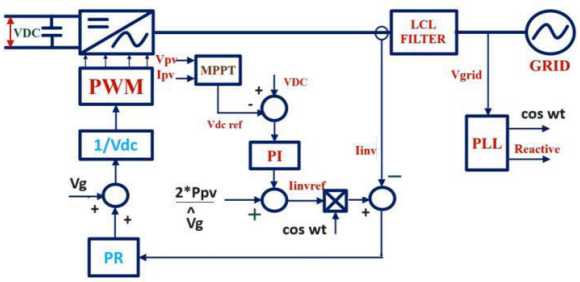


Fig. 3. Shows the structure of the controllers controlling the inverter.

C. Synchronization with the PLL grid

Connecting the PV source to the electrical network through the DC/AC converter requires synchronization of the voltage produced with that of the network, knowledge of the phase and frequency information of the network voltage is therefore essential.

The most efficient method is the use of a PLL (Phase Lock Loop), single-phase PLLs are subject to an additional difficulty than in three-phase, because of the smaller information field (only one phase). To create two orthogonal signals in single-phase, it is necessary to add a "Quadrature" demodulation block, Fig. 5.

The following control loops are provided :

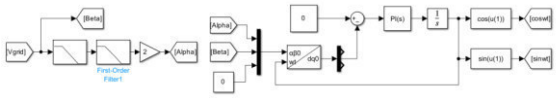


Fig. 4. Basic structure of a single-phase PLL.

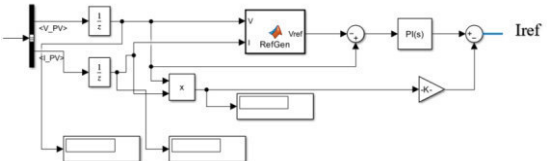


Fig.. 5. MPPT Control and reference current Iref generation loop.

D. Transfer function of the resonant controller

The transfer function of the resonant controller is given by :

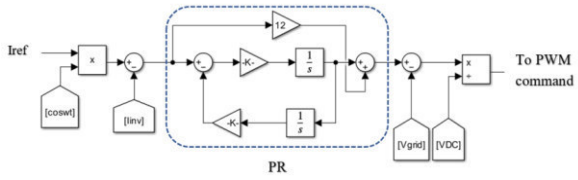


Fig..6. PWM command generation loop.

$$T(\omega) = \frac{k_p s^2 + 2k_i s + k_p \omega_n^2}{s^2 + \omega_n^2}$$

With : $k_p = 12$, $k_i = 200$, $\omega_n = 100\pi$

The amplitude and phase Bode diagrams are represented by Fig. 7.

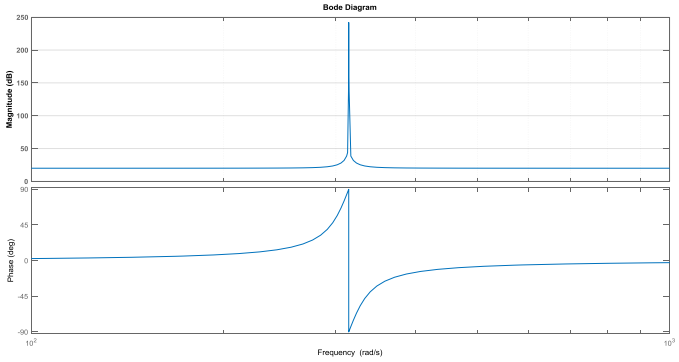


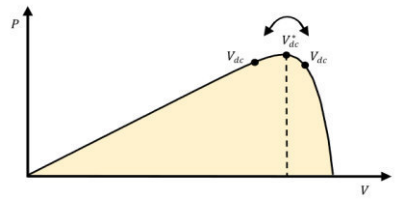
Fig.7. Bode response of the PR controller.

III. Stability Analysis

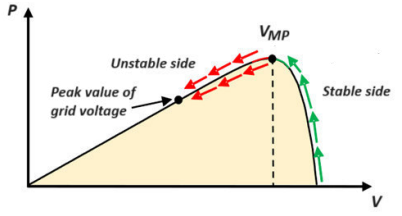
Lyapunov stability theory is generally used in the analysis of nonlinear systems. A Lyapunov function V (x) is a scalar energy that defines system states. Therefore, in this article, it will be formed on the basis of the variation of the energy stored in the intermediate circuit capacitor, assuming that the effect of other components such as the LCL filter is negligible and subsequently have no effect on the stability of the system. the candidate Lyapunov function can be written as follows:

$$V(E)=\frac{C_{dc}^2}{4}\left(V_{dc}^{*2}-V_{dc}^2\right)^2 \tag{11}$$

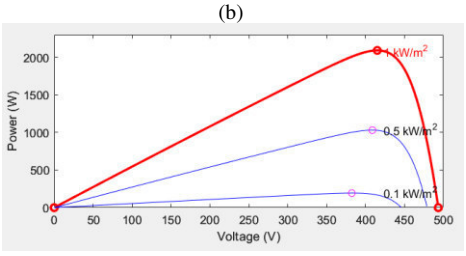
where V(E) is taken as a positive function, E refers to the energy stored across the intermediate circuit capacitor, V^*_{dc} is the reference voltage given by MPPT, shown in Fig. 4, Vdc is the actual circuit voltage measured on the DC line and it can be on either side of the P-V curve, as shown in Figure 3.



(a)



(b)



(c) Series connected Modules = 11 : Zytech Engineering Technologie ZT190S

Fig. 8. Power curve as a function of voltage (P-V).

On the other hand, according to [1], a system is globally stable if V(E) achieves the succeeding conditions as.

$$V(0)=0 \tag{12}$$

$$V(E)>0 \text{ for all } V(E)\neq 0 \tag{13}$$

$$V(E)\to\infty \text{ as } |E|\to\infty \tag{14}$$

$$\frac{dV(E)}{dt}<0 \text{ for all } |E|\neq 0 \tag{15}$$

Note that the last condition implies that the system is globally stable if all stored energy is continuously dissipated (V(E)×(dV(E)/dt) < 0), meaning that the system is constantly sliding towards a point of equilibrium.

Consequently, the chosen Lyapunov function satisfies the properties described by the preceding conditions and its derivative dV(E)/dt is given by (11) in the following form:

$$\frac{dV(E)}{dt}=-C_{dc}^2V_{dc}\frac{dV_{dc}}{dt}\left(V_{dc}^*-V_{dc}\right)\left(V_{dc}^*+V_{dc}\right) \tag{16}$$

Substituting (5) into (11) gives :

$$\frac{dV(E)}{dt}=-C_{dc}\left(P_{gref}-P_{pv}\right)\left(V_{dc}^*-V_{dc}\right)\left(V_{dc}^*+V_{dc}\right) \tag{17}$$

In fact, on the basis of (17), the sign of the derivative of the Lyapunov function can be determined. Clearly, it depends on the power variation in terms of voltage difference. The convergence of system stability is evaluated in two cases, as in [2], the system is stable.
The reference current is given by:

$$I_{invref}=2\frac{P_{pv}+k\left(-V_{dc}^*+V_{dc}\right)}{V_g} \tag{18}$$

IV. Simulation Results and Discussion

The control strategy used is illustrated in Fig. 5 for the single-phase grid-connected PV inverter. The PV system parameters used in the MATLAB/SIMULINK simulation are summarized in Table I. It is worth noting that the selected parameters for the P&O are : f_{MPPT} = 1MHz for tracking frequency, and ΔVMPPT = 0.5 V for increment voltage, in order to reduce the oscillations around the MPPT operating phase and thus the steady-state losses [3]. Fig. 10 shows the DC link voltage start-up waveform, where it is seen that the MPPT algorithm starts tracking the MPP on the right side once the DC link capacitor is fully charged, its value being equal to the open-circuit voltage Voc of the PV string. Then, the DC link capacitor voltage oscillates around the MPP in steady-state.

Simulation parameters :
PV : Voc=44.86V, Isc=5.5A, V_{MPP}=37.73V
LCL Filter : L_1= 4 mH, L_2=4.3 mH, C=6.25 µF
PWM : f= 10kHz
Power (kW) : 2 kW
Grid : Vmax=325V, F=50Hz

P-V curves are recorded for irradiations of 1000W/m^2, 500W/m^2, and 100W/m^2 under a temperature of 25°C (STC conditions) , Fig. 8.

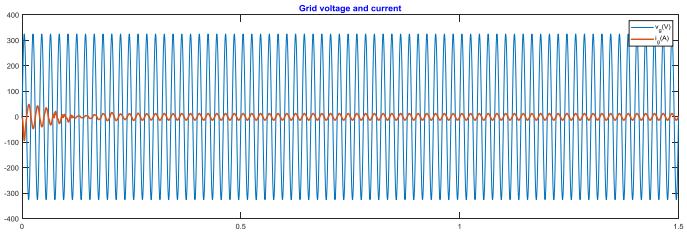


Fig. 9. Voltage and current in the grid.

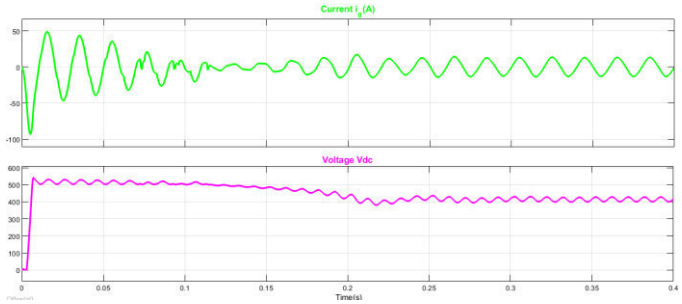


Fig. 10. Current Grid and DC Link Voltage.

Figs. 9, 10 and 11 show that the system is stable, the voltage and current in the network are in phase after an elapsed time of almost 0.2 seconds, the V_{dc} voltage starts from zero but tends towards a stable value of 425 V at time 0.2 seconds, this voltage presents some reduced ripples that do not present any problem.

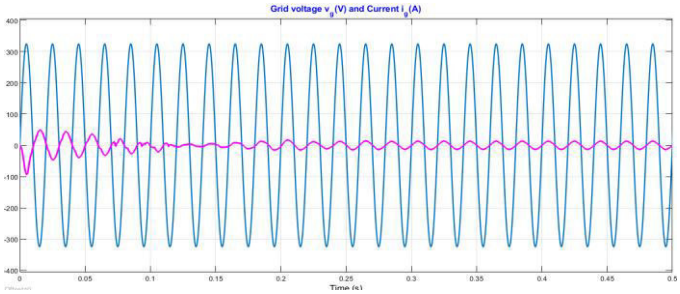


Fig. 11. Voltage and Current Grid.

At the initialization phase of the MPPT controller and before reaching the maximum power, we notice that the electrical network operates as a receiver, which is justified by a phase opposition between the current and the voltage of the network taken in the receiver convention, but this lasts almost 0.2 seconds.

For a variable irradiation profile ranging from 300W/m^2 to 1000W/m^2 and at a temperature of 25°C, we note the following curves:

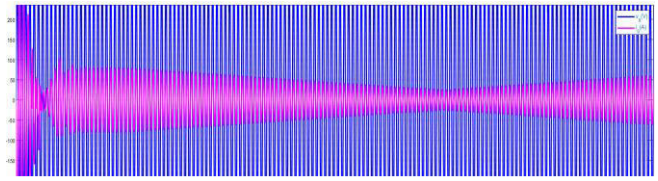


Fig. 12. Voltage and Current Grid for a variable irradiation profile.

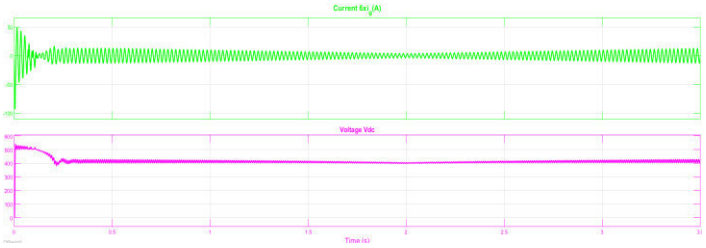


Fig. 13. Vdc voltage of the connection line CC and ig current Grid for a variable irradiation profile.

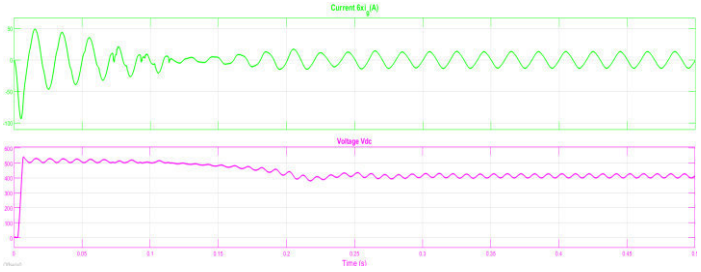


Fig. 14. Vdc voltage of the connection line and ig current Grid for a limited window of 0.5 s.

In the case of a variable irradiation profile, it is clear from Fig. 13 and Fig. 14 that the Vdc voltage tends towards an almost constant value with a possible ripple, it tends towards a value of 425 V, on the other hand the intensity ig(t) of the network current varies and follows the evolution of the irradiation profile as shown in figures Figs 12- 14.

V. CONCLUSION

This paper investigates a single-phase grid-connected photovoltaic (PV) system utilizing a single-phase inverter. The study analytically demonstrates that the nonlinearity of the power-voltage (P-V) curve of the PV string significantly influences system stability. The analysis employs a Lyapunov function based on the energy variation in the DC-link capacitor, with its derivative indicating changes in stored energy. In the absence of a voltage controller, stability is typically observed on the right side of the maximum power point (MPP) on the P-V curve (Vpv > VMP), while the left side (Vpv < VMP) remains unstable. Integrating a PI controller to directly manage the energy via the DC-link capacitor ensures overall system stability, irrespective of the operating point's location.

A theoretical analysis was performed to enhance system performance using a single-phase DC/AC inverter connected to the grid. Simulations in MATLAB/SIMULINK with a power rating of 2 kW were conducted and validated, demonstrating satisfactory performance. The results show that a PR regulator can address the limitations of the PI regulator, such as its inability to follow a sinusoidal reference with zero steady-state error and its poor disturbance rejection capability. The stability analysis indicates that instability arises from the inverse error signal at the input of the DC voltage regulator.

Future work will focus on exploring a single-stage three-phase PV system with a DC/AC inverter connected to the electrical grid, as discussed in other articles.

REFERENCES

[1] R. Au et al., Commande Non Linéaire Avancée d'un Système Photovoltaïque Connecté au réseau. 2016.

[2] A. E. Fadili, F. Giri, A. E. Magri, " Reference voltage optimizer for maximum power point tracking in triphase grid- connected photovoltaic systems " International Journal of Electrical Power & Energy Systems, vol. 60, pp. 293-301, sept. 2014.

[3] C. Aouadi, A. Abouloifa, A. Hamdoun, Y. Boussairi, " Advanced nonlinear control of photovoltaic system connected to the grid ," in *2014* Second World Conference on Complex Systems (WCCS), nov. 2014, pp. 61-66.

[4] M. Barghi Latran, A. Teke, " Investigation of multilevel multifunctional grid connected inverter topologies and control strategies used in photovoltaic systems ," Renewable and Sustainable Energy Reviews, vol. 42, pp. 361-376, févr. 2015.

[5] C. Aouadi, A. Abouloifa, A. Hamdoun, and Y. Boussairi, " Nonlinear controller design for single-phase grid-connected photovoltaic systems ," in 2015 3rd International Renewable and Sustainable Energy Conference (IRSEC), dec. 2015, pp. 1-5.

[6] H. Cha, T. Vu, and J. Kim, " Design and control of Proportional-Resonant controller based Photovoltaic power conditioning system ," in 2009 IEEE Energy Conversion Congress and Exposition, San Jose, CA: IEEE, sept. 2009, pp. 2198-2205.

[7] D. G. Holmes and D. A. Martin, " Implementation of a direct digital predictive current controller for single and three phase voltage source inverters ," in IAS '96. Conference Record of the 1996 IEEE Industry Applications Conference Thirty-First IAS Annual Meeting, San Diego, CA, USA: IEEE, 1996, pp. 906-913.

[8] A. Tilli , A. Tonielli, " Sequential design of hysteresis current controller for three-phase inverter ," IEEE Trans. Ind. Electron., vol. 45, n° 5, pp. 771-781, oct. 1998.

[9] F. E. Aamri et al., " Stability Analysis for DC-Link Voltage Controller Design in Single-Stage Single-Phase Grid-Connected PV Inverters ," IEEE J. Photovoltaics, vol. 13, n° 4, pp. 580-589, july 2023.

[10] A. Kabba, H. El Fadil, A. Yahya, I. Bentalhik, " PQ Theory of SAP Filter in Photovoltaic System Connected to Three-Phase Grid and Controlled Nonlinear ," E3S Web Conf., vol. 351, pp. 1-7, 2022.

[11] M. Rajeev, V. Agarwal, " Analysis and Control of a Novel Transformer-Less Microinverter for PV-Grid Interface ," IEEE J. Photovoltaics, vol. 8, n° 4, pp. 1110-1118, july 2018.

[12] Y. Sun, Y. Liu, M. Su, X. Wenjing, J. Yang, " Review of Active Power Decoupling Topologies in Single-Phase Systems ," IEEE Transactions on Power Electronics, vol. 31, pp. 1-11, jan. 2015.

[13] F. El Aamri, H. Maker, D. Sera, S. V. Spataru, J. M. Guerrero, A. Mouhsen, " A Direct Maximum Power Point Tracking Method for Single-Phase Grid-Connected PV Inverters ," IEEE Trans. Power Electron., vol. 33, n° 10, pp. 8961-8971, oct. 2018.

[14] K. Ma, W. Chen, M. Liserre, F. Blaabjerg, " Power Controllability of a Three-Phase Converter With an Unbalanced AC Source ," IEEE Transactions on Power Electronics, vol. 30, n° 3, pp. 1591-1604, March 2015.

[15] H. Komurcugil, N. Altin, S. Ozdemir, I. Sefa, " An Extended Lyapunov-Function-Based Control Strategy for Single-Phase UPS Inverters ," IEEE Transactions on Power Electronics, vol. 30, p. 1-15, jan. 2014.

Multilevel Inverter Topologies: A Case Study of 9-Level Cascaded H-Bridge Inverter

Salah Anis KRIM, Fateh KRIM
Department of Automation
Laboratory of Power Electronics and Industrial Control
University of Setif - 1
Setif, Algeria
krimsalahanis@gmail.com,
krim_f@ieee.org

Hamza AFGHOUL
Department of Automation
Laboratory of Power Electronics and Industrial Control
University of Setif - 1
Setif, Algeria
hamza.afghoul@gmail.com

***Abstract* — Renewable energy sources and electric vehicles are widely utilized, and their adoption is expected to increase steadily in line with the trend towards using clean, green, and non-polluting energy. Effectively harnessing and utilizing these integrated energy systems requires the involvement of energy converters. Currently, multilevel energy conversion systems play an indispensable role in enhancing energy conversion efficiency and forming a complete system for efficient energy utilization. Therefore, multilevel conversion systems are receiving significant attention and research development from scholars. This paper presents advancements in multilevel energy conversion technology. The fundamental structure of multilevel converters, along with the advantages and disadvantages of each technique, is discussed with a primary focus on modern industrial and practical applications. A case study of a typical 3-phase 9-level cascaded inverter is employed to illustrate through simulation analysis. Results from this case study provide readers with a clear and in-depth understanding of multilevel converters, particularly the current interest in multilevel stacked structure configurations.**

Keywords — Multilevel inverter topologies, Cascaded inverter, 9-level inverter, Energy conversion, Efficient energy utilization

I. INTRODUCTION

Multilevel voltage source inverters present a cost-effective solution in the current energy management market [1]-[2]. These multilevel converters have been extensively researched and widely applied in various industrial fields, including manufacturing, power transmission, energy storage systems for renewable energy sources, motor control systems, and electric vehicles [3]-[5]. In recent years, industrial applications have increasingly demanded higher power devices. Moreover, certain motor drive systems require medium voltage levels, and utility applications demand medium voltage and high power levels, reaching up to several megawatts [6]. Connecting a single semiconductor switch directly to the medium voltage grid is challenging. Consequently, multilevel power converter structures have emerged as a viable alternative for high and medium voltage applications. These converters not only support high power levels but also accommodate the integration of emerging renewable energy sources. Systems such as photovoltaics, wind turbines, and fuel cells can be seamlessly incorporated into multilevel converter setups for high-power applications [7]-[8]. Moreover, for high-quality motor control, a multilevel inverter configuration is essential to achieve optimal output quality [9]-[10].

Research on multilevel converters was introduced in the early 1970s. Initially, the term "multilevel" referred to three-level converters, and numerous multilevel converter structures were developed in the subsequent years [11]-[13]. The fundamental concept behind multilevel converters for achieving higher power ratings involves using multiple power semiconductor switches with several lower voltage DC sources to synthesize a stepped voltage waveform. Capacitors, batteries, and renewable energy sources can serve as versatile DC voltage sources. The switching of these synthesized DC sources allows for the achievement of higher output voltages [14]. It is important to note that the rated output voltage of power semiconductor switching devices is entirely dependent on the rated voltage of the DC source to which they are connected. Multilevel converters offer distinct advantages over conventional two-level converters utilizing high-frequency pulse width modulation (PWM). These benefits include superior staircase waveform quality, capable of producing low-distortion output voltages while mitigating dv/dt stresses, thereby reducing electromagnetic compatibility concerns [15]. They also generate lower common-mode voltages, alleviating stress on motor shaft bearings in multilevel motor drive systems, and can minimize input current ripple [16]. Multilevel converters operate efficiently across both fundamental and high-frequency PWM switching frequencies, where lower switching frequencies typically translate to reduced losses and increased overall efficiency [17]. Despite these advantages, they necessitate a greater number of high-power semiconductor switches, each requiring dedicated gate drive circuits, thereby potentially escalating system complexity and cost. Over the past two decades, diverse configurations of multilevel converters have been proposed and extensively studied. Modern research has explored novel converter topologies and unique modulation schemes. Furthermore, three primary multilevel converter structures have been documented: the neutral point clamped converters, flying capacitor converters, and cascaded H-bridge converters [18]-[20]. Additionally, sophisticated modulation techniques and control models have been developed for multilevel converters, including selective harmonic elimination PWM, sinusoidal pulse width modulation, space vector modulation, and others [21]. Moreover, multilevel converter applications

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

have spanned various domains, focusing on industrial motor drive systems [22], renewable energy integration systems [23], flexible alternating current transmission systems [24], and electric vehicle systems [25].

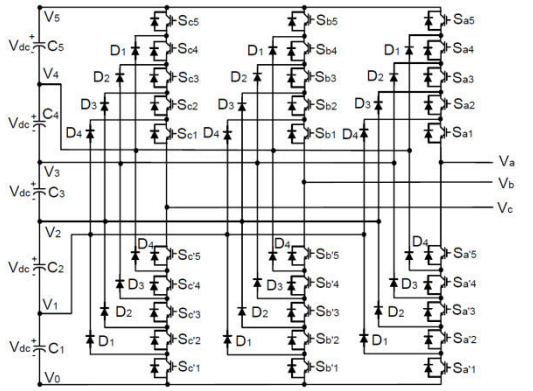
This study assesses the advanced technology of multilevel energy conversion, focusing on multilevel inverters. An overview of fundamental multilevel converter structures will be discussed, encompassing the merits and drawbacks of each technique. Special emphasis will be placed on their practical applications in modern industry and the pragmatic utility of multilevel converters. Potential future developments in multilevel conversion technology will also be highlighted. Finally, a case study of a 9-level inverter will be conducted through simulations. The results of this study provide valuable insights into the control techniques of cascaded multilevel converters.

II. Overview Of Multilevel Inverter Structures

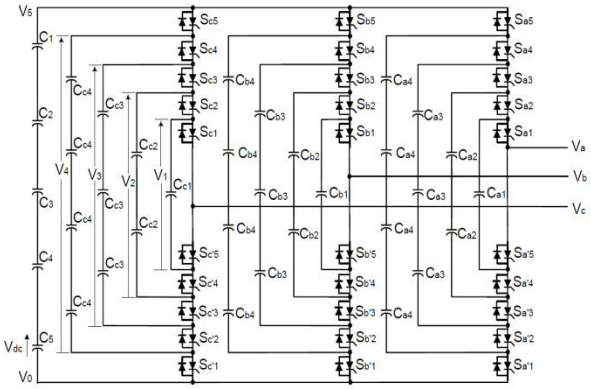
A. Neutral Point Clamped Inverter Structure

The neutral point clamped multilevel inverter (NPCM) [18] was first introduced by Nabae, Takahashi, and Akagi in 1981 as a three-level diode clamped inverter [13]. During the 1990s, numerous researchers published studies presenting experimental results for diode clamped converters with levels ranging from four to six. These converters were applied in various fields such as static var compensation, speed control for electric motors, and high-voltage power transmission systems [26]-[27]. The NPCM effectively utilizes DC sources derived from AC power systems. A multilevel inverter contains pairs of clamping diodes and a DC source divided into smaller voltage levels through a series of capacitors connected in series. A three-phase six-level diode clamped inverter is illustrated in Figure 1(a). Each phase of the inverter shares a common DC bus, divided into five capacitors forming six levels. The voltage across each capacitor is V_{dc}, and the voltage stress on each switching device is limited to V_{dc} through the clamping diodes. Multilevel diode clamped inverters are utilized in various applications. One significant application of multilevel converters is their use in interfacing high-voltage DC transmission lines with AC transmission systems [28]-[29]. These converters are also utilized in variable speed drives for medium-voltage, high-power motors, which typically operate within a voltage range of 2.4 kV to 13.8 kV, as indicated by various studies [30]. Additionally, static var compensation has been proposed as another potential application for these converters by some researchers [31]. The following sections of this paper provide a comprehensive discussion of the key advantages and disadvantages associated with multilevel diode clamped converters.

Remark 1 : All phases share a common DC bus, minimizing the capacitance needs of the converter and making the back-to-back configuration feasible and practical for applications like high-voltage connections or adjustable-speed drives. Capacitors can be pre-charged in groups, ensuring high efficiency in fundamental frequency conversion. Transmitting practical power through a single inverter can be challenging because the intermediate DC voltage tends to become overloaded or discharged if not precisely monitored and controlled. The number of clamping diodes required is proportional to the number of levels in the inverter, which can complicate the design for high-level devices.



(a)



(b)

Fig.1. Typical structure of diode clamped multilevel inverter and flying capacitor multilevel inverter. (a) Diode clamped multilevel inverter; (b) Flying capacitor multilevel inverter with six levels.

B. Flying Capacitor Multilevel Inverter

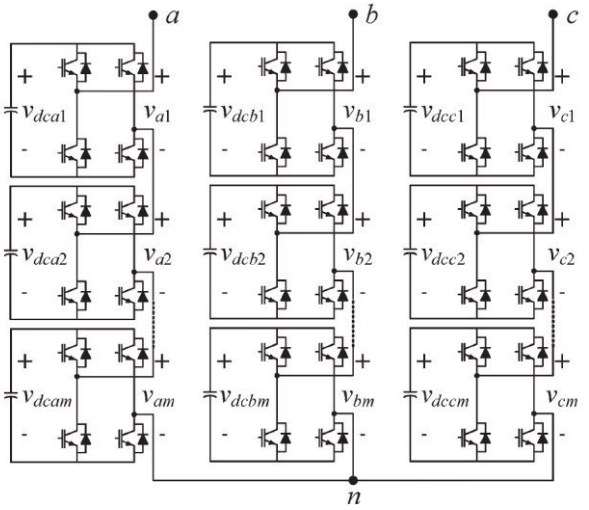
The flying capacitor multilevel inverter (FCM) [19], introduced by Meynard and colleagues in 1992 [32], presents a structure akin to that of the diode clamped inverter, but utilizes capacitors in place of clamping diodes. The circuit topology of the FCM is illustrated in Figure 1(b). This configuration is characterized by a staircase arrangement of capacitors on the DC side, with each capacitor having a different voltage from its neighboring capacitors. The voltage step between two adjacent capacitor terminals determines the voltage steps in the output waveform. A notable advantage of the FCM is its ability to redundantly combine internal voltage levels, allowing multiple valid switch combinations to synthesize the output voltage. Unlike the diode clamped inverter, the FCM does not require all conducting switches to be in series and offers phase redundancy, whereas the diode clamped inverter provides only line redundancy [19,32]. These redundancies enable selective capacitor charging and discharging and can be integrated into control systems to balance voltages across different levels. Proposed applications for the FCM include traction systems, reactive power compensators, electrical

propulsion systems, and the integration of renewable energy sources [33].

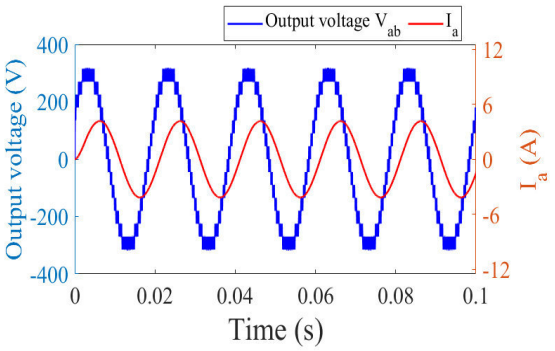
Remark 2 : The FCM offers inherent phase redundancy to balance capacitor voltage levels. Each branch can be independently analyzed without affecting others. Unlike the NPC-type multilevel inverters where voltage balancing across the three phases at the input must be considered, here, active and reactive power flows can be controlled. The large number of capacitors allows the inverter to operate during short-term power loss and deep voltage sags. Monitoring voltage levels across all capacitors in complex systems requires intricate control mechanisms. Furthermore, achieving uniform voltage levels across all capacitors and initiating them simultaneously is highly complex. The efficiency of utilization and switching for real power transmission tends to be suboptimal. The increased number of capacitors not only raises costs but also adds to the bulkiness of the system compared to diode clamped inverters in multilevel configurations. As the number of levels increases, packaging becomes more challenging. Additionally, the presence of numerous power capacitors in the circuit escalates costs and reduces overall reliability.

C. Cascaded H-Bridge Multilevel Inverter

Among the various multilevel inverter structures, the cascaded H-bridge multilevel inverter (CHBM) [20] has garnered significant attention from researchers due to its advantageous features. The CHBM is favored for its simple modular design, which facilitates the generation of multiple voltage levels and results in smoother output waveforms. This leads to reduced harmonic distortion compared to other multilevel structures, such as neutral-point-clamped inverters and flying capacitor inverters. Additionally, the modular structure of the CHBM allows for operation at higher voltages [5,20]. The CHBM synthesizes the desired voltage using multiple independent DC voltage sources, making it versatile for a range of applications, including low-power inverters, high-voltage transmission systems, industrial applications, renewable energy systems, and motor drive systems [9,17,34]. The modularity and flexibility of the CHBM design enable easy scalability to meet specific voltage and power requirements. This design eliminates the need for additional clamping diodes or voltage-balancing capacitors, simplifying the inverter setup. Moreover, the CHBM can operate efficiently in fault-tolerant modes; if one module fails, the remaining modules can continue to function independently. The CHBM also reduces dv/dt stress on insulated gate bipolar transistors (IGBTs), thereby enhancing the lifespan and reliability of the overall system [9]. The reliable, efficient, and adaptable nature of the CHBM makes it an optimal choice for improving the stability and reliability of various systems. Currently, among the discussed structures, the CHBM achieves superior.



(a)



(b)

Fig. 2. Multilevel cascaded inverter. (a) Configuration of three-phase multilevel cascaded H-bridge inverter from connected individual modules. (b) Output voltage waveform and load current of the CHBM.

TABLE I. COMPARISON OF COMPONENT COUNTS AMONG DIFFERENT STRUCTURES (IS THE NUMBER OF LEVELS).

Components of the inverter	Inverter topology types		
	NPCM	FCM	CHBM
IGBTs	$2.(n-1)$	$2.(n-1)$	$2.(n-1)$
Clamping diodes	$(n-1).(n-2)$	0	0
DC-link capacitors	$(n-1)$	$(n-1)$	$(n-1)/2$
Balancing capacitors	0	$(n-1).(n-2)$	0

voltage and power output levels, along with enhanced reliability, due to its modular architecture. The configuration of the CHBM and the output waveform are shown in Figure 2. The AC outputs from each inverter level are connected in series, combining to form a synthesized voltage waveform that represents the sum of the individual inverter outputs. The number of phase voltage levels, denoted as m, in a cascaded inverter is determined by the formula $(m = 1 + 2s)$, where s is the number of separate DC sources. For example, an 11-level cascaded H-bridge inverter with 5 separate DC

sources and 5 full bridges can be represented by the phase voltage $v_{an} = v_{a1} + \dots + v_{a5}$. The Fourier transform of a stepped waveform with s steps is given by

$$V(\omega t) = \frac{4V_{dc}}{\pi} \sum_{n} [cos(n\theta_1) + cos(n\theta_2) + \cdots + cos(n\theta_s)] \frac{\sin(n\omega t)}{n}$$

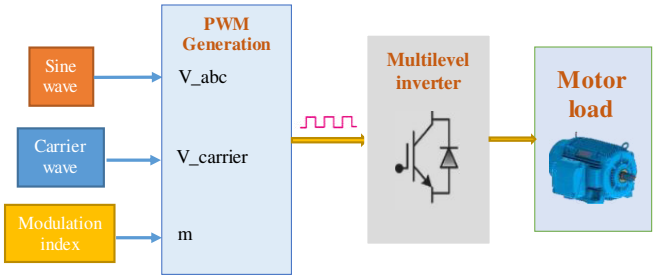
where $n = 1,3,5,7, \cdots$ (1)

From (1), the magnitudes of the Fourier coefficients, normalized with respect to V_{dc}, are as

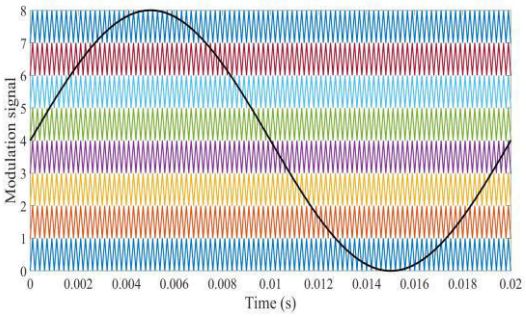
$$H(n) = \frac{4}{\pi n} [cos(n\theta_1) + cos(n\theta_2) + \cdots + cos(n\theta_s)]$$

where $n = 1,3,5,7, \cdots$ (2)

The conduction angles θ_1 , θ_2, ... , θ_s can be selected to minimize the total harmonic distortion (THD) of the output voltage. Generally, these angles are chosen such that the lower order harmonics, specifically the 5th, 7th, 11th, and 13th, are eliminated using(2). In practical industrial applications, cascaded multilevel inverters have been proposed for a variety of uses, including conversion systems for renewable energy sources, static var compensation, power transmission, battery management, motor control systems for electric vehicles, and industrial manufacturing applications. Their modular design makes them particularly suitable for integrating renewable energy sources, such as photovoltaic panels and fuel cells. Cascaded multilevel inverters are also proposed as traction drives in electric vehicles, where multiple batteries or supercapacitors serve as individual DC sources [9,35]. The inverter can function as both a rectifier and a charger for the vehicle's battery when connected to an AC power source. Additionally, it can serve as a rectifier in regenerative braking systems.



(a)



(b)

Fig.3. (a) General block diagram of an integrated multilevel inverter system; (b) Sinusoidal PWM technique for 9-level cascaded inverter.

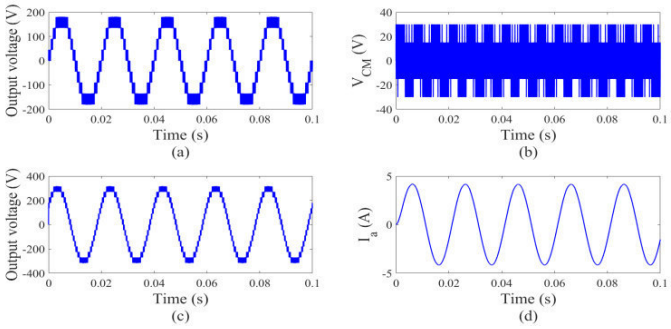


Fig.4. Simulation results with modulation index of 1. (a) Voltage waveform of phase a; (b) Common-mode voltage; (c) Line-to-line voltage V_{ab}; (d) Current of phase a.

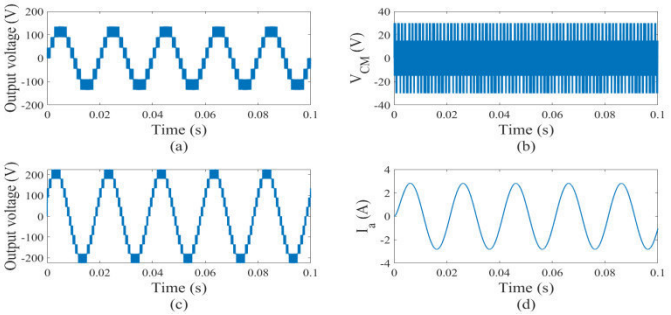


Fig. 5. Simulation results with modulation index of 2/3. (a) Voltage waveform of phase a; (b) Common-mode voltage; (c) Line-to-line voltage V_{ab}; (d) Current of phase a.

Remark 3 : The primary advantages of cascaded H-bridge multilevel inverters (CHBMs) can be summarized as follows. The number of output voltage levels achievable with a CHBM can exceed twice the number of DC sources, according to the formula $(m = 1 + 2s)$. Additionally, the series configuration of H-bridges supports a modular design and packaging approach, which facilitates faster and more cost-effective manufacturing processes. The circuit achieves DC source voltage balancing without the existence of circulating currents between modules. However, the circuit requires the use of output transformers. The main disadvantage of the CHBM is the need for separate DC sources for each H-bridge. This requirement limits its application to products that already have multiple separate DC sources. In summary, the cascaded inverter uses the fewest components. It should be noted that another advantage of the cascaded inverter is its flexible configuration, allowing the number of levels to be easily increased or decreased by adding or removing the appropriate number of single-phase inverter modules. As the power of the inverter increases, the voltage stress on the components decreases, reducing the switching losses. At the same switching frequency, the higher-order harmonic components of the output voltage are smaller compared to inverters with fewer levels. For high-power loads, the voltage supplied to the loads can reach relatively high values. Among the three multilevel inverter topologies mentioned above, each configuration has its own advantages and disadvantages. Table I shows the number of components used in these inverter structures, indicating that the cascaded inverter utilizes fewer components than the other two topologies. This significant advantage makes the

cascaded multilevel inverter more cost-effective compared to the other types.

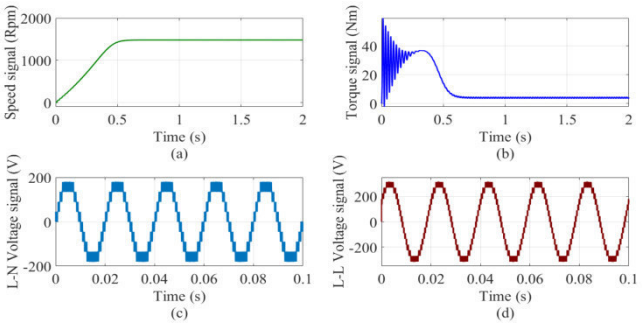


Fig. 6. Results with motor load. (a) Speed signal; (b) Torque signal; (c) L-N voltage signal; (d) L-L voltage signal.

III. Case Study Of 3 – Phase 9 – Level Cascaded H – Bridge Inverter

A. *Overview of the 3-Phase 9-Level Cascaded H-Bridge Inverter System*

In this section, the structure of a 3-phase 9-level H-bridge inverter is examined through both RL loads and motor loads. The simulation diagram of the 3-phase 9-level cascaded inverter is implemented using MATLAB/Simulink with sinusoidal PWM modulation technique. The main blocks of the 3-phase 9-level cascaded inverter model are illustrated in Figure 3(a), comprising the PWM pulse generation block, the power circuit block, and the load block. To create a gating signal for the components in the same phase, the carrier wave (triangular form) and the control signal (sinusoidal form) are used. Figure 3(b) illustrates the PWM technique used in the 9-level cascaded H-bridge inverter. The PWM pulses for the IGBT switches across all three phases of the inverter are generated by comparing the control waveform with the carrier waveform, both operating at a frequency of 5 kHz. The primary parameters used in this case study are detailed as follows: the DC link capacitor is 2200 µF, the DC source voltage is 45 V, and the grid frequency is 50 Hz. The switching frequency is set to 5 kHz, and the modulation index ranges from 0 to 1. Additionally, the motor has power of 2 HP, rated speed of 1500 rpm, and rated load torque of 4 Nm.

B. *Conducting an Analytical Study of a 3 – Phase 9 – Level Cascaded Inverter*

The simulation results for a modulation index $m = 1$ are presented in Figure 4. With $m = 1$, the waveforms for phase voltage (phase a), line-to-line voltage V_{ab}, common-mode voltage, and phase current (phase a) were obtained. With a chosen DC source voltage of 45 V, the theoretical values for phase and line voltages can be calculated. The peak phase output voltages are $V_{ao} = V_{bo} = V_{co} = 180V$ (calculated as $V_{out} = 4mV_{dc} = 4 \times 1 \times 45 = 180\ (V)$. The line voltage is $V_{ab} = 180\sqrt{3} = 312(V)$. Clearly, the simulation results align well with the theoretical expectations for a 9-level cascaded inverter. Further investigation was conducted with a modulation index of 2/3 of the standard modulation index. The simulation results for $m = 2/3$ are shown in Figure 5. With a modulation index m=1, the simulation of the 3-phase 9-level cascaded inverter achieves optimal performance. The 9-level output phase voltage waveform closely approximates a sine wave, significantly outperforming lower-level inverters. As the modulation index is reduced, the amplitude of the output voltage waveform decreases, and the number of voltage levels also decreases. Consequently, the steady-state load current in each phase is reduced. The simulation results confirm the accuracy of the theoretical analysis. An investigation with a motor load is conducted in this study. The results of this investigation are presented in Figure 6. As observed in Figure 6(a), the motor speed can achieve and maintain a stable rated speed of 1500 rpm. The electromagnetic torque generated by the motor is depicted in Figure 6(b), demonstrating its capability to drive the load and achieve a rated load torque of 4 Nm. Additionally, the output voltage values of the 9-level inverter, shown in Figures 6(c) and 6(d), indicate that the output voltage is sufficiently level and the model operates effectively.

IV. Conclusion

The growing emphasis on environmentally friendly green energy and advancements in power electronics technology have significantly propelled the development of energy converters. This study provides a comprehensive examination of recent advancements in multilevel energy conversion technology. The fundamental structure of multilevel converters, along with the advantages and limitations of various techniques, is thoroughly analyzed and evaluated. This discussion provides an overview of contemporary industrial applications of multilevel converters, offering insights for readers interested in understanding the practical uses and advantages of these power conversion technologies. A case study of a 3-phase 9-level cascaded inverter utilizing sinusoidal PWM modulation technique has been conducted within this research. The results from this case study offer clear insights into the characteristics of cascaded multilevel converters, which are increasingly being employed in contemporary applications.

References

[1] A. Arikesh and A. K. Parvathy, "Modular Multilevel Inverter for Renewable Energy Applications," International Journal of Electrical and Computer Engineering (IJECE), vol. 10, no. 1, pp. 1–14, Feb. 2020.pp.1-14.

[2] M. Trabelsi, H. Vahedi, and H. Abu-Rub, "Review on Single-DC-Source Multilevel Inverters: Topologies, Challenges, Industrial Applications, and Recommendations," IEEE Open Journal of the Industrial Electronics Society, vol. 2, pp. 112–127, 2021.

[3] M. S. H. Lipu, M. S. Miah, S. Ansari, S. T. Meraj, K. Hasan, R. M. Elavarasan, A. A. Mamun, M. A. A. M. Zainuri, and A. Hussain, "Power Electronics Converter Technology Integrated Energy Storage Management in Electric Vehicles: Emerging Trends, Analytical Assessment and Future Research Opportunities," Electronics, vol. 11, no. 4, pp.. 562-568, 2022.

[4] S. Kouro, M. Malinowski, K. Gopakumar, J. Pou, L. G. Franquelo, B. Wu, J. Rodriguez, M. A. Pérez, and J. I. Leon, "Recent Advances and Industrial Applications of Multilevel Converters," IEEE Transactions on Industrial Electronics, vol. 57, no. 8, pp. 2553–2580, Aug. 2010.

[5] V. Q. Nguyen and T. L. Le, "Flexible Control with Fuzzy Observer-Based Sliding Mode for Multilevel Inverter," Journal of Electrical Engineering & Technology, 2024.

[6] M. I. Daoud, A. A. Elserougi, A. M. Massoud, R. Bojoi, A. S. Abdel-Khalik, and S. Ahmed, "Zero-/Low-Speed Operation of Multiphase Drive Systems With Modular Multilevel Converters," IEEE Access, vol. 7, pp. 14353–14365, 2019.

[7] M. Sarebanzadeh, M. A. Hosseinzadeh, C. Garcia, E. Babaei, S. Islam, and J. Rodriguez, "Reduced Switch Multilevel Inverter Topologies for Renewable Energy Sources," IEEE Access, vol. 9, pp. 120580–

120595, 2021,
[8] H. Hachemi, A. Allali, and B. Belkacem, “Control of the Power Quality for a DFIG Powered by Multilevel Inverters,” International Journal of Electrical and Computer Engineering (IJECE), vol. 10, no. 5, pp. 4592–4603, Oct. 2020.
[9] V. Q. Nguyen and T. L. Le, “A Robust Control for Five-Level Inverter Based on Integral Sliding Mode Control,” International Journal of Integrated Engineering, vol. 15, no. 4, pp. 177–192, 2023.
[10] V. N. R, K. Gopakumar, and L. G. Franquelo, “A Very High Resolution Stacked Multilevel Inverter Topology for Adjustable Speed Drives,” IEEE Transactions on Industrial Electronics, vol. 65, no. 3, pp. 2049–2056, Mar. 2017.
[11] R. H. Baker and L. H. Bannister, “Electric Power Converter,” U.S. Patent 3,867,643, Feb. 1975.
[12] M. Carpita, M. Marchesoni, M. Pellerin, and D. Moser, “Multilevel Converter for Traction Applications: Small-Scale Prototype Tests Results,” IEEE Transactions on Industrial Electronics, vol. 55, no. 5, pp. 2203–2212, Oct. 2008.
[13] A. Nabae, I. Takahashi, and H. Akagi, “A New Neutral-Point-Clamped PWM Inverter,” IEEE Transactions on Industry Applications, vol. IA-17, no. 5, pp. 518–523, Sep./Oct. 1981.
[14] G. Grandi, C. Rossi, D. Ostojic, and D. Casadei, “A New Multilevel Conversion Structure for Grid-Connected PV Applications,” IEEE Transactions on Industrial Electronics, vol. 56, no. 11, pp. 4416–4426, Nov. 2009.
[15] X. Li, R. Liu, Z. Ke, J. Pan, B. Hu, R. Na, L. Xu, and J. Wang, “Simple Switching Strategies for dv/dt Reduction in SiC-Device-Based Modular Multilevel Converters,” IEEE Transactions on Power Electronics, vol. 38, no. 2, pp. 1485–1493, Feb. 2023.
[16] K. B. Tawfiq, M. Güleç, and P. Sergeant, “Bearing Current and Shaft Voltage in Electrical Machines: A Comprehensive Research Review,” Machines, vol. 11, no. 5, pp. 550-560, May 2023.
[17] M. Priya, P. Ponnambalam, and K. Muralikumar, “Modular-Multilevel Converter Topologies and Applications – A Review,” *IET* Power Electronics, vol. 12, no. 2, pp. 170–183, Feb. 2019.
[18] S. Alepuz, S. Busquets-Monge, J. Nicolás-Apruzzese, À. Filbà-Martínez, J. Bordonau, X. Yuan, and S. Kouro, “A Survey on Capacitor Voltage Control in Neutral-Point-Clamped Multilevel Converters,” Electronics, vol. 11, no. 4, pp. 527-233, Apr. 2022.
[19] M. Humayun, M. M. Khan, M. U. Hassan, and W. Zhang, “Analysis of Hybrid Switches Symmetric Flying Capacitor Multilevel Inverter Based STATCOM,” International Journal of Electrical Power & Energy Systems, vol. 131, pp. 1070-1084, Sep. 2021.
[20] F. A. da Costa Bahia, C. B. Jacobina, N. Rocha, I. R. F. M. Pinheiro da Silva, and R. P. R. de Sousa, “Hybrid Asymmetric Cascaded Multilevel Inverters Based on Three- and Nine-Level H-Bridges,” IEEE Transactions on Industry Applications, vol. 55, no. 6, pp. 6047–6060, Nov./Dec. 2019.
[21] N. Mittal, B. Singh, S. P. Singh, R. Dixit, and D. Kumar, “Multilevel Inverters: A Literature Survey on Topologies and Control Strategies,” in Proceedings of the 2012 2nd International Conference on Power, Control and Embedded Systems (ICPCES), Allahabad, India, Dec. 2012, pp. 1–6.
[22] G. Gontijo, S. Wang, T. Kerekes, and R. Teodorescu, “New AC–AC Modular Multilevel Converter Solution for Medium-Voltage Machine-Drive Applications: Modular Multilevel Series Converter,” Energies, vol. 13, no. 14, pp. 3664-3674, Jul. 2020.
[23] B. Sharma, R. Dahiya, and J. Nakka, “Effective Grid Connected Power Injection Scheme Using Multilevel Inverter Based Hybrid Wind Solar Energy Conversion System,” Electric Power Systems Research, vol. 171, pp. 1–14, Feb. 2019.
[24] M. Diaz, R. Cárdenas Dobson, E. Ibaceta, A. Mora, M. Urrutia, M. Espinoza, F. Rojas, and P. Wheeler, “An Overview of Applications of the Modular Multilevel Matrix Converter,” Energies, vol. 13, no. 21, p. 5546, Nov. 2020
[25] R. S. Tripathi and M. K. Thukral, “Switching Angles Computation of Multi-Level Inverter for Electrical Vehicle Application,” in Proceedings of the 2019 Global Conference for Advancement in Technology (GCAT), Bangalore, India, Oct. 2019, pp. 1–6.
[26] S. Ogasawara and H. Akagi, “Analysis of Variation of Neutral Point Potential in Neutral-Point-Clamped Voltage Source PWM Inverters,” in Proceedings of the 1993 IEEE Industry Applications Conference Twenty-Eighth IAS Annual Meeting, Toronto, ON, Canada, Oct. 1993, pp. 956–963.
[27] P. Li, L. Zhang, B. Ouyang, and Y. Liu, “Nonlinear Effects of Three-Level Neutral-Point Clamped Inverter on Speed Sensorless Control of Induction Motor,” Electronics, vol. 8, no. 4, pp. 402-408, Apr. 2019.
[28] S. L. Gundebommu, O. Rubanenko, and I. Hunko, “Analysis of Three-Level Diode Clamped Inverter for Grid-Connected Renewable Energy Sources,” in Proceedings of the 2019 IEEE 20th International Conference on Computational Problems of Electrical Engineering (CPEE), Lviv-Slavske, Ukraine, Sep. 2019, pp. 1–6.
[29] X. Zeng, D. Gong, M. Wei, and J. Xie, “Research on Novel Hybrid Multilevel Inverter with Cascaded H-Bridges at Alternating Current Side for High-Voltage Direct Current Transmission,” IET Power Electronics, vol. 11, no. 10, pp. 1635–1642, Oct. 2018.
[30] H. Tang and A. Q. Huang, “A Design Investigation of a 1 MVA SiC Medium Voltage Three Phase Rectifier Based on Isolated Dual Active Bridge,” in Proceedings of the 2018 IEEE Applied Power Electronics Conference and Exposition (APEC), San Antonio, TX, USA, Mar. 2018, pp. 2196–2203.
[31] A. Taghvaie, M. E. Haque, S. Saha, and M. A. Mahmud, “A New Step-Up Voltage Balancing Circuit for Neutral-Point-Clamped Multilevel Inverter,” in Proceedings of the 2018 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Chennai, India, Dec. 2018, pp. 1–6,
[32] T. A. Meynard and H. Foch, “Multi-Level Conversion: High Voltage Choppers and Voltage-Source Inverters,” in Proceedings of the 23rd Annual IEEE Power Electronics Specialists Conference (PESC '92), Toledo, Spain, Jun. 1992, pp. 397–403.
[33] C. Sadanala, S. Pattnaik, and V. P. Singh, “A Flying Capacitor-Based Multilevel Inverter Architecture with Symmetrical and Asymmetrical Configurations,” IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 10, no. 2, pp. 2210–2222, Jun. 2022..
[34] M. Sandhu and T. Thakur, “Modified Cascaded H-Bridge Multilevel Inverter for Hybrid Renewable Energy Applications,” IETE Journal of Research, vol. 68, no. 6, pp. 3971–3983, Nov. 2020.
[35] A. Kersten, O. Theliander, E. A. Grunditz, T. Thiringer, and M. Bongiorno, “Battery Loss and Stress Mitigation in a Cascaded H-Bridge Multilevel Inverter for Vehicle Traction Applications by Filter Capacitors,” IEEE Transactions on Transportation Electrification, vol. 5, no. 3, pp. 659–671, Sep. 2019,

Finite element modeling of a self-excited induction generator

Belynda Fares
Department of Electrical Engineering
University of Bejaia
Bejaia, Algeria
belynda.fares@univ-bejaia.dz

Radia Abdelli
Department of Electrical Engineering
University of Bejaia
Bejaia, Algeria
radia.abdelli@univ-bejaia.dz

Ahcene Bouzida
Department of Electrical Engineering
University of Bouira
Bouira, Algeria
a.bouzida@univ-bouira.dz

Safia Azzougui
Department of Electrotechnics
University of Setif
Setif, Algeria
Safia.azzougui@univ-setif.dz

***Abstract*—Modeling different electromagnetic systems is essential for understanding their operation and optimizing their performance. It also makes it possible to predict their behavior under different operating conditions, which is crucial for designing more efficient and reliable systems. In addition, electromagnetic modeling can also help identify potential problems before they occur, thereby reducing maintenance costs and improving system durability. In the present work, we will model a self-excited induction generator using finite element software, enabling us to simulate its operation under different load and speed conditions. This approach will help us gain an in-depth understanding of the generator's behavior and optimize its performance according to its specific needs. In addition, this modeling will enable us to explore different improvement strategies to maximize the system's energy efficiency.**

Keywords— finite element model, induction generator, numerical modeling, self-excitation.

I. INTRODUCTION

The production of electrical energy is a crucial issue, as demand outstrips supply, which is constantly increasing. Today, over 80% of all energy is produced by fossil fuels. The negative consequences of producing energy from fossil fuels are manifold: environmental pollution, which has adverse effects on both the environment and human beings; on-site storage facilities; transport safety (most of these energies being highly flammable by nature); and to add to this, other factors of an economic nature such as the depletion of available resources and ever-increasing prices [1,3]. To remedy this problem, attention has turned to non-conventional sources of production (solar, wind, hydro, etc.). Among these energies, wind power is the one with the greatest potential for future development, as it is clean and renewable. Wind power is also readily available to meet society's energy demands [4,6]. Wind energy conversion systems can operate either in grid-connected mode, or in stand-alone mode to power small isolated loads [7]. Squirrel-cage induction generators appear to be the ideal choice for stand-alone applications, hence their widespread use in small and medium-sized hydroelectric and wind turbines [8]. This generator offers various advantages over other machines, such as low unit cost, brushless rotor (squirrel cage construction), absence of DC excitation and ease of maintenance [9,10]. Induction generators used in isolated wind energy systems are called self-excited induction generators (SEIGs). The self-excited induction generator has been the subject of considerable research in recent decades, as it is perceived as the simplest energy conversion device for producing electricity off-grid, in stand-alone mode [11]. Autonomous operation of an induction generator requires the presence of reactive energy, which in most cases is provided by capacitors connected to each stator phase winding. Equipped with shunt capacitors, a three-phase SEIG is capable of successfully establishing its terminal voltage as soon as it is driven at an appropriate speed [12].

Modeling any system offers the possibility of simulating its behavior and analyzing its performance. This enables us to better understand how it works and to anticipate the consequences of different scenarios. Modeling with finite element software is particularly useful for engineers and researchers to design more efficient and innovative products. It also reduces the cost of physical testing by virtually simulating system behavior under different conditions. In the present work, we will study and simulate the operation of a self-excited induction generator. We will also analyze the performance of this generator under different load conditions in order to better understand its behavior in real operation. The results obtained will enable us to draw conclusions about the efficiency and reliability of this generator.

II. GENERATOR MODE OPERATION OF AN INDUCTION MACHINE

For an induction machine to operate in generator mode and generate power, it must be driven at a speed higher than the synchronous speed. The synchronous speed is given by the following equation:

$$n_s = \frac{120 \times f}{p} \quad (1)$$

f is the frequency Hz.
p is the number of poles.

As the rotor speed exceeds the synchronous speed, this leads to negative slip, which means that the induction machine will produce a negative conjugate, functioning as a generator [13]. Induction generators are commonly used in stand-alone wind power plants. In this case, the rotor of the induction machine is set in motion by the wind turbine blades, and a voltage is generated by connecting a capacitor bank to the stator terminals. This process is called self-excitation of the induction generator [14].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

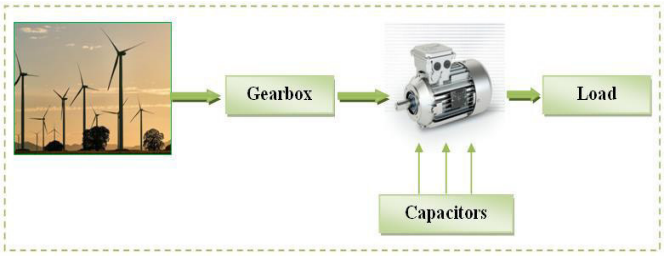


Fig 1. SEIG excited by an isolated-mode capacitor feeding a load.

Self-excitation can occur in an induction machine under certain conditions. When reactive power is supplied by the capacitor bank and the machine has a residual magnetic field, it can generate electricity. The residual magnetism creates an electric current in the stator windings, which is proportional to the rotor speed. This current is then applied to the capacitors connected to the stator terminals, causing a reactive current in the stator windings. This leads to the establishment of a magnetizing flux in the machine, which can generate electricity. Stator voltage is limited by the machine's magnetic saturation. This ability to generate electricity in remote locations without an electrical grid makes induction machines useful in a variety of applications [11].

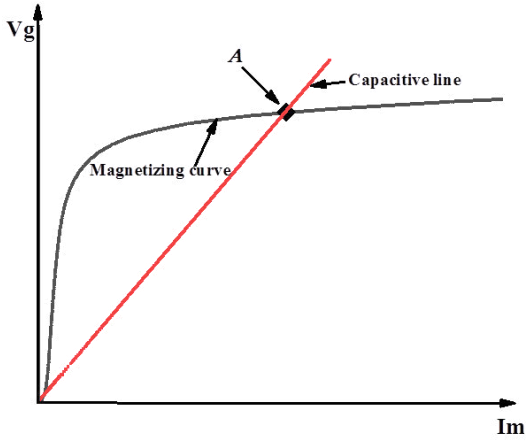


Fig 2. Induction generator magnetization curve.

III. NUMERICAL MODELING OF THE SELF-EXCITED INDUCTION GENERATOR

Finite element modelling is a numerical technique performed on a computer to calculate electromagnetic parameters. This technique allows for the consideration of many complex parameters, and it enables the simulation of the generator's operation under various conditions, which is crucial for comprehending its behavior in real-life situations.

Finite element modelling is essentially based on the solution of Maxwell's equations, using Maxwell's four fundamental partial differential equations, which in their most general form are written:

Maxwell-Gauss $\nabla . D = \rho$

Maxwell-Faraday $\nabla \wedge E = -\frac{\partial B}{\partial t}$

Conservation du flux $\nabla . B = 0$

Maxwell-Ampère $\nabla \wedge H = J_c + \frac{\partial D}{\partial t}$

E and H: electric and magnetic fields
D and B: electric and magnetic induction
Jc and ρ: density of conduction currents and electric charge.

➢ **Presentation of the self-excited induction generator**

The machine studied is an asynchronous squirrel-cage machine, whose characteristics are shown in the table below:

Table I. Characteristics of the machine studied.

Asynchronous generator	
Rated power	400 Kw
Rated voltage	400 V
Rated speed	1500 tr/min
Operating temperature	75°C
Frequency	50 Hz
Coupling type	delta

Table II. Stator and rotor characteristics of the generator studied.

Stator parameters		**Rotor parameters**	
Outside diameter of stator core	580 mm	Rotor core outer diameter	397.2 mm
Stator core inner diameter	400 mm	Rotor core inner diameter	140 mm
Stator core length	690 mm	Stator core length	690 mm
Number of stator core slots	72	Number of rotor bars	56

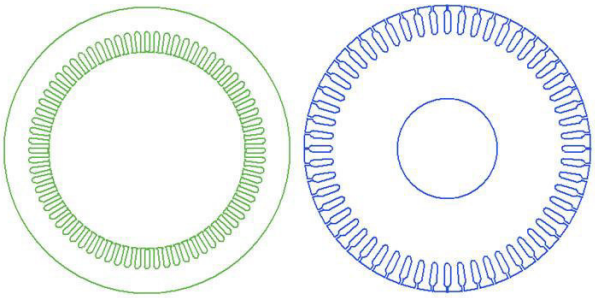


Fig 3. Stator and rotor of the generator studied.

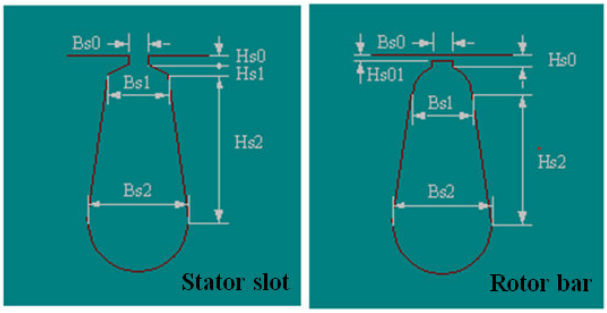


Fig 4. Rotor bar and slot structure

Table III. Slot and rotor bar parameters

	Stator slot	Rotor bar	
Parameters	Value		unite
Hs0	1.2	6.8	mm
Hs1	0.05	0	mm
Hs2	34	25	mm
Bs0	4.6	3.5	mm
Bs1	9.03	10	mm
Bs2	12	8	mm

After designing the generator model, we install a capacitor bank in the stator windings to provide the reactive power needed to operate the generator. The value of the self-excitation capacitance is set at 500µF for all the tests carried out.

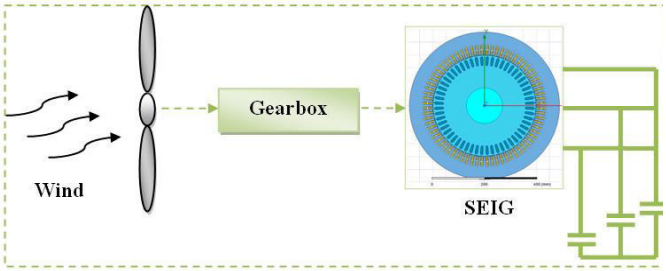


Fig 5. Overview of the study system

IV. Performance of the self-excited induction generator

The SEIG model is simulated at a constant speed of 1570 rpm. To initiate the voltage rise, a small stator current of 1A is used, representing the initial magnetic flux required in the machine core. The SEIG voltage cannot be generated without this current. The model developed is tested and verified for various operating conditions.

- no-load

For a no-load test, the figure below shows the evolution of the electrical quantities: voltage and balanced three-phase current.

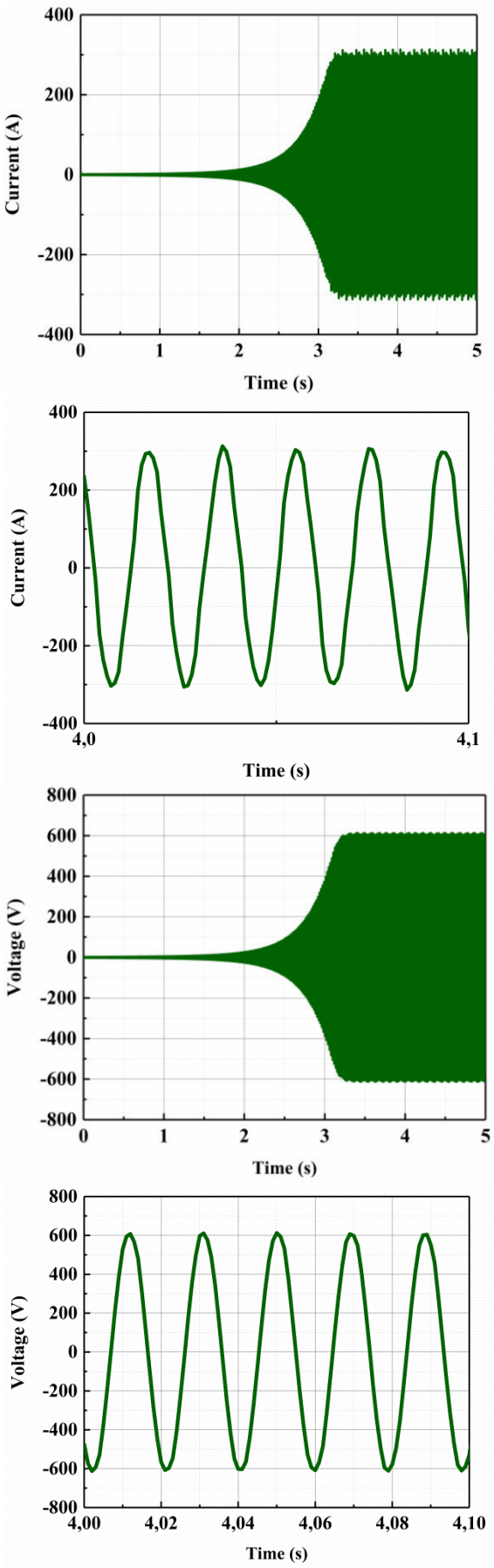


Fig 6. Evolution of electrical quantities without load

The results obtained show that the electrical quantities current, and voltage pass through a self-excitation time

lasting around 2s. The induced voltage rises in exponential form, stabilizing at a fixed value corresponding to the steady-state operating point set by the saturation of the magnetic circuit. The stator current is identical in form to the voltage. Once the quantities have stabilized, they remain constant as long as operating conditions do not change significantly.

- Resistive load

Keeping the same characteristics used for no-load operation. At time t= 4s, we will connect a 3 ohm balanced three-phase load.

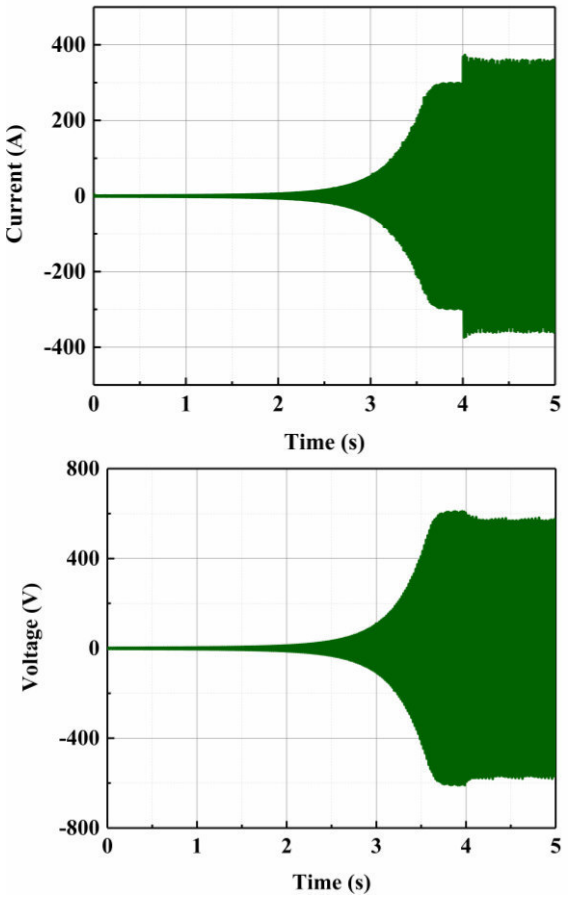


Fig 7. Evolution of electrical quantities with resistive load

At t=4s, when the load R is introduced, there is a noticeable decrease in the phase voltage value and a simultaneous increase in the phase current value.

- RL load

In this second case, we introduce a load RL balanced at time t=4s, and keep the same characteristics as before, with a resistance of value R= 3 ohm and an inductance of value L= 10^{-4}H.

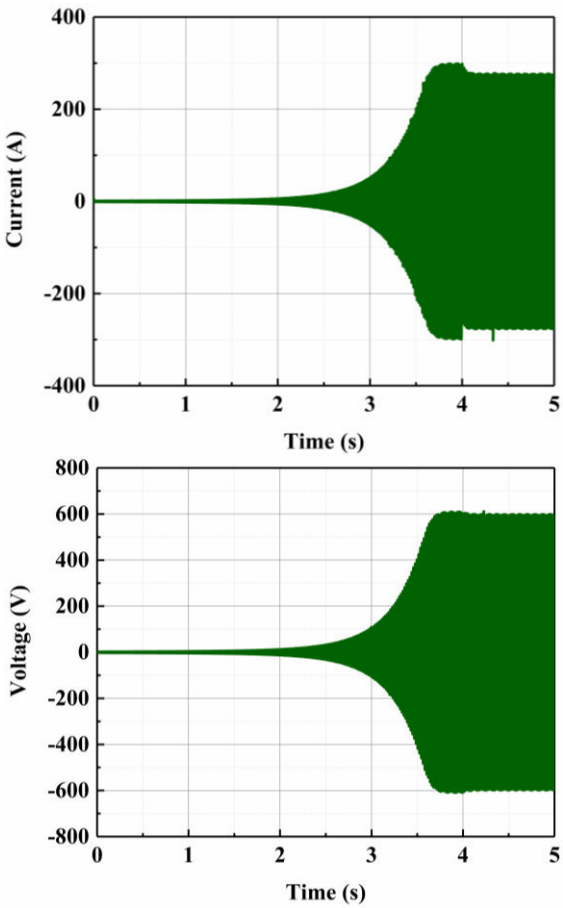


Fig 8. Evolution of electrical quantities with RL load

The results obtained for the RL load are similar to those obtained for an R load. As with the resistive load, connecting an RL load induces a drop in voltage and an increase in phase current, but the latter is more pronounced in the case of an RL load, due to the demagnetizing nature of the load current flowing through the inductance. The introduction of an inductive load results in the consumption of reactive energy.

- variable speed

For the last case, we set the capacitance value to C = 500µF, and the load value to R= 3 ohms, then created a variable speed profile as shown in the figure below.

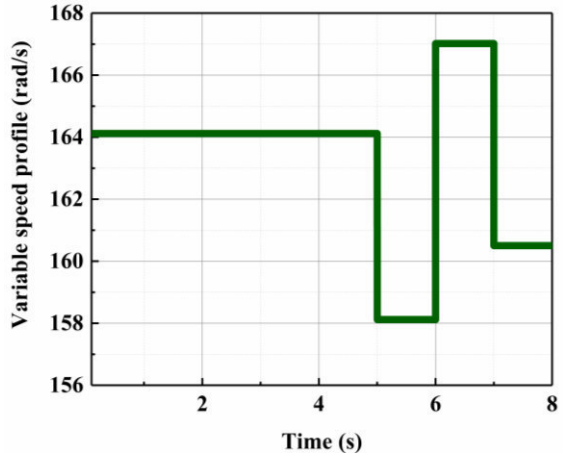


Fig 9. Variable speed profile

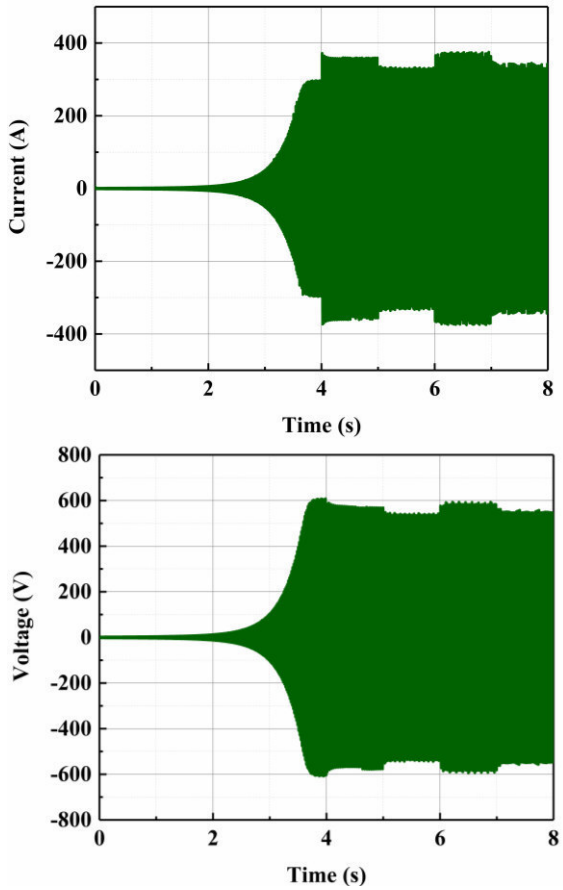


Fig 10. Evolution of electrical quantities with variable speed profile

The speed variation of the rotor affects the electrical properties of the induction generator. If the speed increases, the current and voltage also increase, while they decrease if the speed decreases.

V.Conclusion

A self-excited induction generator is an excellent option for generating electricity in remote areas where grid connection is not possible. It can operate autonomously without relying on an external power source, making it perfect for isolated applications. Additionally, its robust construction enables it to withstand harsh environmental conditions. Modelling this generator using the finite element method offers numerous advantages, including the ability to optimize its design for maximum efficiency and predict its behavior under differing operating conditions. This approach also reduces development costs and improves the generator's durability. Based on the results obtained from various performance tests, we conclude that the self-excited induction generator can operate under different load conditions and at varying rotor drive speeds. As a result, it is particularly suitable for versatile and reliable use in a variety of industrial applications. Furthermore, its finite-element-optimized design ensures high energy efficiency and a long service life, making it a cost-effective and sustainable choice for users.

REFERENCE

[1] K.B. Porate, N. S. Urkude and R. A. Gathe. PERFORMANCE ANALYSIS OF SQUIRREL CAGE INDUCTION GENERATOR USING STATCOM. IJLTEMAS, Volume II, Issue XI, November 2013.

[2] KHADTARE, Sarang A., SAI, Boni Satya Varun, et CHATTERJEE, Debashis. A Model Based Control Strategy for Variable Speed Operation of Three Phase Induction Generator. In : *2020 International Conference on Computational Intelligence for Smart Power System and Sustainable Energy (CISPSSE)*. IEEE, 2020. p. 1-5.

[3] WANNAKAM, Khanittha et JIRIWIBHAKORN, Somchat. Evaluation of Wind Energy Production Using Weibull Distribution and Artificial Neural Networks. In : *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*. IEEE, 2018. p. 1-4.

[4] DHAREPPAGOL,Vishal D. et NAGENDRAPRASAD, S. Modelling and simulation of wecs for maximum power extration and optimal efficincy control using squirrel cage induction generator. In : *2015 IEEE Power, Communication and Information Technology Conference (PCITC)*. IEEE, 2015. p. 912-917.

[5] KHARCHOUF, Ibtissam, ESSADKI, Ahmed, ARBAOUI, Mohammed, *et al.* Modeling and PI Control Strategy of DFIG Based Wind Energy Conversion Systems. In : *2017 International Renewable and Sustainable Energy Conference (IRSEC)*. IEEE, 2017. p. 1-6.

[6] LIU, Yu, YANG, Jian, JIANG, Changming, *et al.* Review on met mast site selection methods in grid-connected wind farm. In : *2019 IEEE 3rd International Electrical and Energy Conference (CIEEC)*. IEEE, 2019. p. 1134-1137.

[7] GUJALWAR, Namdeo S., KASTHA, Debaprasad, et KISHORE, N. K. Comparison of Self Excited Squirrel Cage Induction Machine based Wind Energy Conversion System. In : *2014 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*. IEEE, 2014. p. 1-6.

[8] GÓRSKI, Dominik A. et IWAŃSKI, Grzegorz. Asynchronous grid connection of a cage induction generator excited by a power electronic converter. *IEEE Transactions on Energy Conversion*, 2020, vol. 36, no 1, p. 63-70.

[9] KINNARES,V.et SAWETSAKULANOND, B. Characteristic requirements of a small scale squirrel cage induction generator for effective electricity generation from wind energy. *Energy Procedia*, 2013, vol. 34, p. 26-49.

[10] KARAKASIS, N., MESEMANOLIS, A., et MADEMLIS, C. Performance study of start-up control techniques in a Wind Energy Conversion System with induction generator. In : *International Symposium on Power Electronics Power Electronics, Electrical Drives, Automation and Motion*. IEEE, 2012. p. 547-552.

[11] KUMAWAT, R. K., CHOURASIYA, Seemant, AGRAWAL, Seema, *et al.* Self excited induction generator: A review. *International Advanced Research Journal in Science, Engineering and Technology (IARJSET)*, 2015, vol. 2, no 1, p. 37-42.

[12] KHAN, Faisal, KHAN, Rizwan, et KOUZOU, A. Three phase self-excited induction generator for renewable energy generation. In : *ICPEA 2013, 1st. International Conference on Power Electronics and their Applications*. 2013.

[13] VANCO, Wagner E., SILVA, Fernando B., GONCALVES, Felipe AS, *et al.* Evaluation of the capacitor bank design for self-excitation in induction generators. *IEEE Latin America Transactions*, 2018, vol. 16, no 2, p. 482-488.

[14] SINGH, Sunil, KUMAR, Rohit, et GOEL, Sandeep Kumar. Steady-State Analysis of Three-Phase Self-Excited Induction Generator for Stand-Alone Applications. IJRASET, 2018, Vol 6, Issue I, p. 3341-3347.

Impacts of DC/DC converter faults in the performance of photovoltaic systems

FARES Belynda
Department of Electrical Engineering
university of Bejaia
Bejaia, Algeria
belynda.fares@univ-bejaia.dz

AIBECHE Abderrezak
Department of Electrical Systems Engineering
university of Boumerdes
Boumerdes, Algeria
aibeche_umbb@univ-boumerdes.dz,

AZZOUGUI Safia
Department of Electrotechnics
University of Setif
Setif, Algeria
Safia.azzougui@univ-setif.dz

Abstract_ The goal of this work is to investigate the effects of open-circuit and short-circuit failures in DC/DC converters on all solar system components linked to electrical networks. The model under consideration is realized using the Matlab/Simulink programmers, and the findings produced are confirmed by it.

Keywords_ photovoltaic system, faults, impacts, short circuit, open circuit, DC/DC converter.

I. Introduction

Because of the rapid industrial development, the problem of energy scarcity is becoming increasingly acute. Because the world's energy consumption is primarily provided by non-renewable energies, and because these are finite, energy strategies aimed at increasing the efficiency of electrical systems are being implemented. with the goal of increasing the efficiency of electrical systems, As a result, several nations have shifted to renewable energy (hydraulic, wind, sun, biomass, geothermal, and tidal) [1,2].

Among these green energies, solar photovoltaic energy has garnered the greatest attention due to a variety of stimulating reasons, including lower production costs and support regulations. These motivators make the return on investment of a solar installation more appealing [3]. Unfortunately, like with any other industrial process, a photovoltaic system might be confronted with various faults and anomalies during operation, resulting in a loss in system performance and, in extreme cases, total system unavailability [3].

In this paper, we will look at the effects of open and short circuit faults in dc/dc converters on the various components of a solar system.

II. Modeling the photovoltaic system connected to the electrical networks

The photovoltaic energy production system consists of a photovoltaic panel that converts luminous energy into continuous current, which then passes through a boost hacheur to increase its value before being converted into alternate current and connected to the network. Figure 1 depicts a typical photovoltaic system structure [4].

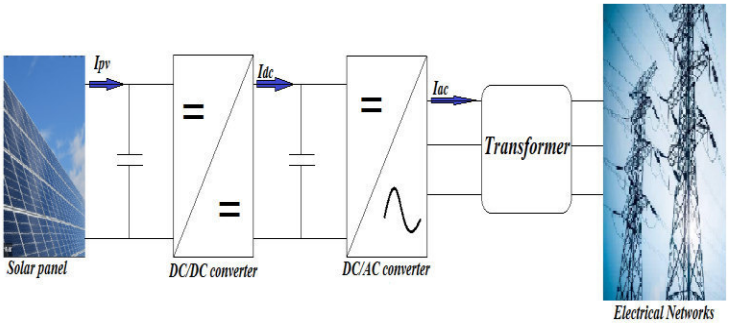


Fig1: Grid-connected photovoltaic system.

- **Model of PV cell**

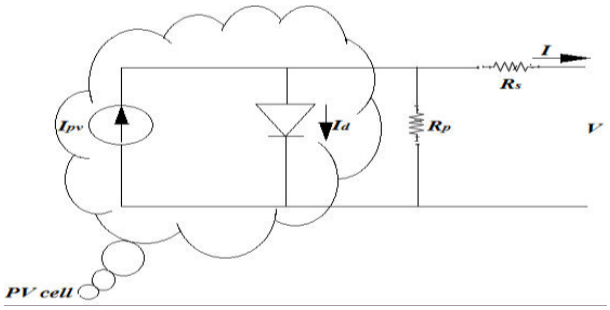


Fig2: Equivalent circuit of a PV solar cell.

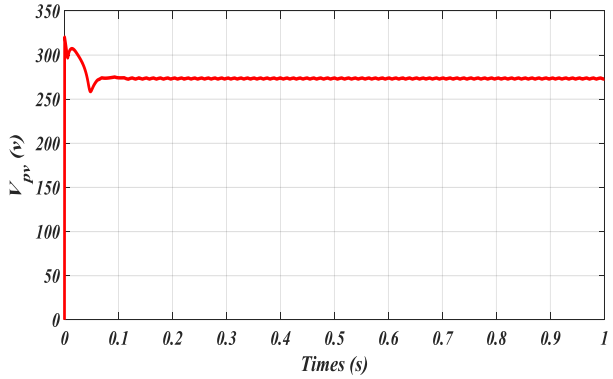
By multiplying the IPV current by the voltage V, we obtain the electrical power produced by a PV cell. The IPV current is given by

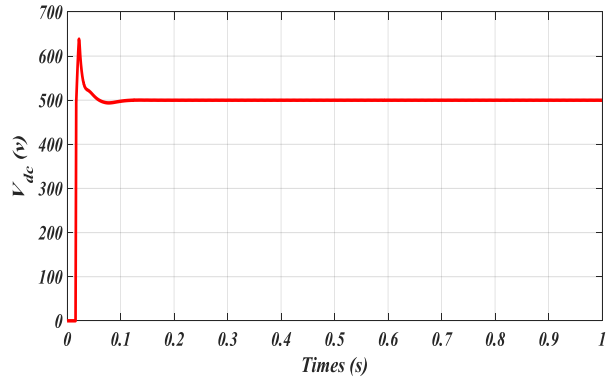
$$I = I_{pv} - I_0 \left[\exp\left(\frac{V+Rs\,I}{Vt\,a}\right) - 1\right] - \frac{V-Rs\,I}{Rp} \qquad (1)$$

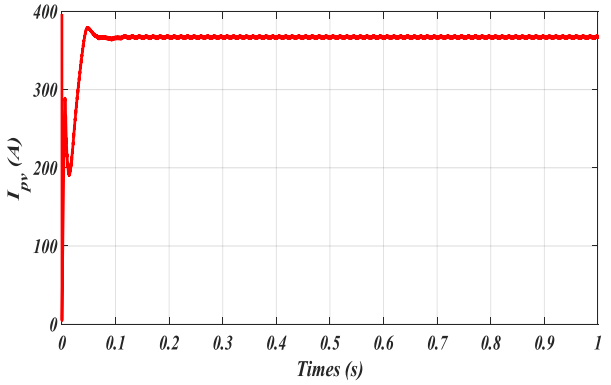
With:
Ipv = the photovoltaic current of the installation
Io = photovoltaic current of the system
Vt = thermal voltage of the generator
Rs = equivalent series resistance
Rp = equivalent parallel resistance
a = generally $1 \leq a \leq 1.5$ and the choice depends on the other parameters of the I-V model [5].

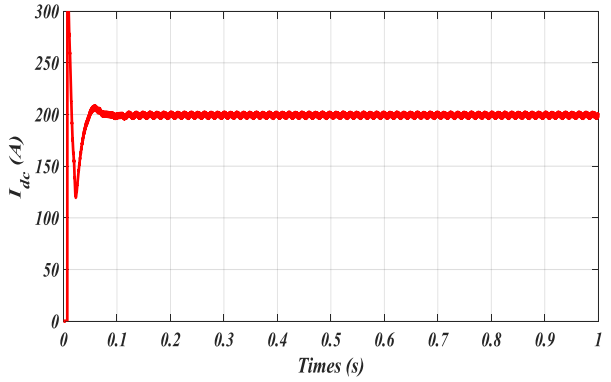
XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

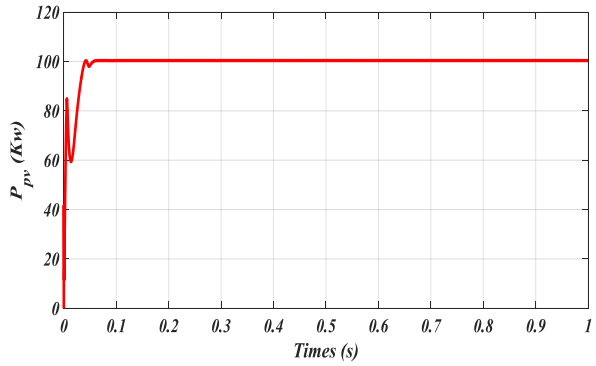
- **Performance of the healthy PV system**











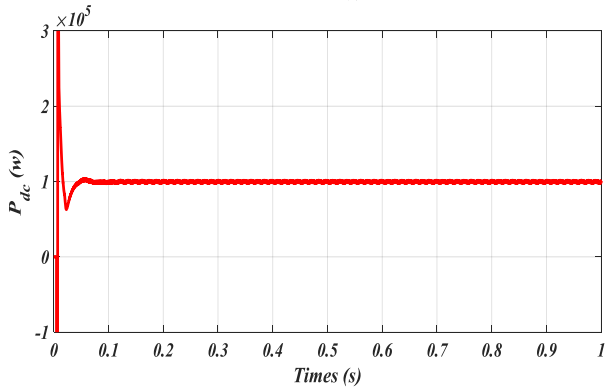


Fig3: Electrical values on the PV side and on the dc/dc converter side.

III Impact of O-C and S-C faults on the PV system

The photovoltaic systems are confronted during their operation with various faults and that at various levels of the production chain, the faults which are repeated most often are the faults circuit-open and short-circuit at the level of the power switches of the dc/dc converters.

- **Open circuit**

At the time t=0.5s we cause the opening of the switch (IGBT) by blocking the control sequences in the 0 state.

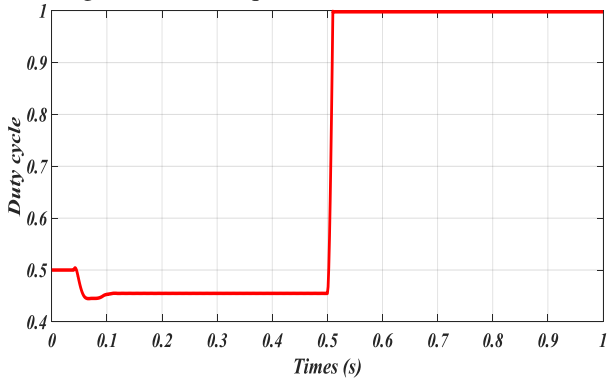
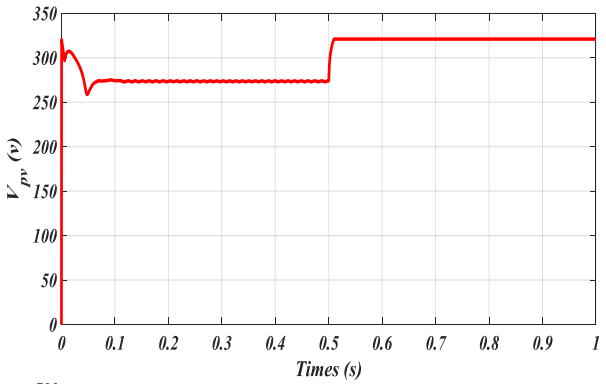
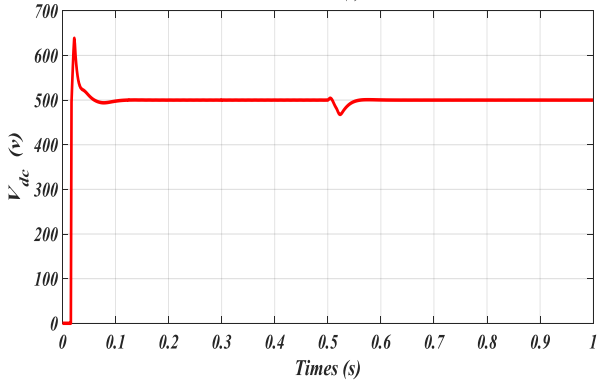
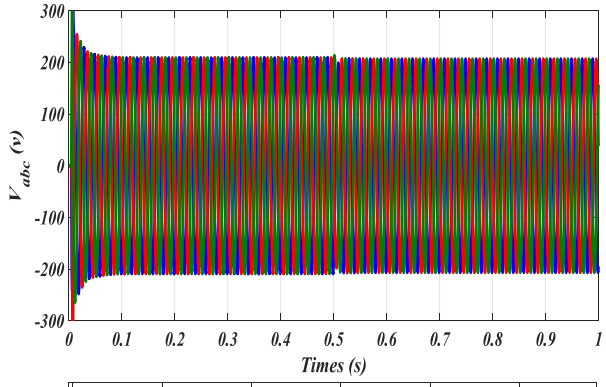
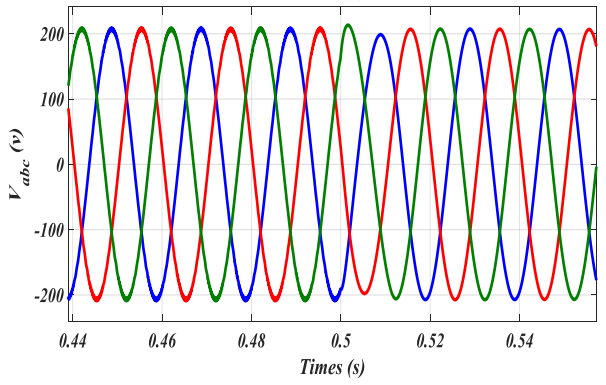


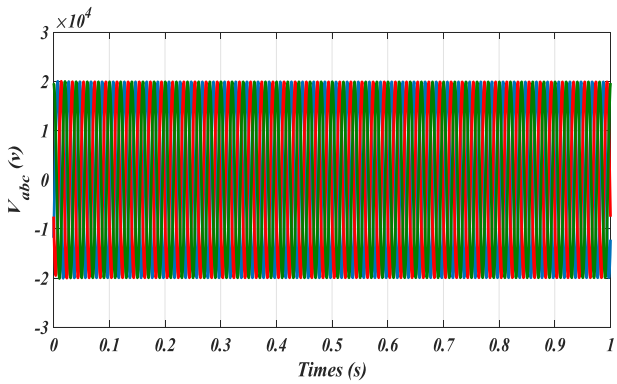
Fig4: Duty cycle during the O-C fault.











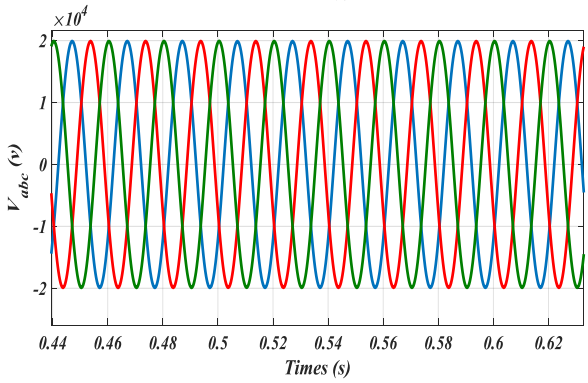
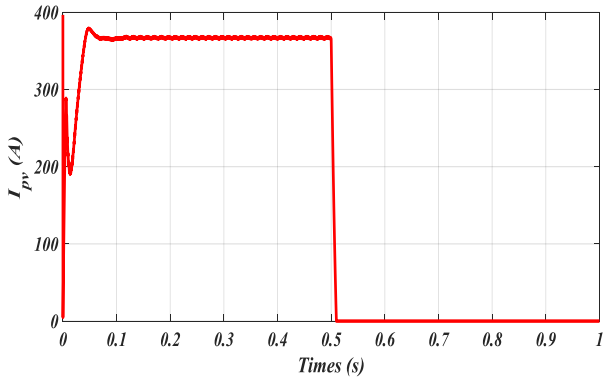
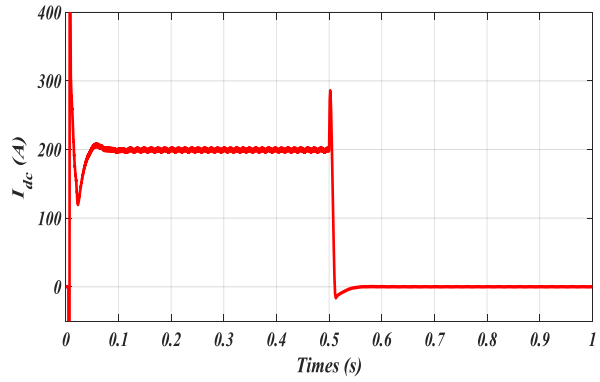


Fig5: Voltage V_{pv} V_{dc} V_{abc} $V_{abc_network}$ during the O-C fault.





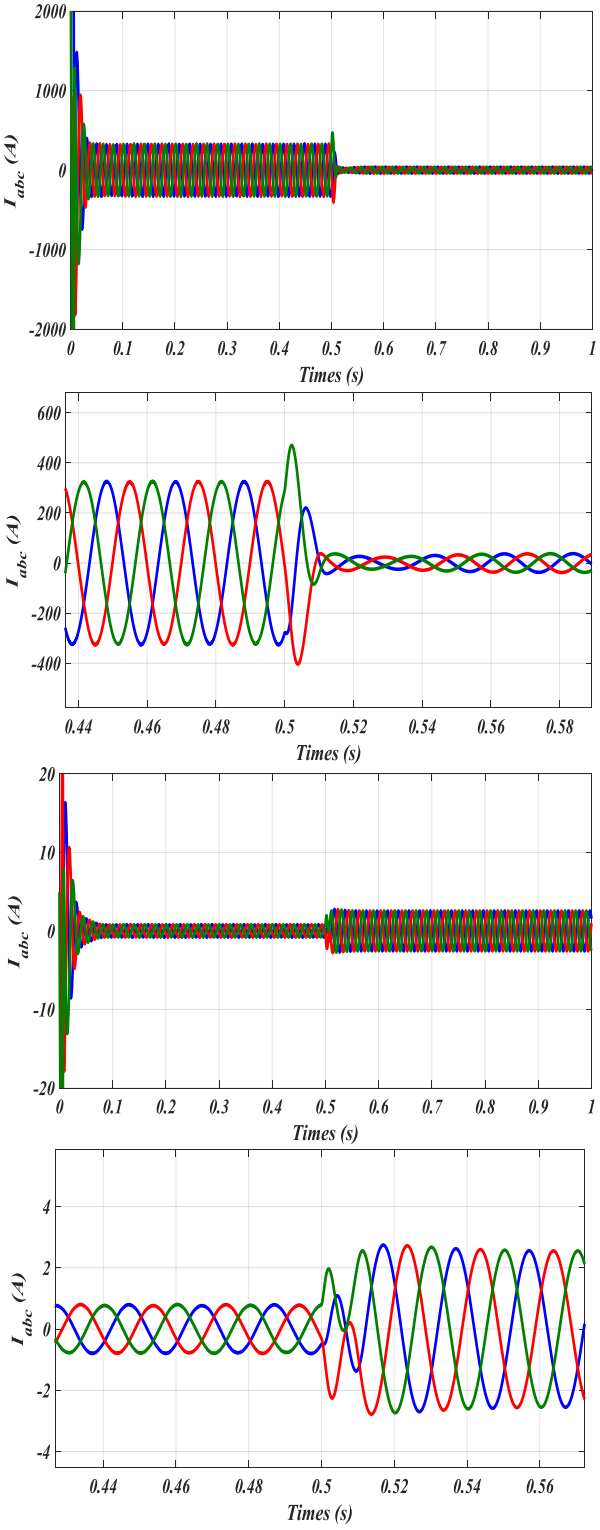


Fig6: Current I_{pv} I_{dc} I_{abc} $Iabc_{_network}$ during O-C fault.

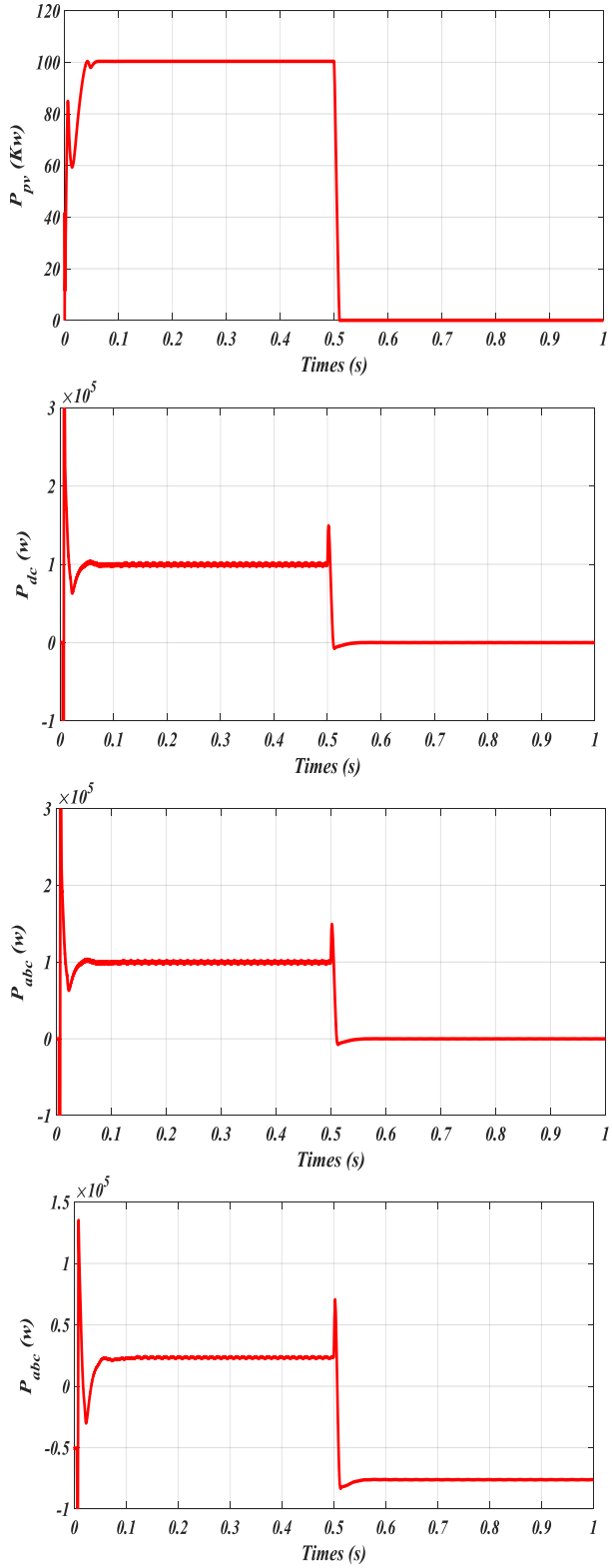


Fig7: Power P_{pv} P_{dc} P_{abc} $P_{abc_network}$ during the O-C fault.

- **Short circuit**

At time t=0.5s we cause a short circuit fault on the switch (IGBT) by blocking the control sequences in state 1.

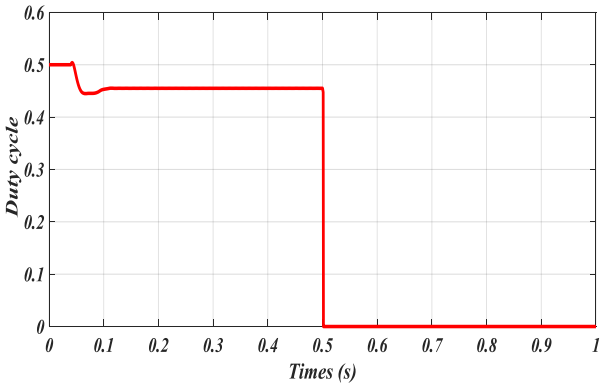
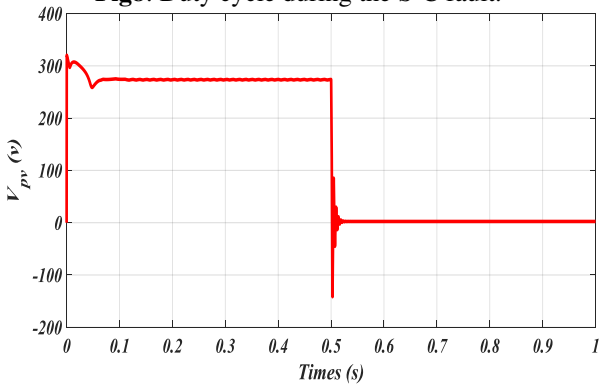
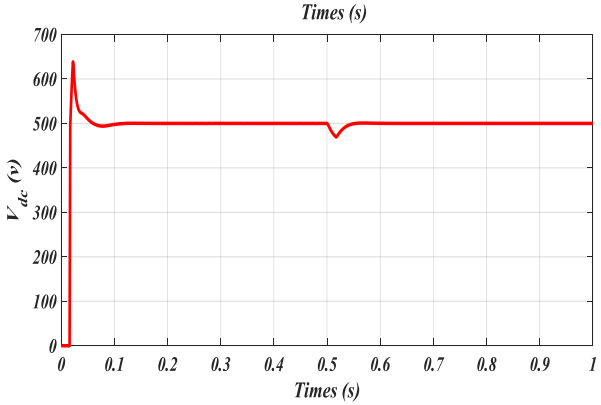
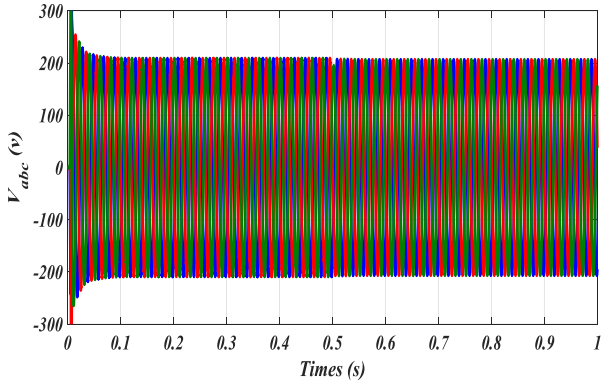
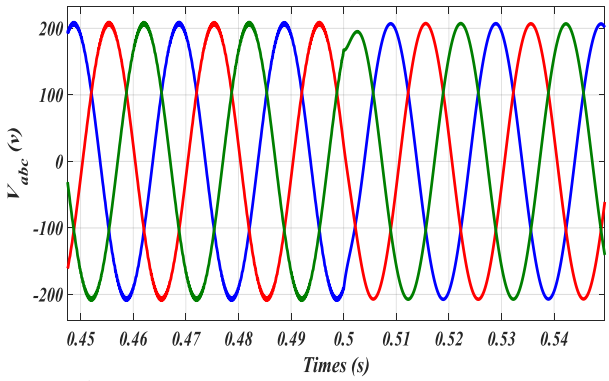


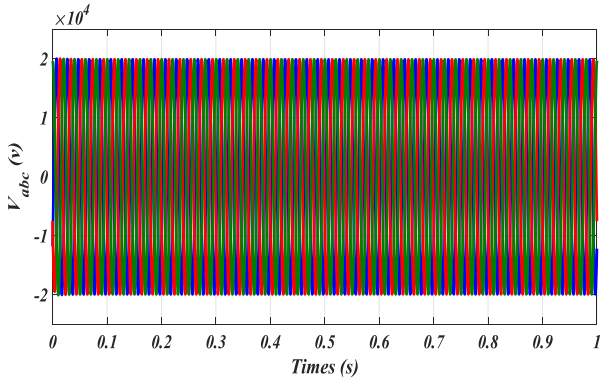
Fig8: Duty cycle during the S-C fault.











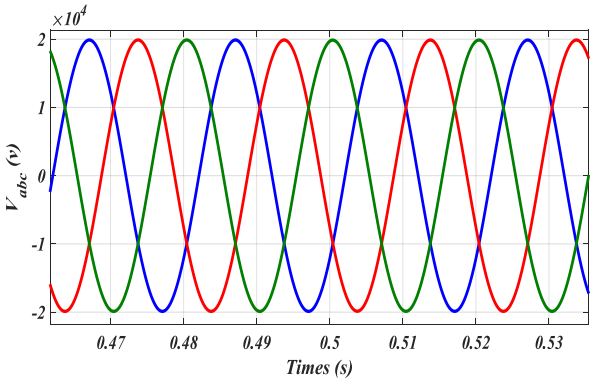
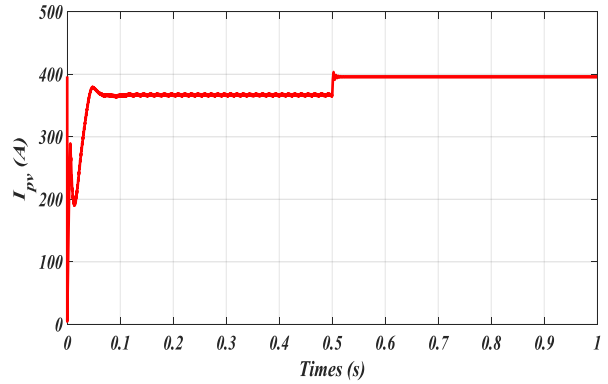
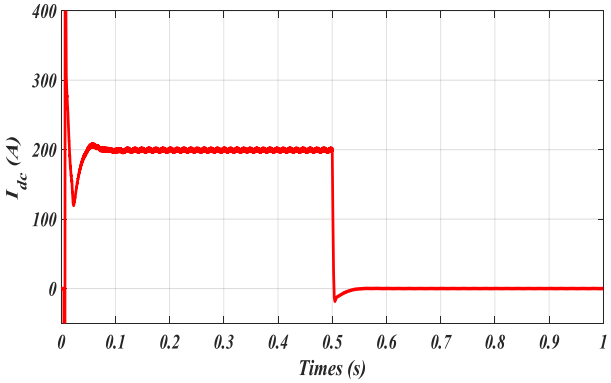
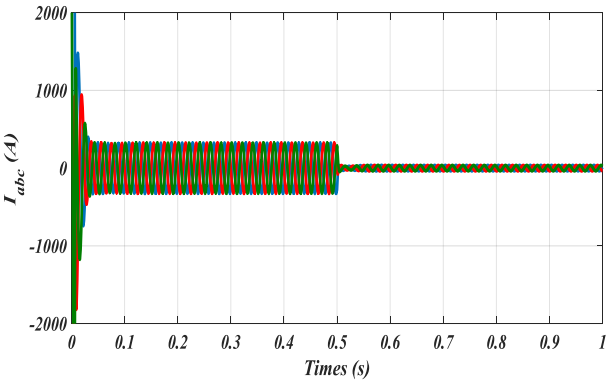
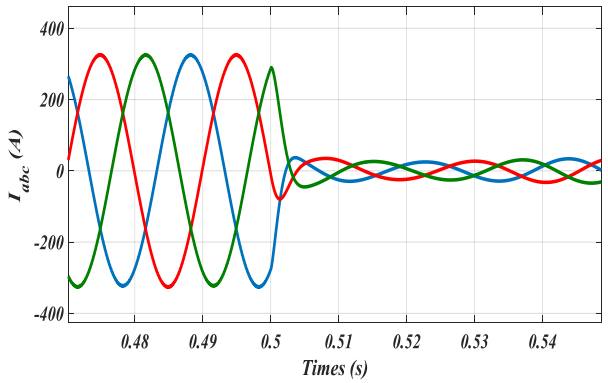


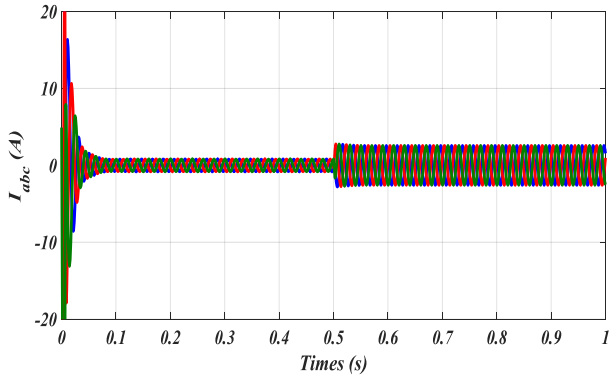
Fig9: Voltage V_{pv} V_{dc} V_{abc} $V_{abc_network}$ during the S-C fault.











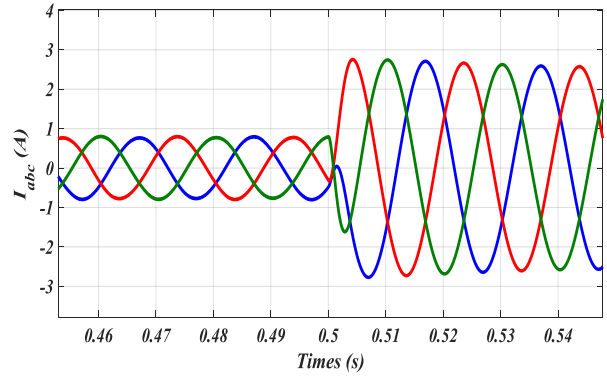
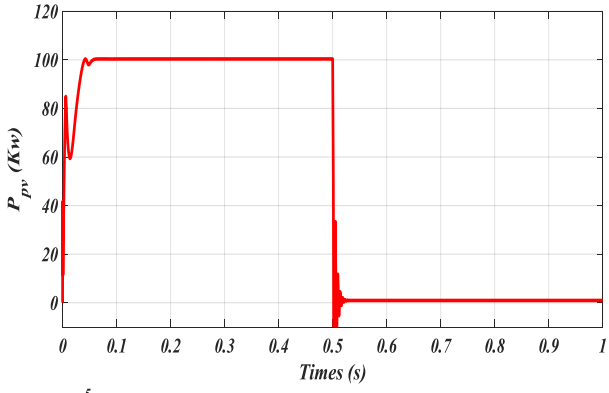
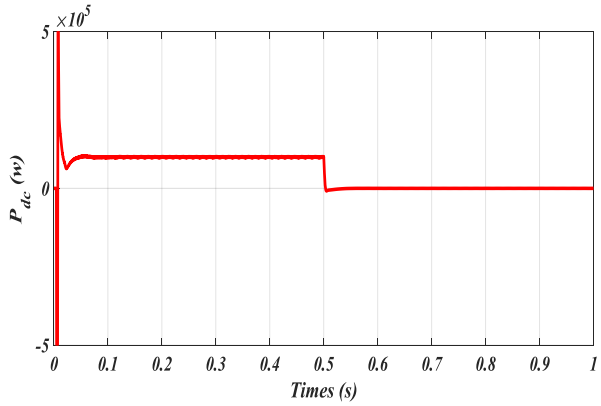


Fig10: Current I_{pv} I_{dc} I_{abc} $Iabc_{_network}$ during S-C fault.





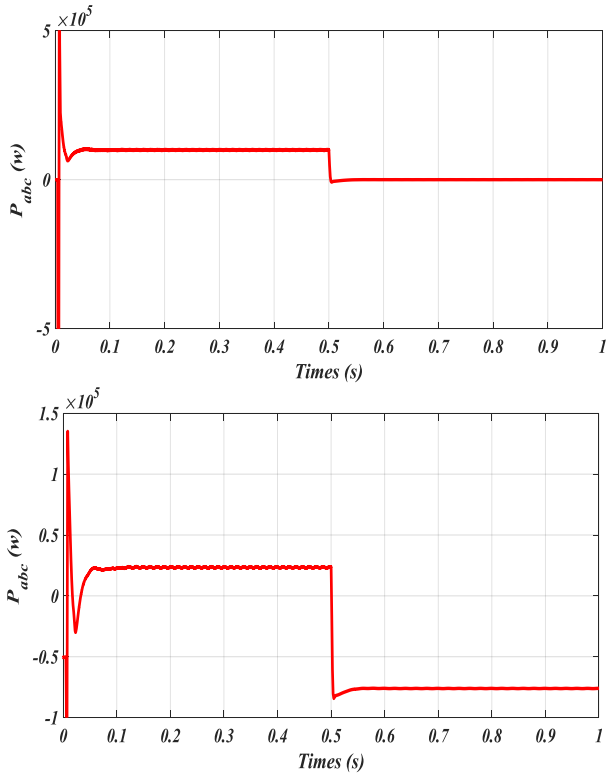


Fig11: Power P_{pv} P_{dc} P_{abc} $P_{abc_network}$ during the S-C fault.

IV. Discussion of the results

We can see from the results that an open circuit fault causes an increase in the voltage Vpv, which corresponds to the voltage in fault, and Vdc also increases at a voltage imposed by the inverter's input, the currents Ipv and Idc are cancelled, giving a power Ppv and Pdc of zero, and the quantities are disturbed in the same way at the inverter's output.

The short circuit fault increases the current Ipv to an excessive (dangerous) value and cancels the voltage Vpv; the dc/dc converter likewise experiences a disturbance at the level of the voltage Vdc only at the time of the fault; the current Idc, as well as the powers Ppv and Pdc, are cancelled.

V. Conclusion and perspective

In this work we have studied the impacts of the open-circuit and short-circuit faults of the IGBT switches of the dc/dc converters on all the components of the photovoltaic conversion chain, we conclude that the open-circuit fault does not have a destructive effect on the system however it can create a decrease in the performances, while the short-circuit fault has a dangerous effect on the installation because it can lead to its total dysfunction. This is why it is essential to implement a diagnostic and reconfiguration system in order to guarantee the protection of all the elements and their continuity of service.

REFERENCES

[1] M.F. Yaden, M. El Ouariachi, T. Mrabti , Ka. Kassmi, B. Tidhaf, E. Chadli et K. Kassm, ' Conception et réalisation d'un système photovoltaïque muni d'une commande MPPT numérique', Revue des Energies Renouvelables Vol. 14 N°1, 2011.

[2] N. Aouchiche , M.S. Aït Cheikh et A. Malek , ' Poursuite du point de puissance maximale d'un système photovoltaïque par les méthodes de l'incrémentation de conductance et la perturbation & observation', Revue des Energies Renouvelables Vol. 16 N°3, 2013.

[3] A.E. Toubal Maamar , S. Ladjouzi , R. Taleb and Y. Kacemi, ' Détection et classification de défauts pour un GPV: Etude comparative entre la méthode de seuillage et réseaux de neurones', Revue des Energies Renouvelables Vol. 21 N°1, 2018.

[4] Ren Jiayu , Li Chunlai , Teng Yun , Yang Xia ,Yao Shengpeng ,' Modeling of photovoltaic grid connected generation system based on parameter identification method', IEEE International Conference on Smart City and Systems Engineering, 2016.

[5] M. Z. C. Wanik , A.A.Ibrahim , A.K.M. Hussin , M.R. Rusli , J.H. Tang, ' Simplified Dynamic Model of Photovoltaic Generation System for Grid Integration Studies', IEEE , 2014.

Assessment of Several Mataheuristic Algorithms for Economic Analysis of a Hybrid Renewable Energy Systems Connected to the Electricity Grid

Oussama Belhadef
Automatic Laboratory of Setif
Department of electrical engineering
Setif 1 University – Ferhat ABBAS
Setif, Algeria
belhadef.oussama@univ-setif.dz

Farid Merahi
Automatic Laboratory of Setif
Department of electrical engineering
Setif 1 University – Ferhat ABBAS
Setif, Algeria
farid.merahi@univ-setif.dz

Abdessalam Badoud
Automatic Laboratory of Setif
Department of electrical engineering
Setif 1 University – Ferhat ABBAS
Setif, Algeria
badoudabde@univ-setif.dz

***Abstract*—The harmful effects of fossil fuels, namely coal, oil and gas, on living creatures and the environment have led the world to turn to greener alternatives, by increasing the integration of renewable energy sources (RES) into power systems. Despite the allure of a greener future, the large-scale integration of these RES still faces many obstacles, mainly due to the irregular nature of most of these sources, in addition to their geographical limitations. A key solution to this problem is the adoption of the concept of hybrid renewable energy systems (HRES), where more than one green source is employed in the same system, in addition to the inclusion of energy storage devices (ESD). However, these HRES increase the complexity of the system, making the energy management strategy for these systems extremely important in order to achieve the best performance. This research was carried out on simulations of photovoltaic/wind/battery systems connected to the main electricity grid, studying their optimization using several intelligent metaheuristic approaches. Numerous algorithms, including particle swarm optimization (PSO) and the whale optimization algorithm, were applied by the energy management system (EMS) to optimize the energy flow based on economic analysis. The EMS returns the optimal operating cost of the system, given the limited renewable energy generation, load demand and energy prices. The comparative analysis demonstrates the superiority of the PSO over other methods, offering the lowest cost, greater resilience, and higher reliability. The results of the study highlight the importance of choosing the right algorithm for energy management in HRES, as it has a significant impact on the overall profitability and sustainability of the system.**

Keywords—hybrid renewable energy system, energy management system, energy storage device, particle swarm optimization, system operating cost

I. INTRODUCTION

The rapid growth in global energy consumption, fueled by demographic expansion and significant technological advances, highlights the urgent need to improve energy production capacity. However, in meeting this demand, the energy sector is becoming a major contributor to environmental challenges, particularly the increase in greenhouse gas emissions, principally CO_2, which is accelerating climate change and causing global warming [1]. This growing demand for energy, combined with the harmful effects of fossil fuels, is forcing us to switch to new, green, and renewable energy sources. Renewable energies exploit natural processes such as solar and wind power, which are constantly renewed and are widely available, making them sustainable sources.

Unlike fossil fuels, RES such as solar, wind, hydro, geothermal, biomass and others do not contribute to environmental degradation and climate change, as they produce negligible amounts of greenhouse gas emissions [2]. The transition to RES is extremely important for reducing dependence on fossil fuels and non-renewable sources, ensuring energy security, and mitigating the health and environmental problems caused by traditional sources of energy production. As technology advances, RES become more efficient and cost-effective, offering a promising solution for a sustainable energy future [3].

Although RES offer considerable advantages, dependence on a single source does not always guarantee the sustainability and reliability of the energy, given the intermittent nature of these resources. If we take solar energy as an example, solar irradiation is absent at night, resulting in zero energy production by photovoltaic panels, in addition to cloudy days when energy production decreases and fluctuates according to the weather.

A key strategy for mitigating the intermittency and variability inherent in RES is the deployment of hybrid systems. By integrating two or more different types of RES, a hybrid system can take advantage of the complementary nature of the integrated sources to create a more reliable and continuous energy supply [4]. For example, solar power tends to be abundant on clear, sunny days, but its availability decreases at night or on cloudy days, while wind power often peaks during colder seasons or when the sky is overcast, thus balancing out the solar resource. Combining these two forms of energy results in more stable and consistent energy production [1].

In addition, energy storage devices (ESDs), including batteries, supercapacitors, fuel cells and pumped water reservoirs, are integrated into HRES to store excess energy produced during periods of low demand or when renewable energy production exceeds consumption requirements [1],[5]. These storage systems play a crucial role in improving the overall performance and flexibility of the energy system, as they allow energy produced during periods of peak renewable energy production to be stored for later use when production is low [5]. ESDs can be used in a wide range of power generation systems, from large-scale industrial to small-scale residential applications, as discussed in [6]. However, the true potential of HRES can only be realized by effectively managing the complex combination of its components to optimize efficiency, reliability, sustainability, and power quality [7]. This complexity highlights the importance of advanced energy management systems (EMS), which play a central role in coordinating the generation, storage and distribution of energy to ensure that the hybrid system performs at its best while meeting the dynamic demands of the grid.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Given the diversity and intermittency of energy sources in HRES, as well as the integration of various ESDs and other system components, EMSs are essential to ensure efficient operation. EMSs are responsible for coordinating and controlling the flow of energy between all units in the system, optimizing performance, minimizing costs, and ensuring that energy is used efficiently [1], [6], [8], [9]. One of the main functions of an EMS is to manage the dynamic balance between energy production, storage, and consumption, which is essential to maintain grid stability and prevent inefficiencies or energy shortages [10]. Researchers in [11] argue that accurate forecasting in EMS is increasingly important to optimize the balance between energy supply and demand, reduce operating costs and ensure the reliability and efficiency of hybrid renewable energy systems. By intelligently controlling this energy distribution, EMSs help to maximize the reliability and efficiency of HRESs while maintaining system security. In addition, EMSs make a significant contribution to the overall stability of the grid by ensuring that energy is delivered consistently to end users without disrupting the balance renewable generation and demand, thus enabling a smoother integration of renewable energy into the broader energy infrastructure [1],[10].

The rest of this paper is organized as follows:

- The second part is designated to the system description of the studied HRES.
- The third part discusses the EMS applied for our HRES, mentioning the utilized optimization algorithms, and presenting the simulation results.
- The last part is a conclusion, summarizing the key points about this study.

II. Hybrid Renewable Energy System Description

Despite significant technological advances in different RES, as well as significant progress in cost reduction and policy incentives for the integration of RES into power systems, relying on a single RES does not fully achieve the energy sector's main objectives of sustainable and efficient power generation. This is largely due to the challenges inherent in each RES [1]. A HRES is an electricity generation framework that integrates two or more RESs to achieve more efficient, sustainable and consistent electricity generation. HRES can integrate energy storage systems (ESS) to store excess energy and manage periods of high demand. They offer a significant solution to the challenges posed by dependence on a single energy source, particularly when they combine complementary sources such as wind and solar power. HRES help to stabilize electricity grids, improve energy availability and provide electricity to rural and remote areas that do not have access to the main electricity grid [3].
In this study, a hybrid system consisting of PV and wind sources, due to the efficiency and complementary nature of these sources, together with an ESS battery was selected, as shown in Fig. 1. The system is used in a grid-connected configuration, as it has a connection to the main electricity grid to import energy when needed. The power profiles of the system under study are shown in Fig. 2, which presents the production and demand. The power output of the photovoltaic system varies from 25kW to 80kW, while the

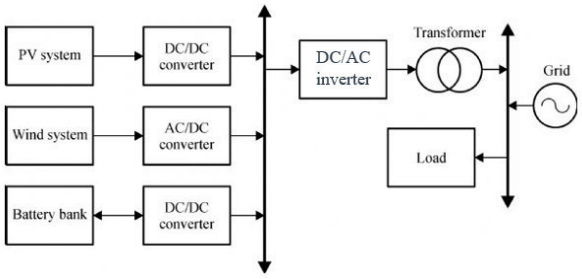


Figure.1. PV/Wind/Battery HRES configuration.

wind power fluctuates between 20kW and 50kW. On the demand side, the load power varies from 60kW to 115kW at peak times.

III. Energy Management System

Combining several RESs in a hybrid system holds great promise for achieving reliable and sustainable energy production. However, this approach presents a few challenges, energy management being one of the most pressing. The inclusion of additional RESs and energy storage units increases the complexity of the system, requiring advanced mechanisms to effectively manage energy and power flows. An effective Energy Management System (EMS) is essential to meet these challenges, as it governs the flow of energy in real time within the HRES components and between the HRES and the main grid [12].

The main objectives of an EMS are to meet demand, reduce operating costs, improve component life and maximize overall efficiency. With the energy sector placing greater emphasis on sustainability and efficiency, the role of energy management has grown rapidly, leading to the development of various EMS models and optimization techniques [13].

To assess the performance of a specific HRES system, it is essential to have clearly defined measures and indicators. These measures are used to evaluate the effectiveness of the EMS solutions and the algorithms used for control and management. For HRES, assessment typically involves three key categories of indicators: economic, environmental and

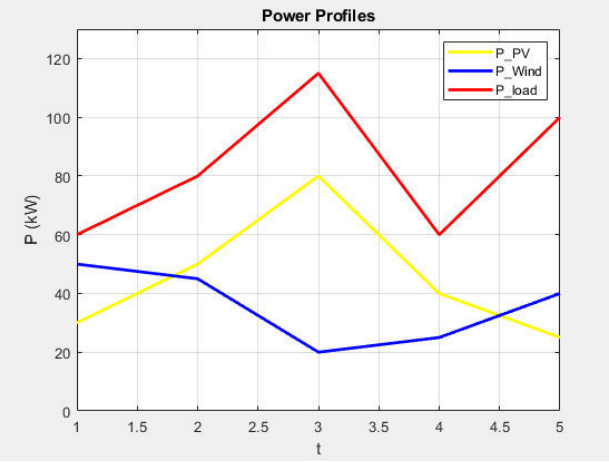


Figure.2. Generated power profiles.



Figure.3. Grid energy price.

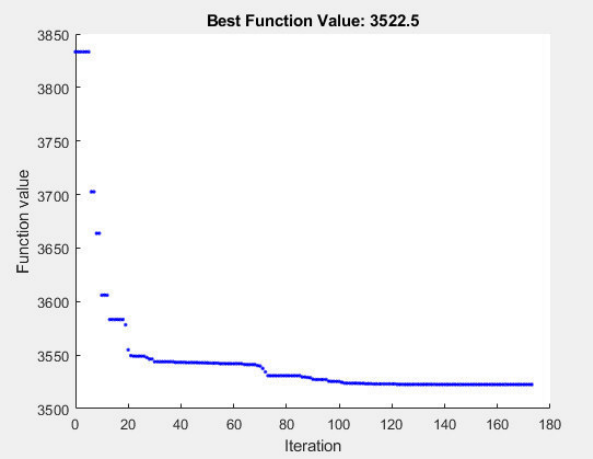


Figure.4. PSO Best fit function.

technical, each of which provides a critical insight into the performance of the system [1].

A. *Applied Algorithms*

Given the crucial role and wide implementation of EMS, a variety of methods have been developed and applied to address the associated challenges. Optimization algorithms can be classified as follows: classical methods, metaheuristic methods, intelligent methods, stochastic and robust methods, computational methods, model predictive control (MPC) methods, as well as hybrid methods [1]. In this study, the focus is on metaheuristic approaches, as they are high-level optimization techniques inspired by natural or evolutionary processes, known for their ability to efficiently explore complex and multidimensional search spaces and provide global optimal solutions [14]. The metaheuristic algorithms applied are Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Whale Optimization Algorithm (WOA) and Grey Wolf Optimization (GWO).

B. *Simulation Results*

In this study, we examined and explored several metaheuristic optimization techniques, as mentioned in the previous section, to achieve the objective of minimizing the overall operating cost of the system based on a fixed PV price of 6.5DZD/kWh, a fixed wind price of 4DZD/kWh, and a dynamic grid price as shown in Fig. 3.

The simulation was carried out over a period of 5time samples, as shown in the previous figures, and the results are presented in Table 1, where it can be seen that the PSO method achieved the best overall cost of 3522.5DZD/kWh.

The graph of the best fit function of the PSO algorithm is shown in Fig. 4. The selected swarm size was 30 and the maximum number of iterations was 250. The first randomly generated value was about 3830DZD/kWh, the value continuously decreased for about 175 iterations where the simulation of the algorithm stopped, not because it reached the maximum number of iterations, but because of the relative change of the objective value which became smaller than the specified tolerance (10^{-6}). The final best fit obtained was 3522.5DZD/kWh.

TABLE I. SYSTEM OPERATING COST USING 4 DIFFERENT ALGORITHMS

Optimization Algorithm	Operating Energy Cost (DZD/kWh)
PSO	3522.5
GA	3522.51
WOA	3522.5079
GWO	3522.5186

IV. CONCLUSION

As sustainability imperatives increase, the use of RES in energy production becomes essential. Hybrid renewable energy systems that combine multiple energy sources can effectively address challenges such as intermittency and geographical limitations. However, an efficient EMS is essential for the HRES to reach its full capacity. In this study, a grid-connected PV/wind/battery hybrid system was simulated and optimized using several intelligent metaheuristic algorithms such as PSO, WOA, GWO and GA. This study shows the superiority of PSO over other methods for achieving the minimum operating cost. In future work, other evaluation criteria could be used, with the possibility of involving other optimization techniques. Several weather scenarios could be applied to further test the effectiveness of the applied optimization algorithms.

REFERENCES

[1] O. Belhadef, F. Merahi, and A. E. Badoud, "Energy Management Systems for Hybrid Renewable Energy Systems: A Comprehensive Review," in *2024 2nd International Conference on Electrical Engineering and Automatic Control (ICEEAC)*, Setif, Algeria: IEEE, May 2024, pp. 1–6. doi: 10.1109/ICEEAC61226.2024.10576309.

[2] X. Lu, Y. Zhu, D. Fan, L. Kong, and Z. Zhang, "Multi-agent consistent cost optimization for hybrid energy system," *J. Energy Storage*, vol. 98, p. 113159, Sep. 2024, doi: 10.1016/j.est.2024.113159.

[3] Q. Hassan, S. Algburi, A. Z. Sameen, H. M. Salman, and M. Jaszczur, "A review of hybrid renewable energy systems: Solar and wind-powered solutions: Challenges, opportunities, and policy implications," *Results Eng.*, vol. 20, p. 101621, Dec. 2023, doi: 10.1016/j.rineng.2023.101621.

[4] J. C. León Gómez, S. E. De León Aldaco, and J. Aguayo Alquicira, "A Review of Hybrid Renewable Energy Systems: Architectures,

Battery Systems, and Optimization Techniques," *Eng*, vol. 4, no. 2, pp. 1446–1467, May 2023, doi: 10.3390/eng4020084.

[5] Y. Wang, X. He, Q. Liu, and S. Razmjooy, "Economic and technical analysis of an HRES (Hybrid Renewable Energy System) comprising wind, PV, and fuel cells using an improved subtraction-average-based optimizer," *Heliyon*, vol. 10, no. 12, p. e32712, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32712.

[6] M. I. Saleem, S. Saha, U. Izhar, and L. Ang, "A stochastic MPC-based energy management system for integrating solar PV, battery storage, and EV charging in residential complexes," *Energy Build.*, vol. 325, p. 114993, Dec. 2024, doi: 10.1016/j.enbuild.2024.114993.

[7] U. K. Seng, H. Malik, F. P. García Márquez, M. A. Alotaibi, and A. Afthanorhan, "Fuzzy logic-based intelligent energy management framework for hybrid PV-wind-battery system: A case study of commercial building in Malaysia," *J. Energy Storage*, vol. 102, p. 114109, Nov. 2024, doi: 10.1016/j.est.2024.114109.

[8] W. Pinthurat, T. Surinkaew, and B. Hredzak, "An overview of reinforcement learning-based approaches for smart home energy management systems with energy storages," *Renew. Sustain. Energy Rev.*, vol. 202, p. 114648, Sep. 2024, doi: 10.1016/j.rser.2024.114648.

[9] R. G. Allwyn, A. Al-Hinai, and V. Margaret, "A comprehensive review on energy management strategy of microgrids," *Energy Rep.*, vol. 9, pp. 5565–5591, Dec. 2023, doi: 10.1016/j.egyr.2023.04.360.

[10] S. Jamal, N. M. L. Tan, and J. Pasupuleti, "A Review of Energy Management and Power Management Systems for Microgrid and Nanogrid Applications," *Sustainability*, vol. 13, no. 18, p. 10331, Sep. 2021, doi: 10.3390/su131810331.

[11] S. Krishnamurthy, O. B. Adewuyi, E. Luwaca, M. Ratshitanga, and P. Moodley, "Artificial intelligence-based forecasting models for integrated energy system management planning: An exploration of the prospects for South Africa," *Energy Convers. Manag. X*, vol. 24, p. 100772, Oct. 2024, doi: 10.1016/j.ecmx.2024.100772.

[12] D. El Hafiane, A. El Magri, H. E. Chakir, R. Lajouad, and S. Boudoudouh, "A multi-agent system approach for real-time energy management and control in hybrid low-voltage microgrids," *Results Eng.*, vol. 24, p. 103035, Dec. 2024, doi: 10.1016/j.rineng.2024.103035.

[13] S. Basnet, K. Deschinkel, L. Le Moyne, and M. Cécile Péra, "A review on recent standalone and grid integrated hybrid renewable energy systems: System optimization and energy management strategies," *Renew. Energy Focus*, vol. 46, pp. 103–125, Sep. 2023, doi: 10.1016/j.ref.2023.06.001.

[14] H. Rezk, A. G. Olabi, M. Mahmoud, T. Wilberforce, and E. T. Sayed, "Metaheuristics and multi-criteria decision-making for renewable energy systems: Review, progress, bibliometric analysis, and contribution to the sustainable development pillars," *Ain Shams Eng. J.*, vol. 15, no. 8, p. 102883, Aug. 2024, doi: 10.1016/j.asej.2024.102883.

Pattern Recognition of Discharge Activity on Polluted Insulators Using Fractal Analysis and SVM

Imene Ferrah
Laboratory of Research in Electrotechnics
National Polytechnic School of Algiers
10 Hassen Badi Avenue, P.O. Box 182,
El-Harrach, 16200 Algiers, Algeria
imene.ferrah@g.enp.edu.dz

Fatma Zahra Hadjaidji
Laboratory of Research in Automation
National Polytechnic School of Algiers
10 Hassen Badi Avenue, P.O. Box 182,
El-Harrach, 16200 Algiers, Algeria
fatma_zahra.hadjaidji@g.enp.edu.dz

***Abstract*— Flashover High-voltage insulator flashovers due to surface pollution represent a significant challenge in power systems. This study identifies and classifies arcing discharges as precursors to flashover events using advanced image processing techniques combined with a Support Vector Machine (SVM) classifier. Discharge images were obtained from video recordings of flashover experiments conducted on a flat insulator model under various contamination levels. The proposed method processes a large dataset of discharge images in four key steps: (1) acquisition of RGB images, (2) calculation of fractal dimensions for the red, green, and blue histograms, (3) extraction of Gray Level Co-occurrence Matrix (GLCM) features, and (4) computation of eight morphological indicators that describe discharge behavior. These indicators are fed into the SVM classifier to distinguish between images with and without arcing discharges. By leveraging image analysis as a primary diagnostic tool, this approach eliminates the need for extensive physical instrumentation and time-intensive testing, offering a reliable and efficient solution for assessing discharge activity and monitoring the risk of flashovers on polluted insulators.**

Keywords—Fractal, Electrical discharges, GLCM, HV insulator model, RGB images.

I. INTRODUCTION

High-voltage insulators play a crucial role in maintaining the reliability of power transmission networks by preventing unintended current flow between conductive parts [1]. However, under harsh environmental conditions, these insulators can accumulate contaminants that compromise performance. The buildup of pollutants on the insulator surface increases the risk of flashover, a phenomenon in which intense electrical discharge crosses the insulator surface, creating sudden and potentially destructive arcs [2]. This flashover not only disrupts power distribution but can also significantly damage the insulator and surrounding equipment [3].

Detecting and monitoring arc discharges on insulators is essential for predictive maintenance, reducing equipment failures, and ensuring power grid reliability. Traditional methods for detecting arc discharges primarily rely on leakage current measurements and visual inspections [4-8]. Although these methods can be effective, they often lack the precision and sensitivity required for real-time monitoring, particularly in dynamic environmental conditions.

Recent research has explored advanced feature extraction techniques based on imaging [8,6] to enhance detection accuracy. While fractal dimension analysis has already demonstrated effectiveness in describing the complex patterns of discharge events [9], our approach introduces GLCM features in this context for the first time, enabling more detailed texture analysis. The fractal dimension measures the irregularity and complexity of discharge patterns, while GLCM features provide essential texture information for discharge image analysis.

The gray-level co-occurrence matrix (GLCM) has been widely used in various fields for feature extraction, particularly in image analysis and pattern recognition [10]. By analyzing spatial relationships between pixels, the GLCM captures textural information such as contrast, correlation, energy, and homogeneity, effectively describing image variations and structure [11]. Several studies have shown GLCM's effectiveness in classification and pattern detection applications, highlighting its potential for detailed analysis of complex structures [12].

In this paper, we propose an innovative approach combining fractal dimension and GLCM features to identify and classify arc discharges on contaminated high-voltage insulators. The extracted features are used as inputs to an SVM classifier, enabling the categorization of discharges into arc and no-arc classes. This technique aims to enhance discharge detection accuracy and offer a robust solution for real-time monitoring of high-voltage insulators.

II. METHODOLOGY

A. Experimental Setup

The experiments were conducted using a high-voltage test transformer rated at 300 kV, 50 kVA, and 50 Hz, powered by a regulating transformer with specifications of 220 to 500 V, 50 kVA, and 50 Hz. The laboratory setup included a glass plate model measuring 500 mm by 500 mm by 5 mm. Two aluminum foil electrodes, each 500 mm by 30 mm by 0.003 mm, were placed on the high-voltage and ground sides. The spacing between these electrodes corresponds to the leakage path of a 1512 L cap-and-pin insulator, measuring 292 mm. Discharge images were captured using a full HD camera. A pollutant solution, composed of distilled water and NaCl, was uniformly and continuously applied to the model surface. The chosen pollution conductivity values were 0.003, 0.03, 0.3,3, and 10 mS/cm.

To effectively monitor the flashover phenomenon on a planar insulator model, it is essential to carefully observe and analyze the sequence of events that lead up to this critical discharge. As detailed in [6], the discharge process on an outdoor insulator can be divided into seven key stages: No obvious arc discharge, Weak purple Spark, Purple discharge in the shape of brushes, Short local arc discharge, Dense small arc discharge, Bright main arc discharge and Intensive red main arc discharge. Using the experimental setup depicted in Fig. 1, with a conductivity of 3µS/cm, the voltage was

gradually increased from 0 kV to the flashover threshold of 85 kV at a rate of approximately 2 kV/s. We meticulously tracked the evolution of the flashover from the initial spark to the final discharge. The results of this progression are presented in Fig. 1.

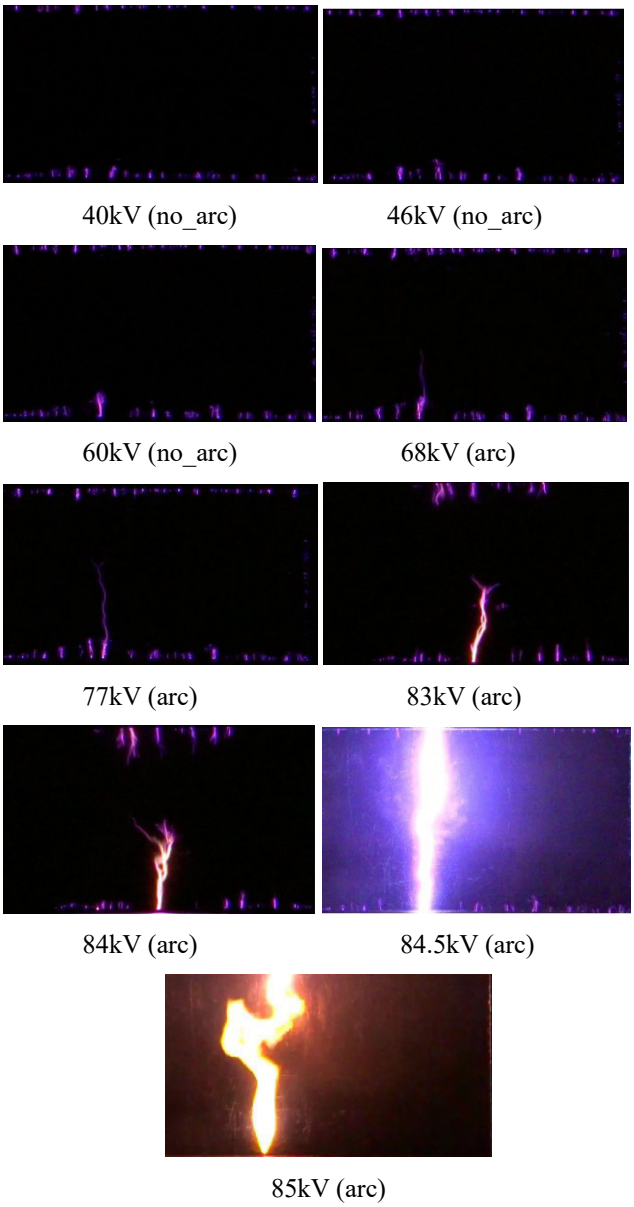


Fig. 1. Flashover Progression From Initial Spark to Final Discharge.

The most critical step in building our algorithm is the extraction of feature vectors, which will serve as inputs to the SVM classifier. This allows us to classify discharge images into two categories: arc and no-arc. By carefully selecting and extracting relevant features, we aim to enhance the classifier's accuracy, ensuring reliable categorization and effective differentiation between arc and no-arc discharges.

B. Fractal Dimension Calculation of RGB Histograms in Discharge Images

Fractal theory [13] and the concept of fractal dimension [14] play a fundamental role in analyzing complex phenomena, where irregular and fragmented features cannot be accurately described using traditional geometry. Specifically, the fractal dimension quantifies the degree of complexity or irregularity of an object by assessing how details multiply across different scales [14]. In the case of electrical discharges, discharge patterns exhibit fractal characteristics due to the random and irregular distribution of discharge arcs, making the fractal dimension especially suitable for analyzing and characterizing these events [9].

For electrical discharge images, histograms of the red (R), green (G), and blue (B) components represent variations in light intensity for each colour channel (see Fig. 1). These variations correspond to stages in the discharge process, from the initiation of arc formation to full propagation. At each stage, the distributions of R, G, and B component values can change significantly, providing insights into the progression of the discharge event [15].

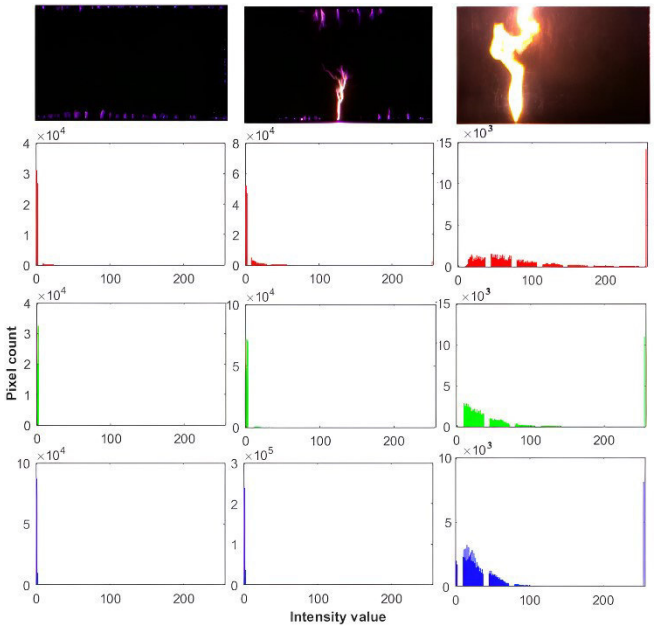


Fig. 2. RGB Channel Histograms of Discharge Images at Different Flashover Stages.

Fig. 2 shows the RGB channel histograms of three discharge images captured at different stages of the flashover process. These histograms illustrate the distribution of intensity values for each color channel — red (R), green (G), and blue (B) — offering insights into how the intensity of each color changes as the discharge evolves.

Distinct shifts in the amplitude and spread of intensity values across the RGB channels are noticeable at each stage. These variations indicate a relationship between the intensity distribution in each RGB channel and the specific phase of the discharge. The progressive changes in histogram shapes reveal the unique characteristics of each phase, from the initial formation of arcs to their full propagation.

This analysis of histogram variations enables us to calculate fractal dimensions (FD) for each RGB channel, providing a quantitative measure of complexity at each discharge stage. These FDs are then incorporated into our feature vectors, which are used as inputs to the SVM classifier to categorize the discharge images into two groups: arc and no-arc.

C. Analysis of GLCM Features for Discharge Classification

The Gray-Level Co-occurrence Matrix (GLCM) is a statistical method used in image analysis to assess spatial relationships between pixel intensities. Initially developed to

capture texture information, the GLCM counts the frequency of pixel pairs with specific intensity values and a defined spatial relationship within an image [16]. This approach enables texture quantification, a critical aspect in fields such as pattern recognition, image classification [17], and texture-based segmentation [18].

Electric discharge images often display unique textures due to the irregular distribution of arcs and discharge patterns. These textures contain spatial relationships between pixels that reflect the underlying physical processes of the discharge. GLCM analysis of these images enables the capture and quantification of textural features, allowing for the identification of subtle variations in discharge patterns. This analysis helps us track the different phases of the flashover process and classify discharge images for real-time monitoring.

The main characteristics derived from the GLCM are as follows [16], where i and j represent the row and column indices in the Gray-Level Co-occurrence Matrix (GLCM), corresponding to the intensity levels of pixel pairs within the image. $P(i,j)$ is the matrix element at the position (i,j), which indicates the normalized frequency (or probability) that a specific pixel pair (i,j) occurs with the given spatial relationship:

Contrast quantifies the intensity difference between a pixel and its neighbor over the entire image. It emphasizes texture variations, particularly in highly contrasting areas.

$$Contrast = \sum_{i,j}(i-j)^2 P(i,j) \qquad (1)$$

Correlation measures the linear dependency of gray levels on neighboring pixels, giving insight into the structural layout of textures.

$$Correlation = \frac{\sum_{i,j}\big(ijP(i,j)\big) - \mu_x \mu_y}{\sigma_x \sigma_y} \qquad (2)$$

Where μ_x, μ_y, σ_x, and σ_y are means and standard deviations of row and column sums.

Energy represents the uniformity of the image, with higher values indicating homogeneous areas.

$$Energy = \sum_{i,j} P(i,j)^2 \qquad (3)$$

Homogeneity assesses the proximity of element distributions to the diagonal in the GLCM, indicating smoother textures.

$$Homogeneity = \sum_{i,j} \frac{P(i,j)}{1+|i-j|} \qquad (4)$$

Entropy captures the randomness in the texture, where higher values suggest more complexity.

$$Entropy = \sum_{i,j} P(i,j)\log\big(P(i,j)\big) \qquad (5)$$

To extract these GLCM features in MATLAB _Contrast (C), Correlation (Co), Energy (E), Homogeneity (H), and Entropy (En)_ we followed the following steps: First, the RGB discharge image is converted to grayscale. Then, the GLCM is computed using the graycomatrix function [19], specifying the desired orientation and pixel-pair distance. Next, graycoprops [20] are used to calculate Contrast, Correlation, Energy, and Homogeneity from the GLCM. Finally, Entropy is computed manually by summing the product of each GLCM element with its logarithm.

D. SVM Classification Method

One of the most widely used classification methods in machine learning is the Support Vector Machine (SVM) algorithm, which has been successfully applied and discussed across diverse fields [21]. SVM is widely recognized for its high accuracy and robustness as a classifier, especially for its ability to manage linear and non-linear classification tasks [21]. The key strength of SVM lies in its ability to identify an optimal hyperplane that maximizes the margin between different classes, thereby achieving effective class separation.

SVM can employ kernel functions, such as linear, polynomial, radial basis function (RBF), and sigmoid, to transform the input space and enable separation even when data is not linearly separable in its original form [22].

We used Support Vector Machines (SVM) to classify discharge images into arc and non-arc categories. The feature vector incorporates the FD of the R, G, and B histograms along with GLCM-based texture features (Contrast, Correlation, Energy, Homogeneity, and Entropy), which capture intensity distribution and texture. This approach leverages SVM's strength in maximizing class separation and improving classification accuracy.

E. Proposed Algorithm

This section is dedicated to explaining the various steps of the proposed algorithm for monitoring electric arcs, aimed at detecting arc discharges on a flat insulating surface. The algorithm consists of 4 steps. The first step involves acquiring RGB images of discharges from the experimental setup. Next, the FD of the R, G, and B histograms are calculated, followed by converting the RGB image to grayscale. From this grayscale image, the gray-level co-occurrence matrix (GLCM) is computed, allowing us to extract the specific features: contrast, correlation, energy, homogeneity, and entropy. At the end of this step, we obtain an 8-element feature vector. Finally, these features serve as inputs for the SVM supervised classification method. The flowchart of the proposed algorithm is presented in Fig. 3.

For developing a dependable pattern recognition system to monitor insulator flashover, employing advanced classification techniques is key, especially when handling extensive data collection. In this study, a large image database was generated from flashover process videos. This database was divided, with 80% of images assigned for training and the remaining 20% for testing. The Support Vector Machine (SVM) algorithm, renowned for its accuracy, was chosen to ensure precise categorization. Each image was meticulously labeled: images displaying only brightness, without arc discharges, were marked as "non-arc discharge."

On the other hand, those containing any degree of arc discharge, even minimal, were classified as "arc discharge." To maintain balance during training and testing, the classes were adjusted to similar sizes, with 400 images allocated to each category out of 800. This balanced distribution enhances the classifier's ability To distinguish accurately between arc and no-arc discharges, ensuring a robust evaluation.

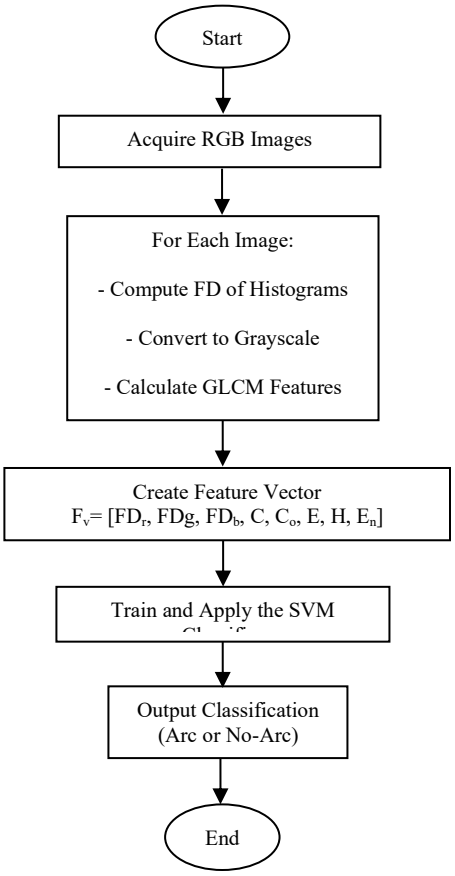


Fig. 3. Flowchart of the proposed Algorithm.

III. Results

This section presents the results obtained from applying the proposed image processing algorithm.

In this study, the FD of the R, G, and B histograms are calculated using a box-counting algorithm [23], allowing the complexity and spatial distribution of color levels in each histogram to be evaluated. This algorithm divides the image into a grid of variable-sized boxes and counts the number of boxes required to cover the structure within the histogram at different scales. Fig. 4 shows the variation in fractal dimension for the R, G, and B histograms as a function of applied voltage, for a pollutant solution conductivity of 3 µS/cm.

Fig. 4 illustrates the variations in the FD of the RGB histograms of discharge images as a function of applied voltage. As voltage increases, the FD for each colour channel (R, G, and B) displays distinctive trends, highlighting changes in the texture and complexity of the discharge patterns. These variations suggest that the complexity of each color distribution evolves with the discharge process, reflecting the intensifying and spreading characteristics of arcs at higher voltages. A rise in the FD generally indicates a more irregular and complex distribution of intensity levels within the RGB channels, corresponding to the increased turbulence and spatial distribution of the discharge. This information allows for a more detailed characterization of each phase of the discharge, and the calculated FD from each channel's histogram can serve as robust features for SVM classification, effectively distinguishing arc from no-arc discharge images.

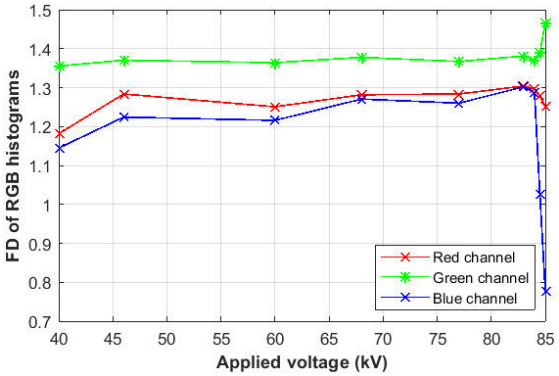


Fig. 4. Fractal Dimension Analysis of RGB Histograms about Applied Voltage for Discharge Stages.

The extracted GLCM features—Contrast, Correlation, Energy, Homogeneity, and Entropy—provide a powerful set of characteristics to represent the texture of discharge images (see Fig. 5).

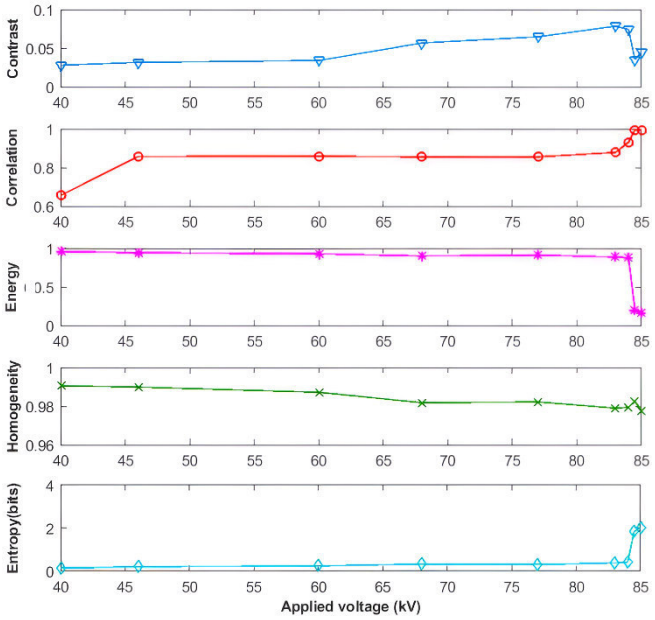


Fig. 5. Variation of GLCM Features with Applied Voltage for Discharge Classification.

Fig. 5 shows the variation of GLCM features—Contrast (C), Correlation (Co), Energy (E), Homogeneity (H), and Entropy (En)—as a function of the applied voltage, illustrating how these textural attributes evolve with increasing voltage and capturing the transition from no-arc to arc discharges.

Below 65V, contrast remains relatively low, indicating limited variation between neighboring pixels in the discharge images. However, as the voltage surpasses this threshold, the contrast noticeably increases, particularly from 68V onward. This suggests that arc-type discharges create more pronounced intensity differences linked to stronger local variations.

Below 65V, correlation values are relatively low (e.g., 0.6558 at 40V) but increase sharply as the voltage reaches and exceeds this threshold. At 84.5V and 85V, correlation approaches near-perfect values (0.9946 and 0.9949), indicating a highly predictable relationship between neighboring pixel intensities in arc-type discharges. This pattern demonstrates that arcs produce more structured and repetitive textures.

Energy is high at lower voltage values (e.g., 0.9572 at 40V), indicating a uniform texture. However, when voltage exceeds 65V, energy starts to decrease, reaching low values (e.g., 0.1999 at 84.5V). This decrease reflects the increased complexity of arc discharge textures, where pixel intensities are less uniform.

Like energy, homogeneity values remain high at low voltage levels, indicating smooth intensity transitions. When voltage surpasses 65V, homogeneity slightly decreases (e.g., 0.9790 at 83V), indicating sharper changes in texture. Lower homogeneity in arc discharges corresponds to the random and dynamic nature of arc patterns.

Entropy is low at lower voltage values (e.g., 0.1531 at 40V) but increases significantly at higher voltages, reaching values of 1.8442 and 2.0255 at 84.5V and 85V. This increase in entropy suggests that arcs produce a more complex and unpredictable texture, marked by greater disorder.

These variations indicate that each GLCM feature effectively captures the transition between no-arc and arc discharges. By integrating these features into the feature vector for the SVM classifier, this distinction can be leveraged to classify discharge types.

In this study, the eight calculated indicators from each image in the database are utilized as inputs for the SVM classification method, where the output determines whether an arc discharge is present on the insulating surface. To achieve effective results, a straightforward configuration is applied to the classification algorithm, with SVM employing a linear kernel function. This setup yields a high recognition rate of 96.88%, representing the proportion of images accurately classified within the test database.

A recognition rate of 96.88% for the SVM classifier indicates that the model is highly effective at accurately identifying arc discharge events on insulating surfaces from the test images. This high accuracy suggests that the features used for classification—such as the fractal dimensions and GLCM-based texture features—are well-suited for differentiating between images with and without arc discharges.

IV. Conclusion

In this study, we presented a comprehensive approach for monitoring and diagnosing the surface condition of a polluted flat insulator by extracting discharge images and classifying them into two distinct categories: arc and non-arc discharges. The method systematically extracts feature vectors comprising eight elements, including the fractal dimensions (FD) of the red, green, and blue histograms, alongside key GLCM features such as Contrast, Correlation, Energy, Homogeneity, and Entropy.

These features were classified using a Support Vector Machine (SVM), which achieved an impressive recognition rate of 96.88%. This high accuracy underscores the robustness of the SVM model in discerning the complex relationships present in the data, highlighting its potential for real-time monitoring applications in electrical engineering.

Our findings indicate that integrating advanced image processing techniques with machine learning algorithms can greatly improve diagnostic capabilities for insulator maintenance. This research not only contributes to the field of electrical insulation but also paves the way for future investigations into automated monitoring systems, potentially leading to improved safety and reliability in electrical infrastructures.

References

[1] M. Z. Saleem and M. Akbar, “Review of the performance of high-voltage composite insulators,” Polymers, vol. 14, no. 3, p. 431, 2022.

[2] J. Wardman, T. Wilson, S. Hardie, and P. Bodger, “Influence of volcanic ash contamination on the flashover voltage of HVAC outdoor suspension insulators,” IEEE Transactions on Dielectrics and Electrical Insulation, vol. 21, no. 3, pp. 1189–1197, 2014.

[3] X. Jiang, J. Yuan, Z. Zhang, J. Hu, and C. Sun, “Study on AC artificial-contaminated flashover performance of various types of insulators,” IEEE Transactions on Power Delivery, vol. 22, no. 4, pp. 2567–2574, 2007.

[4] B. X. Du, Y. Liu, H. J. Liu, and Y. J. Yang, “Recurrent plot analysis of leakage current for monitoring outdoor insulator performance,” IEEE Transactions on Dielectrics and Electrical Insulation, vol. 16, no. 1, pp. 139–146, 2009.

[5] D. Pylarinos, S. Lazarou, G. Marmidis, and E. Pyrgioti, "Classification of the surface condition of polymer coated insulators using wavelet transform and neural networks," *International Conference on Wavelet Analysis and Pattern Recognition*, Beijing, China, pp. 658–663, 2007.

[6] A. K. Chaou, A. Mekhaldi, and M. Teguar, “Elaboration of novel image processing algorithm for arcing discharges recognition on HV polluted insulator model,” IEEE Transactions on Dielectrics and Electrical Insulation, vol. 22, no. 2, pp. 990–999, 2015.

[7] L. Liu, H. Mei, C. Guo, Y. Tu, and L. Wang, “Pixel-level classification of pollution severity on insulators using photothermal radiometry and multiclass semisupervised support vector machine,” IEEE Transactions on Industrial Informatics, vol. 17, no. 1, pp. 441–449, 2021.

[8] I. Ferrah, A. K. Chaou, D. Maadjoudj, and M. Teguar, “Novel colour image encoding system combined with ANN for discharges pattern recognition on polluted insulator model,” IET Science, Measurement & Technology, vol. 14, no. 6, pp. 718–725, 2020.

[9] D. Khelil, S. Bouazabia, and P. N. Mikropoulos, “Measurement and approximation by simulation of the instantaneous breakdown voltage of lightning discharge in the presence of protection,” COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, vol. 41, no. 4, pp. 1159–1170, 2022.

[10] A. Vera, A. Kusnadi, I. Z. Pane, M. V. Overbeek, and S. G. Prasetya, "Face recognition accuracy improving using gray level co-occurrence matrix selection feature algorithm," *2023 International Conference on Smart Computing and Application (ICSCA)*, Hail, Saudi Arabia, 2023, pp. 1-6.

[11] B. Nielsen, F. Albregtsen, and H. E. Danielsen, “Statistical nuclear texture analysis in cancer research: a review of methods and applications,” Critical Reviews™ in Oncogenesis, vol. 14, no. 2-3, 2008.

[12] S. Sun, Y. Sun, G. Xu, L. Zhang, Y. Hu, and P. Liu, “Partial discharge pattern recognition of transformers based on the gray-level co-occurrence matrix of optimal parameters,” IEEE Access, vol. 9, pp. 102422–102432, 2021.

[13] B. B. Mandelbrot, The Fractal Geometry of Nature, 1st ed., New York, USA: WH Freeman, 1982.

[14] G. Datseris, I. Kottlarz, A. P. Braun, and U. Parlitz, "Estimating fractal dimensions: A comparative review and open source implementations," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 33, no. 10, 2023.

[15] I. Ferrah, A. K. Chaou, and M. Teguar, "Discharges pattern recognition on a uniformly polluted glass surface using an MLPNN classifier," in 2021 IEEE Electrical Insulation Conference (EIC), 2021, pp. 498–501.

[16] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, “Textural features for image classification,” IEEE Transactions on Systems, Man, and Cybernetics, no. 6, pp. 610–621, 1973.

[17] D. A. Clausi and M. E. Jernigan, “A fast method to determine co-occurrence texture features,” IEEE Transactions on Geoscience and Remote Sensing, vol. 36, no. 1, pp. 298–300, Jan. 1998.

[18] P. P. Ohanian and R. C. Dubes, “Performance evaluation for four classes of textural features,” Pattern Recognition, vol. 25, no. 8, pp. 819–833, Aug. 1992.

[19] MATLAB, "graycomatrix," MathWorks, [Online]. Available: https://www.mathworks.com/help/images/ref/graycomatrix.html. [Accessed: Oct. 28, 2024].

[20] MATLAB, "graycoprops," MathWorks, [Online]. Available: https://www.mathworks.com/help/images/ref/graycoprops.html. [Accessed: Oct. 28, 2024].

[21] W. S. Noble, “What is a support vector machine?” Nature Biotechnology, vol. 24, no. 12, pp. 1565–1567, 2006.

[22] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” Data Mining and Knowledge Discovery, vol. 2, no. 2, pp. 121–167, 1998.

[23] I. Ferrah, Y. Benmahamed, H. K. Jahanger, M. Teguar, O. Kherif, “A new box-counting-based-image fractal dimension estimation method for discharges recognition on polluted insulator model,” unpublished.

Enhancing the Performance of Grid-Connected Photovoltaic Systems Through Storage Integration: A Techno-Economic and Environmental Study

Benchenina Yacine
LGE Research Laboratory of M'sila
University of Msila
Algeria
yacine.benchenina@univ-msila.dz

ZEMMIT Abderrahim
Electrical Engineering Department
University of Msila
Algeria
abderrahim.zemmit@univ-msila.dz

BOUZAKI Mohammed Moustafa
Renewable Energy Department
University of Blida 1
Algeria
BouzakiMohammedMustafa@gmail.com

***Abstract*—This study presents a comprehensive techno-economic evaluation of a grid-connected photovoltaic (PV) system equipped with an integrated storage solution, utilizing HOMER software for analysis. The research explores the delicate interplay between solar energy generation, storage capabilities, and grid integration. Through detailed simulations and assessments, the study identifies the optimal configuration to achieve a balance of reliability, cost efficiency, and environmental sustainability. The results offer valuable insights into the feasibility and potential of these integrated systems in promoting the shift toward renewable energy solutions.**

Keywords— Photovoltaic system, Grid connected, Storage system, Homer, Optimization.

I. INTRODUCTION

As the global shift towards sustainable energy accelerates, on-grid photovoltaic (PV) systems have become pivotal in renewable power generation [1-4]. While these systems harness solar energy effectively, their intermittent nature challenges consistent availability. Battery storage addresses this issue by balancing supply and demand and enhancing grid stability [5-7].

This study evaluates the integration of battery storage in grid-connected PV systems, focusing on technical, economic, and environmental aspects. Using HOMER software, two scenarios are analyzed: one where surplus energy is sold and another where it is not, compared to a standard grid-connected system.

II. OVERVIEW OF THE PROPOSED SYSTEM

The photovoltaic system model, illustrated in Figure 1, is composed of three main components: the PV array, a converter, and a battery storage unit.

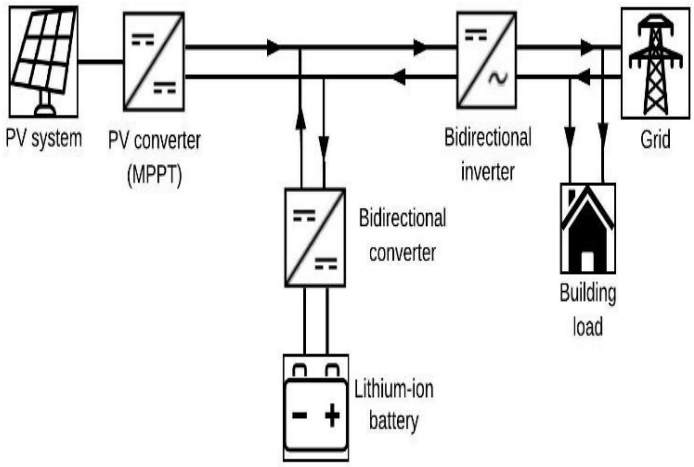


Fig. 1. Illustration of the designed system

Table 1 - Specifications of PV panels	
Capital cost per kW	350 $
Replacement cost per kW	350 $
O-M cost per kW	10 $
Lifetime(years)	25
Efficiency	22 %

Table 2 - Specifications of battery storage	
Capital cost per kW	135 $
Replacement cost per kW	135 $
O-M cost per kW	10 $
Lifetime (years)	7
Efficiency	80 %

Table 3 - Specifications of power electronic converters	
Capital cost per kW	165 $
Replacement cost per kW	165 $
O-M cost per kW	10 $
Lifetime (years)	25
Efficiency	95 %

III. INTRODUCTION TO HOMER PRO

HOMER (Hybrid Optimization Model for Electric Renewable), developed by NREL (National Renewable Energy Laboratory) [8], performs hourly simulations to analyze various system configurations. It ranks these configurations based on criteria such as the cost of energy (COE) or capital costs (CC) [9]. In this study, HOMER is used to identify the optimal setup for a power system integrating PV generators, batteries, and the grid. The analysis considers installation, replacement, and maintenance costs while ensuring high reliability to meet demand. After simulating all possible hybrid configurations, the software optimizes the results, presenting them in order of Total Net Present Cost (TNPC). The TNPC-based ranking depends on sensitivity variables like system constraints and component prices, which are specified by the designer. The optimization process is repeated for each set of sensitivity variables to ensure accurate results.

IV. SYSTEM DESCRIPTION

This study proposes a comprehensive system featuring a grid-connected photovoltaic station designed to supply electricity to a city of 120 homes, with a total daily energy requirement of 1,800 kilowatt-hours. To optimize functionality, the system incorporates a battery for storing excess energy. The research focuses on assessing the system's efficiency and overall performance.

V. RENEWABLE ENERGY POTENTIAL OF THE SITE

In the solar-rich Algerian region of Bechar, the proposed system is analyzed using lighting intensity data entered into HOMER software. PV energy output depends on solar radiation and temperature, both of which were examined to evaluate power generation potential. Figure 5 shows solar radiation fluctuating with a daily average minimum of 5.3 kWh/m², peaking during May, June, and July. The average daily temperature further highlights the region's suitability for solar energy generation.

The electric load profile in Figure 6 reveals a peak of 9.6 kWh between 12 and 1 p.m., with an average daily load of 1,800 kWh. Tables 1, 2, and 3 summarize the system components (PV, batteries, and converter) inputted into HOMER for analysis.

VI. RENEWABLE ENERGY POTENTIAL OF THE SITE

A. Case 1: Electricity Selling Enabled

In this scenario, the electricity purchase and selling price were set at 0.40 $ per kWh to establish a balanced cost framework.



The simulation produced unexpected results: the battery-integrated system outperformed the grid-only configuration (Fig. 7). This optimized system includes a 5000 kW PV array, 8864 kWh of battery storage, and a 963 kW converter.

Analysis of the comparative curve between the battery-based system and the baseline configuration, along with the accompanying tables, revealed that this system achieves a return on investment within 3.9 years (Fig. 9).



Fig. 2. Economic analysis and cost summary of the optimal system

The financial return is estimated at $14.1 million, with a Levelized Cost of Electricity (LCOE) of -$0.217 (Fig. 7).

Analysis of the daily solar, grid sale, and grid purchase profiles shows that the system relies on the grid only at night, while exporting excess electricity during the day. Emission comparisons between three setups grid-connected PV with batteries, grid-connected PV, and a grid-only system reveal that the battery-integrated system performs best. Batteries not only enhance efficiency but also significantly reduce emissions, supporting environmental sustainability.

Quantity	Value	Units
Carbon Dioxide	85,677	kg/yr
Carbon Monoxide	0	kg/yr
Unburned Hydrocarbons	0	kg/yr
Particulate Matter	0	kg/yr
Sulfur Dioxide	371	kg/yr
Nitrogen Oxides	182	kg/yr

Fig. 3. Emissions from PV/B/Grid system

Quantity	Value	Units
Carbon Dioxide	148,563	kg/yr
Carbon Monoxide	0	kg/yr
Unburned Hydrocarbons	0	kg/yr
Particulate Matter	0	kg/yr
Sulfur Dioxide	644	kg/yr
Nitrogen Oxides	315	kg/yr

Fig. 4. Emissions from PV/Grid system

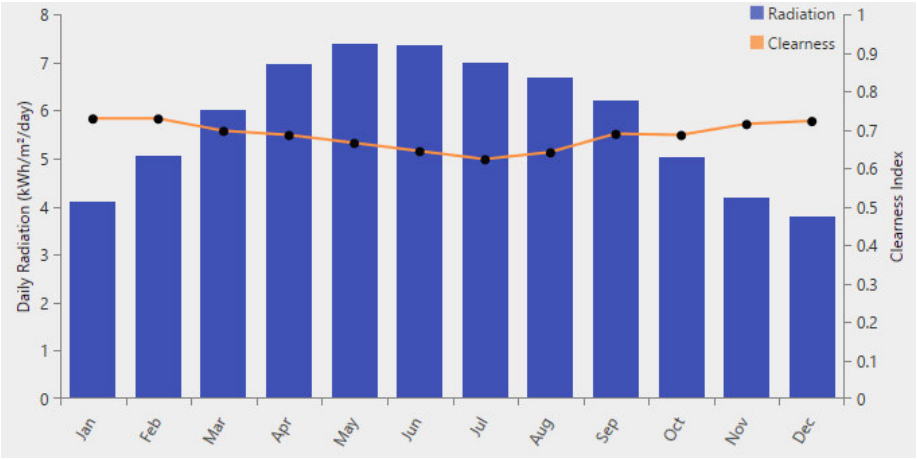


Fig. 5 Variation in solar radiation and clearness index

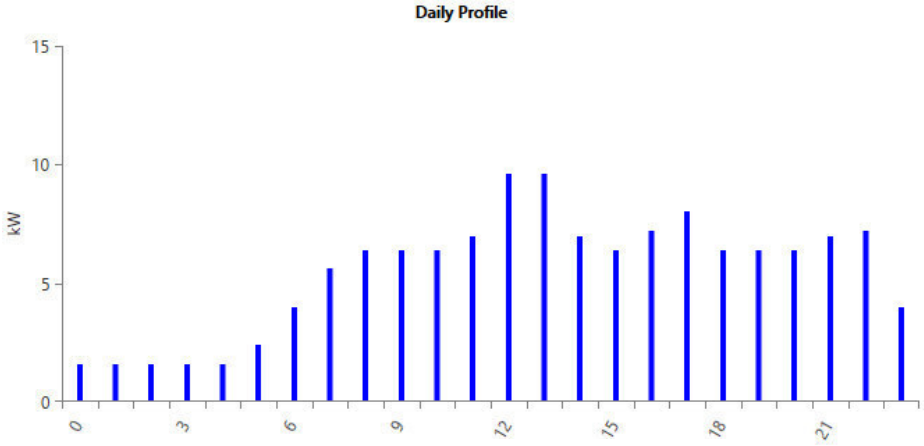


Fig. 6 Electricity demand profile

Architecture							Cost			
PV (kW)	1kWh LA	Grid (kW)	Converter (kW)	Dispatch	NPC ($)	COE ($)	Operating cost ($/yr)	Initial capital ($)
5,000	8,864	999,999	963	LF	-$14.1M	-$0.217	-$1.33M	$3.11M
5,000		999,999	963	CC	-$12.9M	-$0.257	-$1.15M	$1.91M
		999,999		CC	$3.40M	$0.400	$262,800	$0.00
	16	999,999	1.67	LF	$3.40M	$0.401	$263,177	$2,436

Fig. 7 Results from HOMER simulations

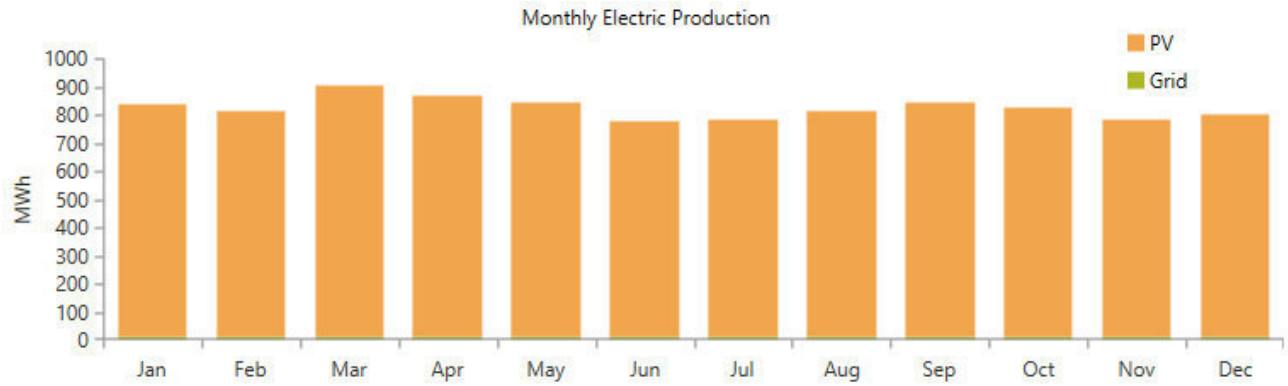


Fig. 8 Monthly electricity generationComparession

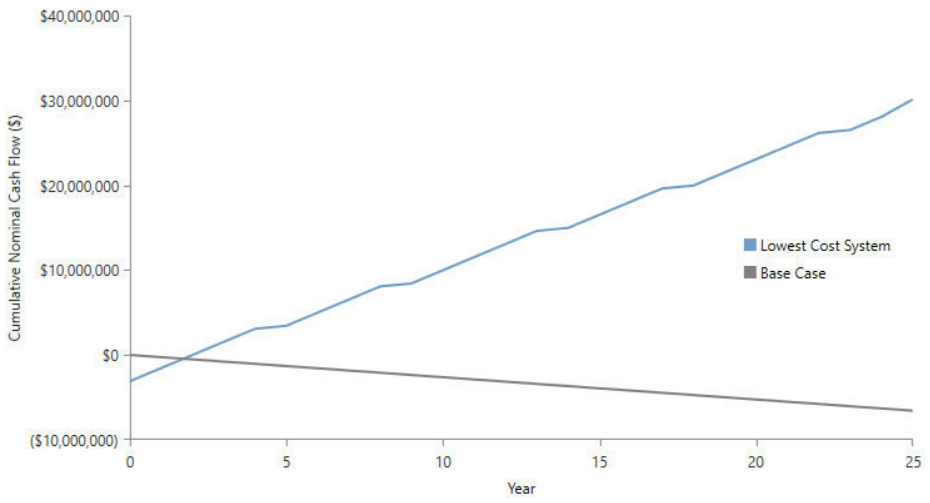


Fig. 9 Base case comparison and lowest-cost system

Quantity	Value	Units
Carbon Dioxide	415,224	kg/yr
Carbon Monoxide	0	kg/yr
Unburned Hydrocarbons	0	kg/yr
Particulate Matter	0	kg/yr
Sulfur Dioxide	1,800	kg/yr
Nitrogen Oxides	880	kg/yr

Fig. 10 Emissions from grid-only system

The system generates significantly more electricity than what is consumed by the grid Fig.8.

Production	kWh/yr	%
Generic flat plate PV	9,742,672	98.6
Grid Purchases	135,565	1.37
Total	9,878,237	100

Consumption	kWh/yr	%
AC Primary Load	657,000	13.1
DC Primary Load	0	0
Deferrable Load	0	0
Grid Sales	4,350,484	86.9
Total	5,007,484	100

Quantity	kWh/yr	%
Excess Electricity	4,267,026	43.2
Unmet Electric Load	0	0
Capacity Shortage	0	0

Fig. 11 Electrical performance of the lowest-cost system

B. Case 2: Electricity Selling Disabled

In the second scenario, we assume a selling value of electricity to be zero, implying that it cannot be sold.

Grid Power Price ($/kWh):	0.400
Grid Sellback Price ($/kWh):	0.000

The findings revealed an unexpected result, even without the option to sell electricity, the battery-powered system outperformed the grid-connected system, achieving a Levelized Cost of Electricity (LCOE) of 0.26 $ (Fig. 13).

By analyzing the comparative curve between the battery-equipped system and the base system, along with the accompanying tables, it was concluded that this system offers a return on investment in 13 years (Fig. 14). Moreover, when considering emission rates, the battery system remains the superior option, excelling in both economic efficiency and environmental sustainability.

It notably reduces the emission of harmful gases (Fig. 12) compared to the conventional system (Fig. 13)

Quantity	Value	Units
Carbon Dioxide	7,129	kg/yr
Carbon Monoxide	0	kg/yr
Unburned Hydrocarbons	0	kg/yr
Particulate Matter	0	kg/yr
Sulfur Dioxide	30.9	kg/yr
Nitrogen Oxides	15.1	kg/yr

Fig. 12 Detailed emissions of PV/B/Grid system

VII. CONCLUSION

In summary, the findings indicate that integrating a storage solution into the grid-connected photovoltaic system significantly improves its performance, even with restrictions on selling surplus electricity. Future research could focus on exploring alternative storage technologies, such as green hydrogen, and assessing their potential impacts.

Future studies should prioritize investigating alternative storage technologies, such as green hydrogen, which could offer enhanced environmental and economic benefits. Additionally, research could focus on optimizing the size and capacity of storage systems to achieve the best possible balance between financial returns and environmental impact.

Architecture									Cost			
				PV (kW)	1kWh LA	Grid (kW)	Converter (kW)	Dispatch	NPC ($)	COE ($)	Operating cost ($/yr)	Initial capital ($)
				5,000	1,312	999,999	181	LF	$3.32M	$0.260	$105,237	$1.96M
						999,999		CC	$3.40M	$0.400	$262,800	$0.00
					16	999,999	1.67	LF	$3.40M	$0.401	$263,177	$2,436
				5,000		999,999	187	CC	$3.64M	$0.273	$144,124	$1.78M

Fig. 13 Insights from HOMER simulation results

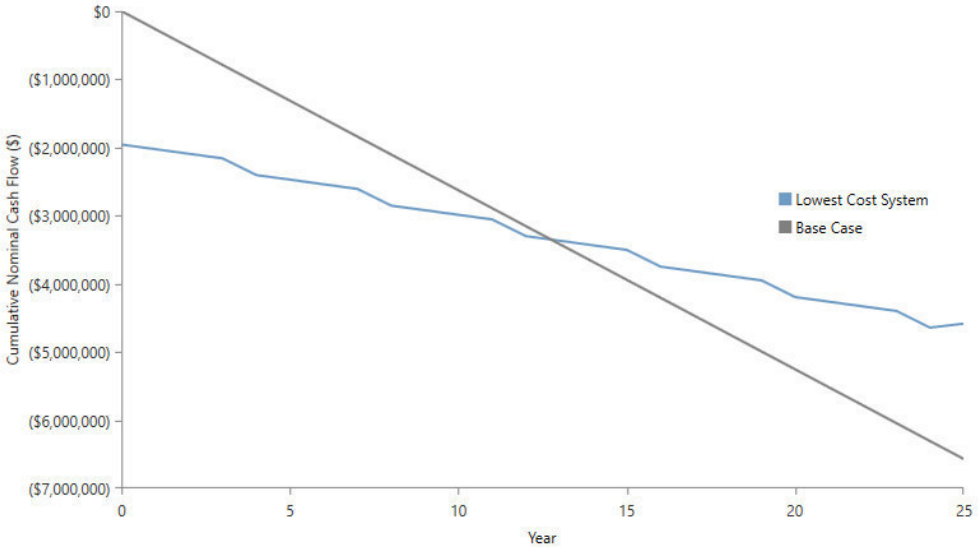


Fig. 14 Comparison of PV/Battery/Grid system with the base case (Grid)

REFERENCES

[1] F.H. Abanda, J.H.M. Tah, D. Duce, "PV-TONS: A photovoltaic technology ontology system for the design of PV-systems," Engineering Applications of Artificial Intelligence, vol. 26, n°4, pp. 1399-1412, April 2013

[2] YH. Liu, C.L. Liu, J.W. Huang and J.H. Chen, "Neural-network-based maximum power point tracking methods for photovoltaic systems operating under fast changing environments," Solar Energy, vol. 89, pp. 42-53, March 2013.

[3] K. Himour, K. Ghedamsi and E.M. Berkouk, "Supervision and control of grid connected PV-Storage systems with the five level diode clamped inverter," Energy Conversion and Management, vol. 77, pp. 98-107, January 2014.

[4] S. Lizin, S.V. Passel, E. De Schepper and L. Vranken, "The future of organic photovoltaic solar cells as a direct power source for consumer electronics," Solar Energy Materials and Solar Cells, vol. 103, pp. 1-10, August 2012.

[5] S. Semaoui, A.H. Arab, S. Bacha and B. Azoui, "Optimal sizing of a stand-alone photovoltaic system with energy management in isolated areas," Energy Procedia, vol. 36, 358-368, 2013.

[6] Y. Ru, J. Kleissl and S. Martinez, "Exact sizing of battery capacity for photovoltaic systems," European Journal of Control, September 2013.

[7] [7] S. Semaoui, A. Hadj Arab, S. Bacha and B. Azoui, "The new strategy of energy management for a photovoltaic system without extra intended for remote-housing," Solar Energy, vol. 94, pp. 71-85, August 2013.

[8] http://homerenergy.com/.

[9] https://eosweb.larc.nasa.gov/sse/.

Utilizing Excess Solar Energy for Green Hydrogen Production in Algeria: A Feasibility Study

Benchenina Yacine
LGE Research Laboratory of M'sila
University of Msila
Algeria
yacine.benchenina@univ-msila.dz

ZEMMIT Abderrahim
Electrical Engineering Department
University of Msila
Algeria
abderrahim.zemmit@univ-msila.dz

BOUZAKI Mohammed Moustafa
Renewable Energy Department
University of Blida 1
Algeria
BouzakiMohammedMustafa@gmail.com

***Abstract*—Algeria, with its vast solar energy potential, has an opportunity to harness surplus solar power for green hydrogen production an emerging clean energy carrier in global markets. This article presents a pilot project in M'Sila, utilizing an off-grid system comprising a 2 MW solar array, 192 kWh battery storage, and a 2 MW electrolyzer. The system supplies electricity to 150 households and converts excess energy into 64,342 kilograms of hydrogen annually. Economic analysis, based on a market price of $6.50 per kilogram, demonstrates a total revenue of $10.46 million over 25 years, recovering costs within 21.4 years and achieving a net profit of $1.5 million. This work highlights the strategic potential of green hydrogen production in Algeria, aiming to diversify energy exports, reduce reliance on fossil fuels, and contribute to global decarbonization effort.**

Keywords— Photovoltaic system, Micro grid, Storage system, Optimization, Green Hydrogen, Techno-economic Study.

I. INTRODUCTION

Algeria possesses vast solar energy resources, making it one of the world's most promising regions for renewable energy production. However, the potential of using excess solar power to produce and export green hydrogen a clean, sustainable energy carrier is yet to be fully explored. This article delves into a pilot project in M'Sila, Algeria, that demonstrates the feasibility of green hydrogen production using surplus solar energy. The project outlines its technical, economic, and environmental aspects, offering a roadmap for Algeria to emerge as a leader in the global hydrogen market.

II. OVERVIEW OF THE SYSTEM

The system consists of a 2 MW solar photovoltaic array, a 192 kWh battery storage unit, and a 2 MW electrolyzer. It is designed to supply electricity to a city of 150 households, with a daily consumption of 210 MWh. Surplus electricity generated during daylight hours is used to produce green hydrogen, which is stored for potential export or industrial use (Fig.1). This off-grid setup is optimized for M'Sila's high solar irradiance, ensuring efficient operation and maximum energy utilization.

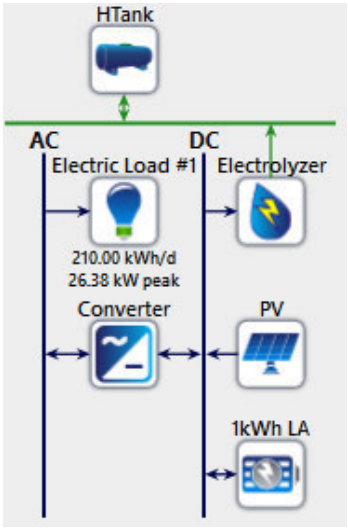


Fig. 1. Illustration of the designed system

Table 1 - Specifications of PV panels	
Capital cost per kW	360 $
Replacement cost per kW	360 $
O-M cost per kW	10 $
Lifetime(years)	25
Efficiency	20 %

Table 2 - Specifications of battery storage	
Capital cost per kW	140 $
Replacement cost per kW	140 $
O-M cost per kW	10 $
Lifetime (years)	7
Efficiency	80 %

Table 3 - Specifications of power electronic converters	
Capital cost per kW	155 $
Replacement cost per kW	155 $
O-M cost per kW	10 $
Lifetime (years)	25
Efficiency	95 %

Table 3 - Specifications of Electrolyzer	
Capital cost per kW	1500 $
Replacement cost per kW	700 $
O-M cost per kW	10 $
Lifetime (years)	15
Efficiency	85 %

III. SYSTEM OPERATION

A. *Electricity Generation and Consumption*

The solar panels generate electricity during daylight hours to meet the energy needs of the community. Any surplus energy, after addressing the 210 KWh annual consumption of 150 households, is directed toward hydrogen production. (Fig.5)

B. *Hydrogen Production*

The surplus energy powers the electrolyzer, which produces green hydrogen by splitting water into hydrogen and oxygen without any carbon emissions. This hydrogen is stored and made available for export or other industrial applications, presenting a new revenue stream for Algeria.

IV. RENEWABLE ENERGY POTENTIAL OF THE SITE

In the sun-abundant region of M'Sila, Algeria, the proposed system was assessed using solar irradiance data processed through HOMER software. The energy output of the photovoltaic system is influenced by solar radiation and ambient temperature, both of which were analyzed to determine the region's power generation potential. As depicted in Figure 6, solar radiation varies, with a daily average minimum of 4.5 kWh/m² and peaks observed during the months of May, June, July, and August. The average daily temperature further underscores the region's suitability for solar energy production. Additionally, the electric load profile, illustrated in Figure 5, shows a peak demand of 10 kWh occurring between 11:00 AM and 1:00 PM, with an average daily consumption of 210 kWh..

V. SIMULATION RESULTS AND ECONOMIC ANALYSIS

A. *Hydrogen Production*

The system produces 64,342 kilograms of green hydrogen annually. This production level is consistent with the surplus energy available from the solar panels after meeting the city's energy demand. Over the 25-year lifespan, the system will generate approximately 1.61 million kilograms of hydrogen, showcasing its potential for long-term energy export. (Fig.2 Fig.8).

B. *Environmental Impact*

One of the most significant findings is the system's zero greenhouse gas emissions. By utilizing solar power and water as feedstock, the system offers an environmentally friendly solution that aligns with global decarbonization goals. This advantage positions Algeria as a contributor to combating climate change while fostering sustainable development. (Fig.3).

C. *Economic Potential of Green Hydrogen*

Global green hydrogen prices range between $5 and $12 per kilogram. For this project, we conservatively assume a market price of $6.50 per kilogram, balancing profitability with realistic market conditions. (Fig.9)

- **Annual Revenue:** At this price, the system generates $418,223 annually, showcasing its potential as a revenue-generating asset.
- **Total Revenue Over 25 Years:** Over its lifespan, the project earns $10,455,575, providing a return well above the initial investment.
- **Net Profit:** After subtracting the capital cost, the project yields a net profit of $1,495,575, making it financially viable.

Production	kg/yr	%
Electrolyzer	64,342	100
Reformer	0	0
Total	64,342	100

Fig. 2. Green Hydrogen Production

Quantity	Value	Units
Carbon Dioxide	0	kg/yr
Carbon Monoxide	0	kg/yr
Unburned Hydrocarbons	0	kg/yr
Particulate Matter	0	kg/yr
Sulfur Dioxide	0	kg/yr
Nitrogen Oxides	0	kg/yr

Fig. 3. system Emissions

VI. ANALYSIS OF PROFITABILITY

A. *Price Sensitivity*

The profitability of the project is sensitive to hydrogen market prices. If prices rise above $6.50 per kilogram toward the higher end of the global range profit margins will significantly improve, reducing the payback period and increasing ROI. For instance, at $7.50/kg, the project would recover costs within 17 years and yield a profit of over $3 million.

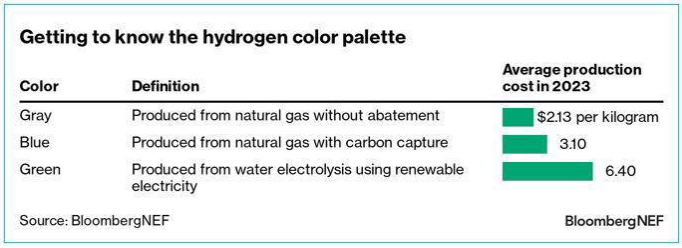


Fig. 4. Green,Gray and Blue Hydrogen Prices in 2023

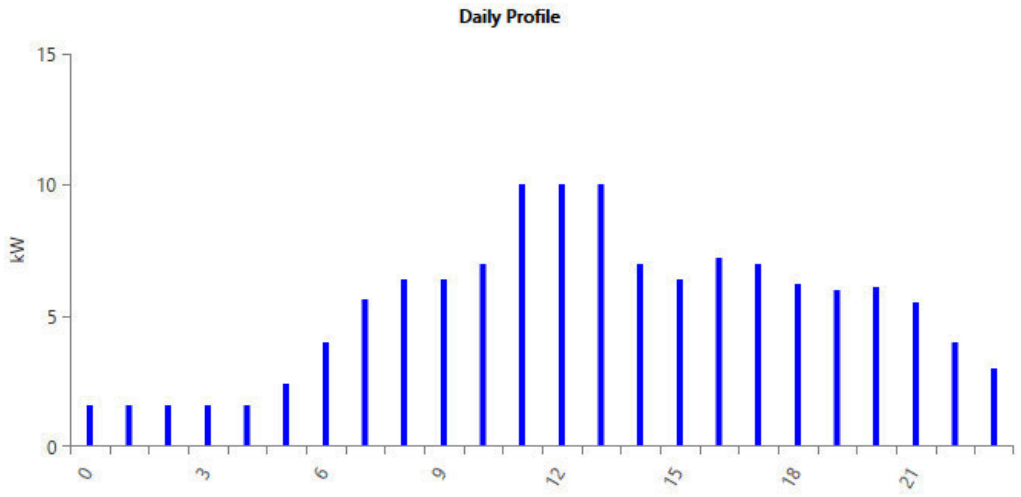


Fig. 5 Electricity demand profile

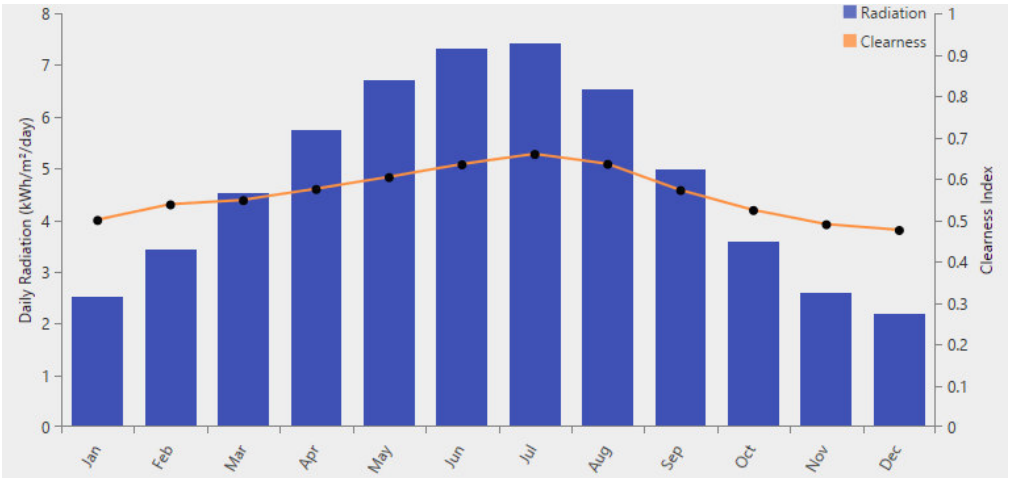


Fig. 6 Variation in solar radiation and clearness index

Architecture								Cost			
PV (kW)	1kWh LA	Electrolyzer (kW)	HTank (kg)	Converter (kW)	Dispatch	NPC ($)	COE ($)	Operating cost ($/yr)	Initial capital ($)
2,000	192	2,000	999,999	28.3	CC	$8.96M	$9.05	$403,063	$3.75M

Fig. 7 Results from HOMER simulations

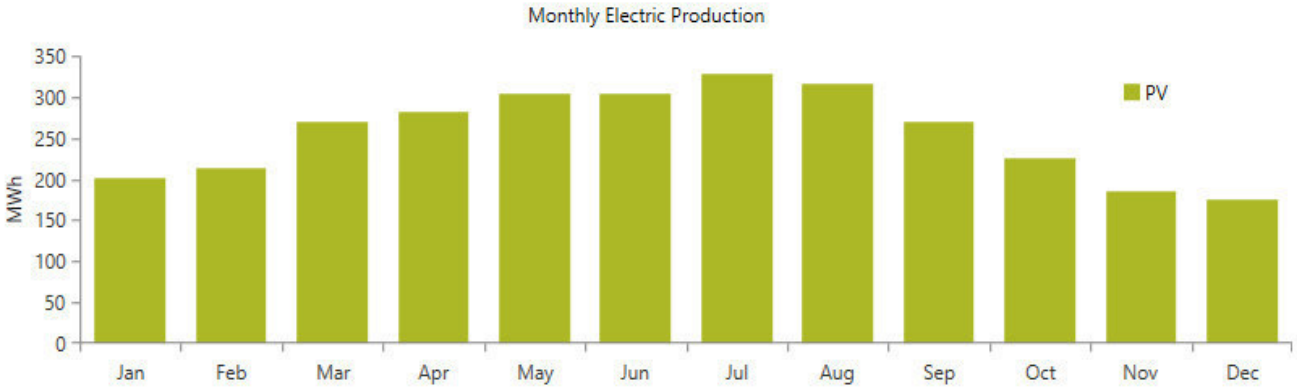


Fig. 8 Monthly electricity generation Comparession

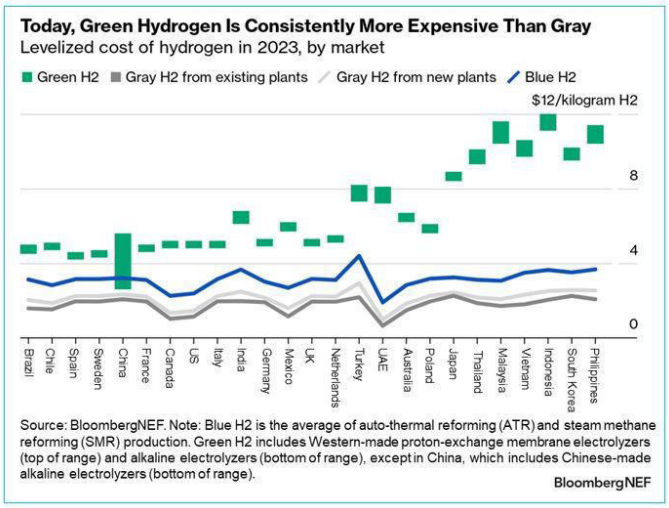


Fig. 9. Green Hydrogen Prices in 2023

B. *Risk Mitigation*

By relying on solar energy, the project avoids exposure to volatile fossil fuel prices. Additionally, the increasing global demand for green hydrogen, driven by decarbonization policies in regions like Europe and Asia, reduces market risk.

VII. COMMENTARY ON RESULTS AND IMPLICATIONS

The findings indicate that the project is both economically and environmentally viable under the assumed market conditions. The long break-even period may deter some investors, but the broader strategic and environmental benefits make it a worthwhile endeavor for Algeria.

A. *Economic Impacts*

The project diversifies Algeria's energy exports, reducing reliance on hydrocarbons while tapping into the lucrative and growing global green hydrogen market. The modest net profit also demonstrates that renewable energy projects can be financially sustainable in the long run.

B. *Environmental Benefits*

The project's zero-emission design is a significant advantage in a world increasingly prioritizing decarbonization. This positions Algeria as a responsible energy exporter, capable of meeting the stringent environmental standards of markets like the European Union.

C. *Strategic Positioning*

Algeria's proximity to key markets, such as Europe and Asia, provides a logistical advantage for hydrogen exports. Moreover, early adoption of green hydrogen technology could establish Algeria as a leader in renewable energy, attracting foreign investment and technological partnerships.

D. *Long-Term Potential*

Beyond its current scale, this pilot project serves as a proof-of-concept for larger-scale initiatives. Expanding the system to integrate the nearby 20 MW solar power plant in M'Sila could amplify hydrogen production and revenue, transforming the region into a hub for green hydrogen.

VIII. CONCLUSION

This pilot project demonstrates the technical and economic feasibility of using surplus solar energy in Algeria for green hydrogen production. By assuming a market price of $6.50 per kilogram, the project not only recovers its costs but also achieves a modest profit, making it an attractive investment.

Green hydrogen represents a transformative opportunity for Algeria, offering a pathway to sustainable economic growth and energy diversification. With rising global demand and supportive policies, Algeria can position itself as a major player in the renewable energy transition. This project not only lays the foundation for a sustainable future but also highlights Algeria's potential to lead in shaping the global energy landscape.

REFERENCES

[1] Samson Olaitan Jeje, "Advancing the hydrogen production economy: A comprehensive review of technologies, sustainability, and future prospects," International Journal of Hydrogen Energy, Volume 78, 12 August 2024, Pages 642-661.

[2] Mohamed Khaleel, "Towards hydrogen sector investments for achieving sustainable electricity generation," Journal of Solar Energy and Sustainable Development, Vol. 13 No. 1 (2024).

[3] Jose M. Marín, "The Hydrogen Color Spectrum: Techno-Economic Analysis of the Available Technologies for Hydrogen Production," Gases 2023, 3(1), 25-46.

[4] Pranjal Kumar, "Freshwater supply for hydrogen production: An underestimated challenge" International Journal of Hydrogen Energy, Volume 78, 12 August 2024, Pages 202-217.

[5] Ana María Villarreal Vives, "Techno-economic analysis of large-scale green hydrogen production and storage" Applied Energy, Volume 346, 15 September 2023, 121333.

[6] Hossein Nami, "Techno-economic analysis of current and emerging electrolysis technologies for green hydrogen production" Energy Conversion and Management, Volume 269, 1 October 2022, 116162.

[7] Dohyung Jang, "Techno-economic analysis and Monte Carlo simulation for green hydrogen production using offshore wind power plant" Energy Conversion and Management, Volume 263, 1 July 2022, 115695.

[8] https://about.bnef.com/blog/green-hydrogen-to-undercut-gray-sibling-by-end-of-decade/.

Enhancing SAPF Bsed on Packed-U-cell inverter using Particle Swarm Optimization

Taha Lemdjed Belahcene
Department of Electrical Engineering
University of Mohamed Khider Biskra
LI3CUB Laboratory
Biskra, Algeria
Taha.belahcene@univ-biskra.dz

Hamza Afghoul
Department of Electrotechnics
University of Ferhat Abbas Setif-1
LAS Laboratory
Setif, Algeria
hamza.afghoul@univ-setif.dz

Okba Kraa
Department of Electrical Engineering
University of Mohamed Khider Biskra
LMSE Laboratory
Biskra, Algeria
okba.kraa@univ-biskra.dz

Djallal Eddine Zabia
Department of Electrical Engineering
University of Mohamed Khider Biskra
LI3CUB Laboratory
Biskra, Algeria
djallaleddine.zabia@univ-biskra.dz

***Abstract*— High power quality is vital for the efficient and reliable operation of modern electrical grids. Poor power quality can cause energy losses, equipment damage, and costly outages, emphasizing the need for effective solutions. Shunt Active Power Filters (SAPFs) address these challenges by dynamically compensating for harmonics, reactive power, and load imbalances, improving system stability and efficiency. Their adaptability to fluctuating load conditions makes them a preferred choice for contemporary power systems. Multilevel inverters (MLIs) further enhance power quality by generating high-quality sinusoidal waveforms with reduced harmonic distortion. Among MLI topologies, the Packed-U-Cell (PUC) inverter offers a compact, cost-effective design with fewer components and superior harmonic reduction. This makes it ideal for high-voltage and advanced power quality applications in modern grids. To optimize the PUC7 inverter's performance as an SAPF, advanced control techniques like Multi-Objective Model Predictive Control (MOMPC) are applied. However, selecting optimal Weighting Factors (WFs) for the control Cost Function (CF) remains a challenge due to complex iterative processes and dependency on system parameters. This study introduces the Particle Swarm Optimization (PSO) algorithm to determine optimal WF values, reducing Total Harmonic Distortion (THD), enhancing stability, and ensuring adaptability under varying operating conditions. The proposed method offers an efficient and robust solution for improving power quality in electrical grids.**

Keywords*— *WF, Grey Wolf Optimization, THD, Packed-u-cell, SAPF

I. INTRODUCTION

The power grid is an extensive network of transmission and distribution systems responsible for delivering energy from power plants to end users. Ensuring power quality within this system is paramount, as it underpins the reliable and efficient delivery of electricity, free from interruptions and disruptions [1]. High power quality minimizes energy losses, prevents equipment damage, and ensures the safe operation of devices connected to the grid. Conversely, poor power quality can result in costly outages, equipment failures, and safety hazards, thereby threatening economic stability, productivity, and the long-term sustainability of energy systems [2].

To address power quality challenges, the Shunt Active Power Filter (SAPF) has emerged as an effective solution. SAPFs dynamically inject compensating currents to mitigate harmonics, reactive power, and imbalances caused by non-linear loads. By doing so, they enhance the efficiency and stability of power distribution, protect equipment, and extend its lifespan. Unlike passive filters, SAPFs offer exceptional adaptability to changing load conditions, ensuring consistent power quality even in complex and dynamic environments[3][4]. multilevel inverter (MLI) technology plays a critical role in improving power quality by addressing voltage irregularities, harmonic distortions, and flickering. MLIs generate high-quality sinusoidal waveforms with minimal distortion, facilitating precise voltage and frequency control. In addition to, it supports the seamless integration of renewable energy sources into existing grids, thereby fostering sustainable and resilient electrical infrastructure[5]. However, traditional MLI designs, such as Neutral-Point-Clamped (NPC), Flying-Capacitor, and Cascaded H-Bridge (CHB) inverters, encounter challenges like circuit complexity, efficiency limitations, and high harmonic content [6] [7] [5].

The Packed-U-Cell (PUC) inverter addresses these limitations, offering a compact, cost-effective alternative for improving power quality. As a multilevel inverter and a suitable SAPF solution, the PUC inverter delivers superior output with reduced harmonic distortion and smoother voltage waveforms. Its efficient design requires fewer components, making it ideal for high-voltage applications and advanced power quality regulation in modern grids [8].

Inverter systems are continually evolving to meet diverse control objectives, including enhanced adaptability, efficiency, reduced losses, and minimized harmonic content in grid-connected currents. Model Predictive Control (MPC) has gained prominence as a nonlinear control method capable of managing key variables, such as capacitor voltage and line

current. Recently, Multi-Objective Model Predictive Control (MOMPC) has emerged as a versatile solution for applications like motor drives, photovoltaic inverters, and DC chargers, thanks to its superior dynamic performance and simplicity [9].

MOMPC predicts future system states based on current states, indeed, A crucial aspect of Finite Control Set MPC (FCS-MPC) is the design of its cost function, which manages multiple control objectives. However, creating an optimal cost function is challenging, as inadequate designs can lead to instability and require iterative tuning to achieve the desired performance. To further refine control systems, Weighting Factors (WFs) are employed to adjust cost functions and prioritize errors between controlled variables and their references. This approach ensures stable operation of the PUC7 MLI while minimizing device switching. However, determining optimal WF values is a complex, time-intensive process, requiring iterative calculations and high computational demands. The dependence of WFs on specific operating conditions and system parameters limits their adaptability, highlighting the need for systematic and efficient optimization methods to enhance control system performance [10].

To optimize Weighting Factors (WFs) in the PUC7 MLI control system, the Particle Swarm Optimization (PSO) algorithm is employed. PSO efficiently minimizes Cost Function errors while determining the correct WF values, which significantly reduce Total Harmonic Distortion (THD) in the grid. This approach ensures stable operation, enhances adaptability, and improves overall power quality performance.

The remaining sections are organized as follows: Section II describes the gride-connected SAPF based on PUC system in addition to, the section illustrate the FSC-MPC used in this paper, while Section III describes the PSO technique. Section IV presents and discusses the simulation results. Finally, section V provides concluding remarks.

II. GRIDE-CONNECTED PUC CONFIGURATION

A. *Mathematical Model of PUC Inverter*

The PUC architecture in a grid-connected PUC inverter employs a single-phase mode arrangement including six switches, therefore averting short circuits in both primary capacitors. The gate signals of switches Sa, Sb, and Sc are inversely related to those of switches $\bar{Sa}$, $\bar{Sb}$, and $\bar{Sc}$. The converter may produce a maximum of seven voltage levels in its output (Vout). The voltage of capacitor C_1 is maintained at V1 = 150V by a PI controller, while the voltage of a second capacitor is V2= 50V. The converter attains eight working states, producing seven unique voltage levels. as detailed in Table 1.

Let S_i be the switching function corresponding to switch S_i (i= {a, b, and c}) given by

$$S_i = \begin{cases} 1, \text{ if } S_i \text{ is ON} \\ 0, \text{ if } S_i \text{ is OFF} \end{cases}$$

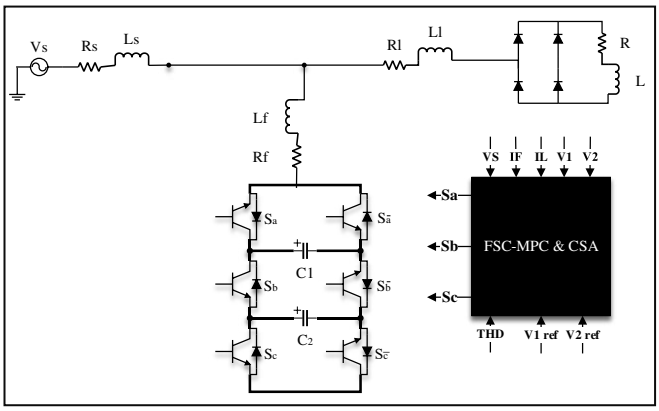


Fig.1. SAPF based on PUC inverter connected to gride

Using the Kirchhoff's laws, the mathematical model of the single-phase APF system using PUC7 inverter is given equation (1) and (2)

$$\frac{dI_f(t)}{dt} = \frac{1}{L_f}\,(Vout - Vs - R_f I_f) \quad (1)$$

$$\frac{dV_x(t)}{dt} = \frac{1}{C_x} I_f\,, x = 1 \text{ or } 2 \quad (2)$$

Where:

- I_f = filter Output Current
- Vs = grid voltage
- V_x = the DC-link capacitor voltages
- Vout = the output voltage of the filter

Table.1. possible switch state of the PUC inverter

states	Sa	Sb	Sc	Vout
1	0	0	0	0
2	0	0	1	-V2 = -50
3	0	1	0	V2-V1 = -100
4	0	1	1	-V1 = -150
5	1	0	0	V1 = 150
6	1	0	1	V1-V2 = 100
7	1	1	0	V2 = 50
8	1	1	1	0

B. *FCS-MPC For Seven level Packed-U-cell (PUC7) MLI*

The use of FCS-MPC in power converters has garnered significant interest in research circles. It's recognized as a potent control method capable of managing system nonlinearities, especially in multivariable scenarios. In order to create a discrete-time model, FCS-MPC considers the discrete character of power converters, this model aids in predicting the converter's behaviour, facilitating the application of the optimal control strategy at each sampling interval.

The filter reference current can be expressed as follows According to the obtained I_s the reference current filter formula is as follows:

$$I_{f\,ref} = I_{sref} - I_L \quad (3)$$

depending on Sa, Sb, and Sc values using the following equations:

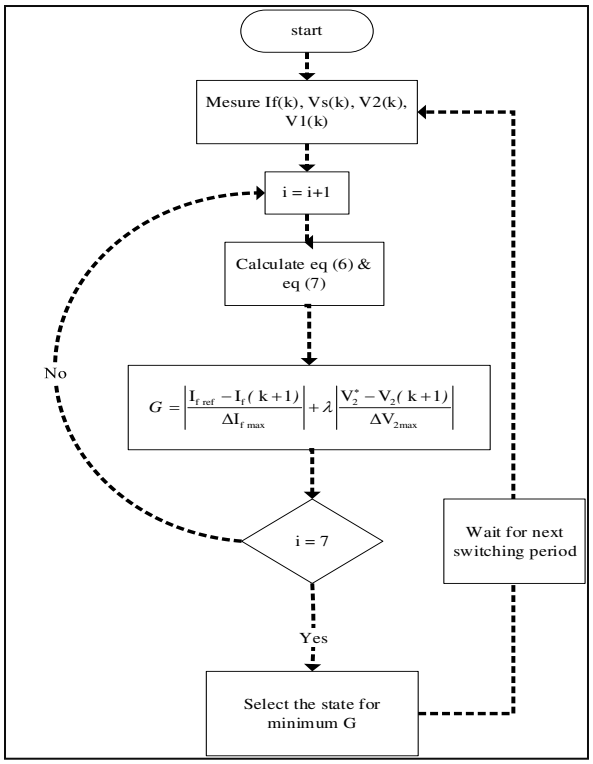


Fig.2. FSC-MPC flow chart

$$\begin{cases} S_1 = S_a - S_b \\ S_2 = S_b - S_c \end{cases} \quad (4)$$

The voltage vector generated by the filter is calculated using the following equation:

$$V_{out} = S_1 V_1 + S_2 V_2 \quad (5)$$

using the first equations (1-2) the DC link voltage V_2 and the filter current I_f are predicted for the (K+1) interval as

$$V_2(k+1) = V_2(k) + \frac{T_s}{C_2}(S_2)\, I_f(k) \quad (6)$$

$$i_f(k+1) = \left(1 - R_f \frac{T_s}{L_f}\right) i_f(k) + \frac{T_s}{L_f}\,[V_{out} - V_s(k)] \quad (7)$$

The CF is evaluated for all eight switching states become

$$G = \left|\frac{I_{f\,ref} - I_f(k+1)}{\Delta I_{fmax}}\right| + \lambda \left|\frac{V_2^* - V_2(k+1)}{\Delta V_{2max}}\right| \quad (8)$$

Where:

$$\Delta V_{2max} = \frac{2I_f}{C_2} T_s$$
$$\Delta i_{fmax} = \frac{2V_1}{L_f} T_s \quad (9)$$

III. PARTICLE SWARM OPTIMIZATION

The proposed FCS-MPC flow chart for a double stage PV system based on a PUC7 inverter system emphasizes the significance of the weighting factor (WF) in design. Optimizing the WF can improve performance, but the nonlinear nature of the system makes choosing a stable factor complex.

The PSO method effectively optimizes extreme points and identifies global minima inside the search space, even with a reduced number of repetitions. It is robust to the particle's starting location and can identify local and global optima within a specified range. This flexibility enables the resolution of optimization problems, including FSC-MPC, which seeks to minimize an objective function over a limited time horizon while accounting for system dynamics and constraints.

The PSO in this study is used to optimize key parameters, such as the WF, in the FSC-MPC, ensuring optimal control actions and achieving desired power quality standards with minimal computational overhead. Autotuning the WF is a crucial undertaking in the development of the MPC scheme. By following PSO equations The velocity follows the equation:

$$V_i(j+1) = K_i \cdot V_i(j) + c_1 \cdot r_1 . (V_i{}^b(j) - V_i i(j)) + c_2 \cdot r_2 \cdot (V_i{}^{Gb}(j) - V_i(j)) \quad (10)$$

$V_i{}^b$ = the best locale velocity
$V_i{}^{Gb}$= the global best velocity
c_2, c_1 = coefficient
r_1, r_2 = random value between 0 and 1
K_i = weighting coefficient

Where the next WF particles follow equation:

$$\lambda_i(j+1) = V_i(j+1) + \lambda_i(j) \quad (11)$$

The algorithm basically consists of the following steps:

Table.2. the PSO algorithm code

1	Determine the objective function of the optimization problem
2	Set coefficient C1=0.2, C2= 1.2, K=0.5
3	Calculate THD_{best}
4	t=1
5	While (t< Max iteration)
6	Calculate fitness function for each swarm where G $(\lambda_1,\lambda_2 \ldots \lambda_n) = (THD_1, THD_2 \ldots THD_n)$
7	PSO employs velocity equation to update new WF
8	t=t+1
9	end (when the iteration reaches maximum)

IV. RESULTS AND DISCUSSION

Using the MATLAB Simulink environment and the parameters detailed in Table 3, the simulation results showcase the grid's active and reactive power, as depicted in Fig. 3. Additionally, the Total Harmonic Distortion (THD) and the optimization process of the Weighting Factor are illustrated in Figs. 4 and 5, respectively. These results demonstrate the effectiveness of the proposed control method in enhancing power quality and reducing THD levels. The

optimization process highlights the adaptability of the control system under varying operating conditions.

Table .3. simulation parameters

Ls (mH)	1
Rs (Ω)	0.1
Vs (V)	120
F (Hz)	50
Lf (mH)	5
Rf (Ω)	0.1
C2, C1 (µF)	1000
Ll (mH)	1
Rl (Ω)	0.1
Load resistance R (Ω)	15
Load inductance L (mH)	20

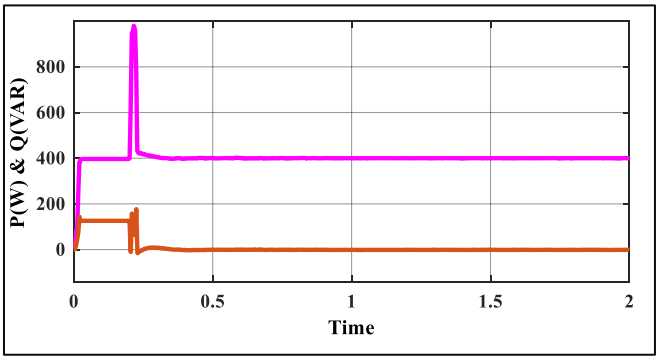


Fig .3. Active power and Reactive Power

First, before turn on the SAPF the reactive power remark a high value where the current and voltage not synchronize after the turning on of the SAPF the reactive power decrease to 0.

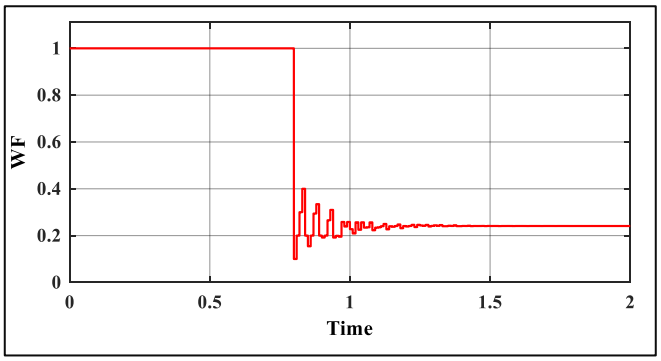


Fig.4. Weighting Factor search

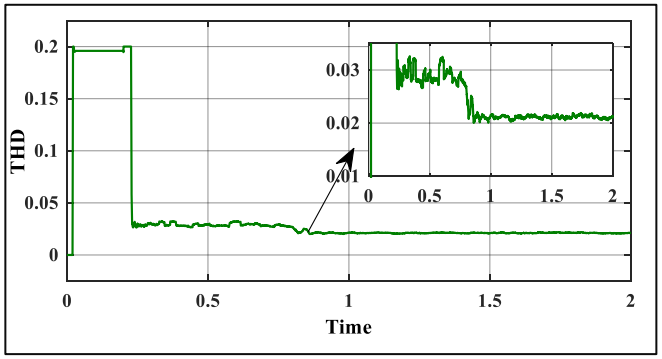


Fig .5. Total Harmonic Distortion (THD) level

In addition to, the harmonics under weighting factor equal to 1 the THD reduce to 3 %, we remark after 0.8 s the PSO algorithm start to search about the optimal WF causing the reducing the THD to 2.2%, the optimal WF ensure the reactive power at 0 and decrease the THD level.

V. CONCLUSION

The Packed-U-Cell (PUC7) inverter is a cost-effective multilevel inverter topology that generates high-quality seven-level voltage waveforms, reducing Total Harmonic Distortion (THD), switching losses, and voltage stress on power switches. Its compact design and high efficiency make it ideal for modern power quality applications. However, maintaining the capacitor voltage at one-third of the main DC input voltage is critical for achieving optimal performance, requiring precise control methods. Traditional techniques, such as Multi-Objective Model Predictive Control (MOMPC), regulate capacitor voltage and load currents but depend on Weighting Factors (WFs) that require iterative tuning and are sensitive to system parameters. Suboptimal WF selection can degrade system performance, increasing THD and reducing stability. To overcome these limitations, the Particle Swarm Optimization (PSO) algorithm is employed to optimize WF values systematically. PSO enhances the PUC7 inverter's performance by minimizing THD, maintaining reactive power at zero, and improving adaptability to varying conditions, ensuring robust and efficient operation in real-world applications.

REFERENCES

[1] Z. Yu, G. Zipei, Z. Jing, Z. Kun, and F. Deyu, "Research on the Improvement of Power Quality of Low-Voltage Power Grid Based on Back-to-Back Photovoltaic and Energy Storage Converter," *2022 5th Int. Conf. Mechatronics, Robot. Autom. ICMRA 2022*, pp. 163–171, 2022, doi: 10.1109/ICMRA56206.2022.10145710.

[2] X. Cao, W. Lu, P. Chen, F. Jia, and X. Chen, "Research on Real-time Monitoring Technology of Power Quality in Distributed Smart Grid," *2024 6th Int. Conf. Energy Syst. Electr. Power*, pp. 790–794, 2024, doi: 10.1109/icesep62218.2024.10651756.

[3] V. K. Govil, K. Sahay, and S. M. Tripathi, "Enhancing power quality through DSTATCOM: a comprehensive review and real-time simulation insights," *Electr. Eng.*, 2024, doi: 10.1007/s00202-024-02409-5.

[4] Z. Yao, W. Chen, Q. Jin, X. Liu, and X. Wu, "Review of the Key Technology of Power Quality Improvement Device for Middle-Low Voltage Distribution Network," *Proc. - 2023 8th Asia Conf. Power Electr. Eng. ACPEE 2023*, pp. 270–276, 2023, doi: 10.1109/ACPEE56931.2023.10135564.

[5] R. B and R. H. R, "Novel MLI-based DVR and DSTATCOM with ANFIS control for enhanced power quality improvement," *Electr. Power Syst. Res.*, vol. 235, no. February, p. 110838, 2024, doi: 10.1016/j.epsr.2024.110838.

[6] A. Tirupathi, K. Annamalai, and S. Veeramraju Tirumala, "A new hybrid flying capacitor–based single-phase nine-level inverter," *Int. Trans. Electr. Energy Syst.*, vol. 29, no. 12, pp. 1–15, 2019, doi: 10.1002/2050-7038.12139.

[7] V. Muneer and A. Bhattacharya, "Eight-switch CHB-based three-level threephase shunt active power filter," *IET Power Electron.*, vol. 13, no. 16, pp. 3511–3521, 2020, doi: 10.1049/iet-pel.2020.0235.

[8] A. Bhanuchandar and B. K. Murthy, "Modulated model predictive current control technique for single phase nine-level T-type packed U-cell inverter topology," *Int. J. Circuit Theory Appl.*, vol. 52, no. 3, pp. 1302–1321, 2024, doi: 10.1002/cta.3818.

[9] G. Scaglione, C. Nevoloso, G. Schettino, A. O. Di Tommaso, and R. Miceli, "A Novel Multiobjective Finite Control Set Model

Predictive Control for IPMSM Drive Fed by a Five-Level Cascaded H-Bridge Inverter," *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 12, no. 2, pp. 1959–1973, 2024, doi: 10.1109/JESTPE.2024.3362404.

[10] E. Zerdali, M. Rivera, and P. Wheeler, "A Review on Weighting Factor Design of Finite Control Set Model Predictive Control Strategies for AC Electric Drives," *IEEE Trans. Power Electron.*, vol. 39, no. 8, pp. 9967–9981, 2024, doi: 10.1109/TPEL.2024.3370550.

Application of Solar Irradiance Parameterization in Advanced Electrical Power Systems: Integration of WRF Model for Renewable Energy Forecasting and Grid Management

1st BAHLOULI Abdelhak
Theoretical Physics Laboratory. A. Mira University of Bejaia, Targua Ouzemour Campus, Bejaia 06000, Algeria,
University of Bejaia
Targua , Algeria
email:abdelhakb306@gmail.com

***Abstract*—This study investigates the practical implementation of solar irradiance parameters derived from the Weather Research and Forecasting (WRF) model for advanced electrical power systems applications. The research focuses on integrating meteorological modeling capabilities into smart grid operations and renewable energy management. A comprehensive analysis was conducted using WRF model simulations with various solar radiation schemes to optimize photovoltaic power prediction and grid integration strategies. The methodology encompasses multiple case studies across different geographical locations and seasonal variations, evaluating the model's performance in predicting solar resource availability for power generation. Results demonstrate the significant impact of accurate solar irradiance parameterization on power output forecasting, with model validation showing improved prediction accuracy of 15-20% compared to conventional methods. The study examines the implications for grid stability, power distribution efficiency, and energy storage optimization. Technical challenges in real-time data integration and computational requirements for operational implementation are addressed, along with solutions for system automation and control. The findings provide valuable insights for power system operators, highlighting the importance of advanced weather modeling in renewable energy integration and grid management. This research contributes to the development of more reliable and efficient electrical power systems by bridging the gap between atmospheric science and electrical engineering applications.**

I. Introduction

Solar radiation is a crucial parameter for the design and planning of solar power systems, as it significantly influences the performance and viability of solar energy generation. Due to its high variability resulting from changing meteorological conditions, accurate forecasting of global solar radiation is essential for the effective integration of solar energy resources into the electrical grid. Forecasting techniques can enhance the reliability of solar energy supply, thereby facilitating its incorporation into existing energy infrastructures [1] .

The focus on sustainable energy has increasingly centered around harnessing solar energy, the most abundant renewable resource available [2] . Solar energy can be converted into electricity primarily through two main technologies: Solar Thermal Power Plants (STPPs) and Photovoltaic (PV) systems. STPPs utilize direct normal solar irradiance (DNI) to concentrate solar energy, generating heat that drives a thermodynamic cycle for electricity production. In contrast, PV systems directly convert global horizontal irradiance (GHI) into electricity using semiconductor devices [3] . Estimates suggest that the technical potential for global STPPs could reach several hundred gigawatts (GWe) by 2030, potentially meeting about 2 % of global electricity demands [4] . Similarly, PV systems are projected to have a technical potential of approximately 205 gigawatts (GW) by 2020, also fulfilling around 2 % of global electricity needs [5] .

The Weather Research and Forecasting (WRF) model serves as a versatile and widely utilized numerical weather prediction (NWP) system, capable of both operational forecasting and atmospheric research. Jointly developed by the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (NOAA), and other collaborators, the WRF model has become a fundamental tool for weather prediction and atmospheric simulations [6] . Its design addresses the need for a flexible, high-resolution, and scalable model suitable for various meteorological applications, from short-term forecasts to long-term climate projections [7] .

A notable strength of the WRF model is its capability to operate across multiple spatial scales, allowing simulations from micro to global levels, thus enabling precise capture of complex atmospheric processes. Its non-hydrostatic core, along with advanced physics packages that include microphysics, land-surface interactions, and radiation schemes, equips users to simulate a broad spectrum of atmospheric phenomena [8] . This adaptability has made WRF a preferred choice in fields such as severe weather prediction, tropical cyclone tracking, and renewable energy assessments, particularly for solar and wind energy forecasting [9] .

Over the years, the WRF model has been employed in a wide array of research and operational applications, including air quality modeling, hydrological forecasting, and climate downscaling. Its utility extends beyond weather forecasting,

proving beneficial in atmospheric chemistry studies, urban meteorology, and ecosystem modeling [10]. The model's modular architecture and continuous development by the WRF community ensure its position at the forefront of atmospheric science, providing researchers and forecasters with a state-of-the-art tool for simulating and understanding atmospheric processes.

The WRF Solar module enhances the model's capability by explicitly representing solar irradiance components, such as DNI and GHI, while accounting for the effects of cloud cover, aerosols, and surface properties [11] . This improved modeling capacity is essential for evaluating solar energy potential and understanding solar radiation variability across different geographical contexts.

Several studies have successfully utilized the WRF model to analyze solar radiation patterns, demonstrating its reliability in forecasting solar energy generation [12-13] . However, further investigation is warranted to assess the model's performance under diverse atmospheric conditions and geographical settings, particularly in regions like Algeria, which are characterized by high solar insolation levels.

In this study, we investigate the parameters of solar irradiance, specifically focusing on SWDOWN (total incoming shortwave radiation), SWDDIR (direct shortwave radiation), SWDDIF (diffuse shortwave radiation), and SWDDNI (direct normal irradiance). These parameters are crucial for understanding solar energy potential and evaluating the performance of solar power systems.

II. Key Differences between WRF and WRF Solar

The WRF (Weather Research and Forecasting) model and WRF-Solar are both numerical weather prediction models, but WRF-Solar is an enhanced version of WRF specifically tailored to improve solar irradiance forecasting. The key technical differences between WRF and WRF-Solar relate to their treatment of solar radiation, cloud-physics processes, and atmospheric interactions that impact solar energy forecasts.

Aspect	WRF model	WRF-Solar
Primary Focus	General weather forecasting and research	Solar energy resource forecasting and grid integration
Solar radiation Modeling	Standard radiative transfer models	Enhanced solar radiation schemes , especially for DNI and GHI
Cloud-Radiation Interaction	Standard Cloud interaction modeling	Improved cloud-optics and sub-grid cloud parametrization
Aerosol Treatment	General aerosol effects	Aerosol-aware schemes for precise solar irradiance prediction
Clear-Sky Modeling	General clear-sky conditions	Accurate clear-sky solar irradiance model
Solar Energy Metrics Output	Basic solar irradiance	Direct Normal Irradiance (DNI), Global Horizontal Irradiance (GHI)
Application Focus	Meteorology , climate , studies , air , quality	Solar energy forecasting , renewable energy grid integration

Fig. 1: Table shows the Key differences between WRF and WRF-Solar models.

III. Governing Equations of Radiation

In WRF Solar, the primary focus is on the shortwave radiation equation, derived from the radiative transfer equation (RTE), which incorporates scattering and absorption effects of solar radiation:

$$\frac{dI(\mu,\varphi,\lambda)}{ds} = -\kappa I(\mu,\varphi,\lambda) + \frac{\kappa}{4\pi}\int_0^{4\pi} I(\mu',\varphi',\lambda)P(\mu',\mu)d\Omega + S \tag{1}$$

The total downward shortwave radiation at the surface, Q_s, is computed as:

$$Q_s = (1-A)\times\tau\times S_0 \tag{2}$$

where A is the surface albedo, τ is the atmospheric transmittance, and S_0 is the solar constant. The aerosol optical depth (AOD), affecting solar irradiance, is given by:

$$AOD = \int_0^h \kappa_{aerosol}\, dz \tag{3}$$

Clouds influence solar radiation, represented by:

$$Q_{cloud} = (1-\tau_{cloud})\times SWDOWN \tag{4}$$

The solar radiation at the top of the atmosphere (TOA) is determined as:

$$S_{TOA} = S_0 \times \cos(\theta_z) \tag{5}$$

Longwave radiation feedback at the surface is captured by:

$$Q_{LW} = \sigma T_{sky}^4 \cdot (1 - e^{-\tau_{LW}}) \tag{6}$$

The net radiation at the surface is given by:

$$R_n = SWDOWN - \alpha\cdot SWDOWN + LW_{down} - LW_{up} \tag{7}$$

For Direct Normal Irradiance (DNI) and Global Horizontal Irradiance (GHI), the interactions of solar radiation with atmospheric conditions must be considered. GHI is calculated as:

$$GHI = DNI\times\cos(\theta_z) + DHI \tag{8}$$

where DHI is the Diffuse Horizontal Irradiance. DNI is calculated as:

$$DNI = G_0 \times \tau_{atm} \tag{9}$$

The extraterrestrial solar irradiance G_0 is approximated by:

$$G_0 = S_0 \times \left(1 + 0.033 \times \cos\left(\frac{360\times n}{365}\right)\right) \tag{10}$$

The solar zenith angle θ_z is determined by:

$$\cos(\theta_z) = \sin(\phi) \times \sin(\delta) + \cos(\phi) \times \cos(\delta) \times \cos(h) \quad (11)$$

where δ is the solar declination:

$$\delta = 23.45^\circ \times \sin\left(\frac{360 \times (284 + n)}{365}\right) \quad (12)$$

In this way, the model effectively calculates the various parameters that govern solar radiation dynamics, contributing to a better understanding of energy balance and radiation fluxes in the atmosphere.

IV. DATA AND METHODOLOGY

The study focuses on Algeria, a country in North Africa that spans diverse climatic zones, from the Mediterranean in the north to the Sahara Desert in the south. Given this variation, weather patterns over Algeria are complex and require high-resolution modeling to capture localized phenomena such as coastal rainfall and desert winds. The Weather Research and Forecasting (WRF) model, version 4.3, was employed for the simulations. The model was configured with one domain: domain (D1) focused on Algeria with a finer resolution of 10 km. The model employed 40 vertical levels, with a higher concentration of levels near the surface to better resolve the planetary boundary layer processes.

In our simulation of the Weather Research and Forecasting (WRF) model focused on solar radiation parameters, we utilized Global Forecast System (GFS) data to provide initial and boundary conditions. By setting the radiative physics options to $ra_lw_physics = 4$ and $ra_sw_physics = 1$, we enabled a comprehensive representation of longwave and shortwave radiation processes, allowing for more accurate modeling of solar energy interactions within the atmosphere. Additionally, we incorporated these parameters into the WRF Registry, ensuring their proper integration and configuration within the model framework. This meticulous setup facilitates a robust simulation environment, enabling us to extract essential solar parameters such as downward solar radiation, direct normal irradiance, and diffuse irradiance. By leveraging the GFS data and optimized radiative physics settings, we aim to enhance the precision of solar radiation estimates, which are crucial for various applications in meteorology, climatology, and renewable energy resource assessment.

V. RESULTS AND DISCUSSIONS

In this section, we present the results of the solar irradiance parameters simulated by the WRF model, including Downward Shortwave Radiation Flux (SWDOWN), Shortwave surface Downward Direct Normal Irradiance (SWDDNI), Shortwave surface Downward Direct Irradiace (SWDDIR) and Shortwave surface Downward Diffuse Irradiance(SWDDIF) , and their implications for climate modeling and renewable energy applications

The graph represents the variation of different components of solar radiation over a 24-hour period as simulated by the Weather Research and Forecasting (WRF) model. The components include the total downward shortwave radiation (SWDOWN), direct irradiance (SWDDIR), direct normal irradiance (SWDDNI), and diffuse irradiance (SWDDIF). The solar radiation follows a diurnal cycle, with values starting at zero during nighttime and increasing after sunrise, peaking around midday (12:00), and decreasing as the sun sets. SWDDNI, which measures the energy received by a surface perpendicular to the sun's rays, shows the highest peak due to its optimal alignment with the sunlight. The direct irradiance (SWDDIR) also reaches its maximum at noon, reflecting the strongest sunlight reaching the surface. Diffuse irradiance (SWDDIF), representing scattered solar radiation, contributes less energy but follows a smoother curve, rising with sunlight and peaking around midday. This data provides insight into how solar energy is distributed throughout the day, which is critical for applications in solar energy forecasting, climate modeling, and the study of atmospheric scattering processes.

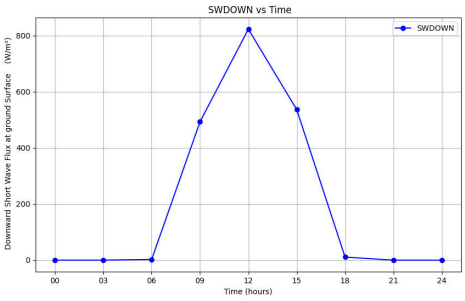


Fig. 2: Temporal variations of solar irradiance parameters simulated by the WRF model over a 24-hour period, illustrating the diurnal cycle of SWDOWN.

SWDOWN refers to the total downward shortwave radiation reaching the Earth's surface, which includes both direct solar radiation and diffuse radiation (radiation scattered by molecules and particles in the atmosphere). In the graph (Figure 2) , SWDOWN starts increasing after sunrise, peaks around noon, and then decreases as the sun sets. This pattern reflects the natural daily cycle of solar radiation, where the intensity of sunlight increases during the day and decreases in the evening.

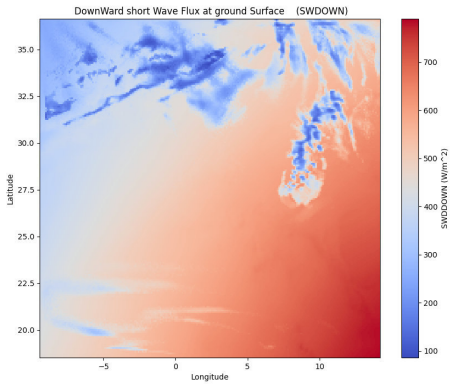


Fig. 3: Shortwave Surface Downward Diffuse Irradiance (SWDDOWN) map at 09 hour .

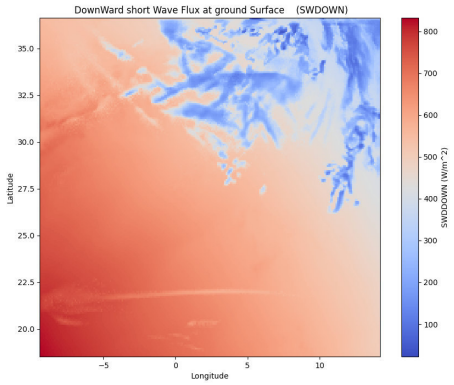


Fig. 4: Shortwave Surface Downward Diffuse Irradiance (SWDDOWN) map at 15 hour .

The map of Shortwave Downward Radiation (SWDOWN) at 09:00 and 15:00 hours illustrates the spatial distribution of solar energy received at the Earth's surface. At 09:00 Figure 3 , the SWDOWN values generally show lower intensity across most regions, as the sun is still relatively low in the sky. This results in longer solar paths and higher scattering, especially in areas with atmospheric interference like clouds or aerosols.

By 15:00 hour (Figure 4) , the map typically indicates a decrease in SWDOWN after reaching its peak around midday. Solar radiation values start to decline as the sun moves lower on the horizon, reducing the direct solar intensity. Regional variations can occur due to cloud cover, terrain, or other local atmospheric conditions, which modulate the shortwave radiation received. Comparing both times provides insight into the diurnal pattern of solar energy, which is essential for understanding the availability of solar power generation throughout the day.

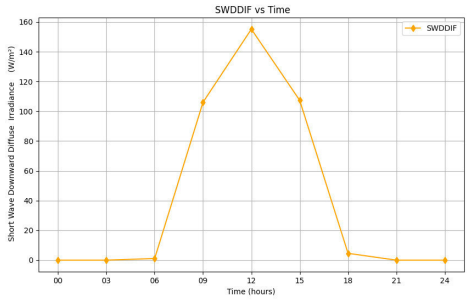


Fig. 5: Temporal variations of solar irradiance parameters simulated by the WRF model over a 24-hour period, illustrating the diurnal cycle of SWDDIF

SWDDIF refers to the diffuse component of solar radiation that reaches the Earth's surface. This is the sunlight that has been scattered by the atmosphere and clouds, arriving indirectly rather than in a straight line from the sun. In the graph (Figure 5) , the SWDDIF values increase after sunrise, peak around midday, and then decrease toward sunset, following a similar trend to SWDOWN but at a lower magnitude.SWDDIF remains significant even when direct sunlight is reduced due to clouds or atmospheric scattering. This is why it does not reach zero during the day, unlike the direct irradiance (SWDDIR), which may drop significantly under overcast conditions.

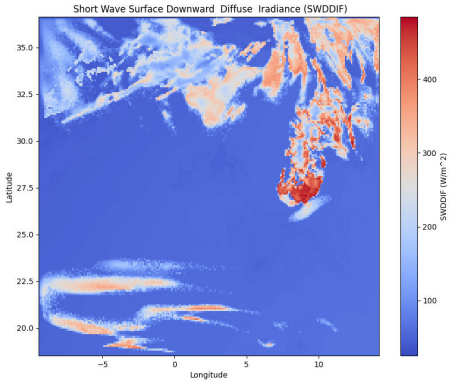


Fig. 6: Shortwave Surface Downward Diffuse Irradiance (SWDDIF) at 09 hour .

Between 09:00 and 15:00, the map of shortwave downward diffuse irradiance (SWDDIF) reveals a clear diurnal pattern influenced by the sun's position in the sky and atmospheric scattering processes.

At 09:00 hour (Figure 7) , SWDDIF values are relatively lower as the sun is still low on the horizon, causing sunlight to pass through a thicker layer of the atmosphere. This increases the scattering of solar radiation, reducing the overall diffuse irradiance reaching the surface. As the sun climbs higher, SWDDIF increases, reaching its peak around midday.

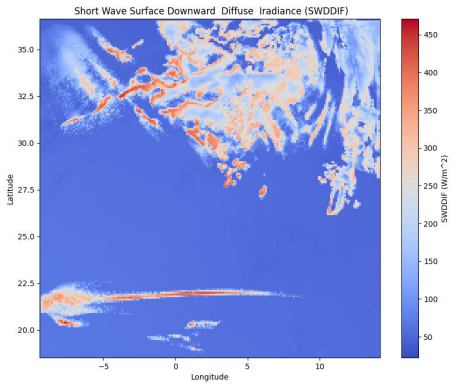


Fig. 7: Shortwave Surface Downward Diffuse Irradiance (SWDDIF) at 15 hour .

By 15:00 hour , (Figure 8) although the sun is still relatively high, the diffuse irradiance begins to decline as the solar angle decreases, resulting in more scattering and attenuation. Additionally, cloud cover and atmospheric aerosols contribute to variations in SWDDIF across different regions, with cloudier areas exhibiting higher diffuse irradiance due to increased scattering. The map highlights the gradual transition of SWDDIF through the day, with clear spatial and temporal variability driven by meteorological and atmospheric conditions.

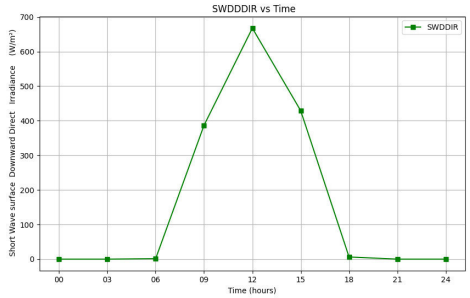


Fig. 8: Temporal variations of solar irradiance parameters simulated by the WRF model over a 24-hour period, illustrating the diurnal cycle of SWDDIR.

SWDDIR represents the direct irradiance—the component of solar radiation that travels in a straight line from the sun to the Earth's surface without being scattered by the atmosphere. In the graph (Figure 8) , SWDDIR follows a diurnal cycle, with no values during nighttime and rising sharply after sunrise, reaching a peak around noon when the sun is at its highest point. This peak occurs because the sunlight has the least atmospheric interference at midday, allowing more direct sunlight to reach the ground. After noon, the values of SWDDIR gradually decrease as the sun sets, reducing the amount of direct solar radiation received at the surface. This information is crucial for understanding solar energy availability for photovoltaic systems and other solar-dependent technologies, as direct irradiance provides the most significant contribution to solar power generation during clear sky conditions.

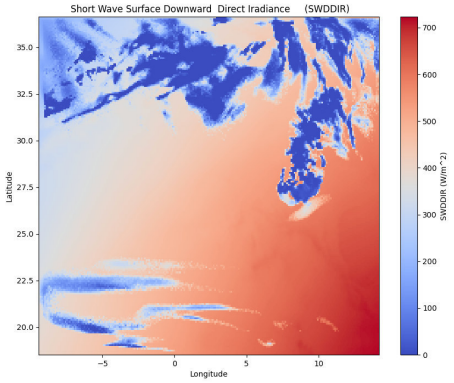


Fig. 9: Short Wave Surface Direct Irradiance (SWDDIR) at 09 hour .

In the morning at 09:00 hour (Figure 10) , the sun is relatively low, resulting in lower SWDDIR values. During this time, sunlight must travel through a longer atmospheric path, increasing the likelihood of scattering and absorption by aerosols, clouds, and other atmospheric components. Consequently, less direct solar radiation reaches the surface.

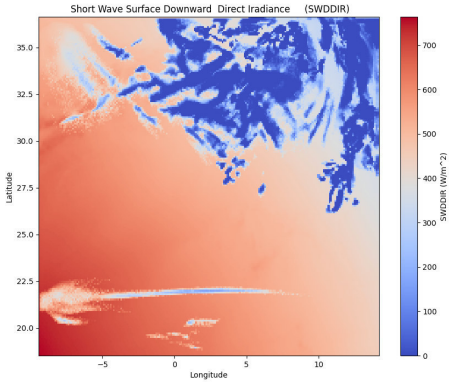


Fig. 10: Short Wave Surface Direct Irradiance (SWDDIR) at 15 hour .

By 15:00 hour (Figure 11) , as the sun rises higher in the sky, the direct solar irradiance reaching the surface increases significantly. The shorter atmospheric path reduces the scattering effect, allowing more solar energy to directly strike the surface. However, as the afternoon progresses, SWDDIR begins to decline as the sun starts its descent, with irradiance levels dropping accordingly.

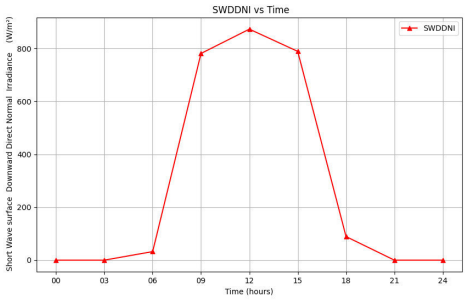


Fig. 11: Temporal variations of solar irradiance parameters simulated by the WRF model over a 24-hour period, illustrating the diurnal cycle of SWDDNI.

SWDDNI refers to the Direct Normal Irradiance (DNI), which measures the solar radiation that comes directly from the sun in a straight line and is received on a surface perpendicular to the incoming rays. This value represents the intensity of direct sunlight without any scattering by the atmosphere or clouds.

In the graph (Figure 11) , SWDDNI starts at zero during the night, increases rapidly after sunrise, and peaks around noon when the sun is highest in the sky. It then decreases again towards sunset. The high values of SWDDNI around midday indicate the strength of direct sunlight during clear-sky conditions. Direct Normal Irradiance is crucial for solar power applications, especially for systems that use solar concentrators, as they rely on direct sunlight for efficiency.

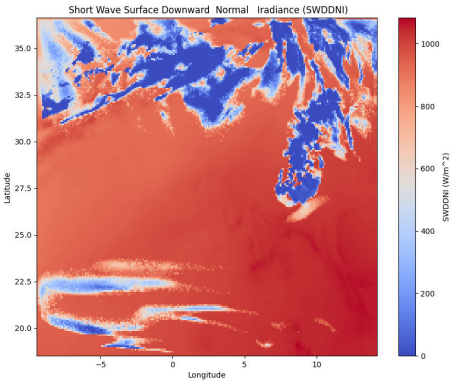


Fig. 12: Variation of Shortwave Downward Direct Irradiance (SWDDNI) at 09 hour .

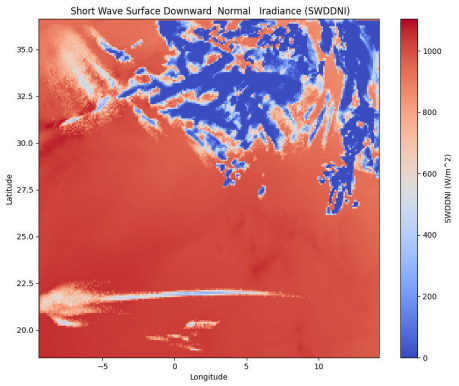


Fig. 13: Variation of Shortwave Downward Direct Irradiance (SWDDNI) at 15 hour .

Between 09:00 and 15:00 (Figure 12) and (Figure 13) , the Shortwave Downward Direct Irradiance (SWDDNI) exhibits a marked increase, reflecting the progression of solar elevation throughout the day. During the early morning hours, the irradiance levels are relatively low due to the sun being positioned closer to the horizon, leading to longer atmospheric paths and greater scattering of solar radiation. However, as the day progresses towards noon, the solar angle rises, resulting in more direct sunlight reaching the surface. This increase in SWDDNI indicates enhanced solar energy availability for potential conversion into electricity, particularly for photovoltaic systems. The spatial distribution across the region reveals variations influenced by local topography and atmospheric conditions, with certain areas experiencing higher direct irradiance values than others. By 15:00, SWDDNI generally peaks or approaches peak values, demonstrating the importance of this parameter in evaluating solar energy potential during peak sunlight hours.

VI. Conclusion

In conclusion, the analysis of solar radiation components within the Weather Research and Forecasting (WRF) model provides critical insights into the interactions between solar energy and atmospheric processes. The evaluation of downward solar radiation, both direct and diffuse, along with the assessment of the normal irradiance, highlights the model's capability to simulate the complex dynamics of solar energy transfer in the atmosphere. The findings demonstrate that accurate representation of these solar parameters is essential for improving the model's predictive skill, particularly in short-term forecasting scenarios. Moreover, understanding the variations in solar radiation throughout the day enhances our knowledge of energy distribution at the Earth's surface, which has profound implications for temperature forecasting, energy balance assessments, and climate modeling. Future work should focus on refining the parameterization of solar radiation components in the WRF model, as well as validating model outputs against high-resolution observational data, to

further enhance the reliability of weather predictions and climate simulations.

REFERENCES

[1] Kaldellis, J. K., Zafirakis, D. (2011). "The role of solar energy in the energy mix of the future." Renewable and Sustainable Energy Reviews, 15(8), 3063-3071.

[2] Szuromi, P., et al. (2007). "Solar Energy Potential: An Overview." Renewable Energy, 32(14), 2240-2252.

[3] Dincer, I., Rosen, M. A. (2011). "Thermal Energy Storage: Systems and Applications." John Wiley Sons.

[4] International Energy Agency (IEA). (2006). Energy Technology Perspectives 2006.

[5] Electric Power Research Institute (EPRI). (2003). "Photovoltaic Technology and Applications."

[6] Skamarock, W. C., et al. (2008). "A Description of the Advanced Research WRF Version 3." NCAR Technical Note.

[7] Powers, J. G., et al. (2017). "The Weather Research and Forecasting Model: A Technical Overview." Bulletin of the American Meteorological Society, 98(12), 2447-2471.

[8] Jimenez, P. A., et al. (2016). "Applications of the WRF Model for Renewable Energy Assessment." Renewable Energy, 99, 59-71.

[9] Grell, G. A., et al. (2005). "A description of the Advanced Research WRF Version 2." *NCAR Technical Note.

[10] Chuang, Y., et al. (2013). "A review of the WRF Model applications." Atmospheric Research, 122, 295-310.

[11] Wang, J., et al. (2016). "WRF Solar: A New Module for Solar Radiation Calculation." Journal of Atmospheric and Oceanic Technology , 33(1), 109-122.

[12] Hirsch, A., et al. (2013). "Assessing the Accuracy of WRF Solar Radiation Forecasts." Journal of Solar Energy Engineering , 135(3), 031006.

[13] Pérez, C., et al. (2015). "Solar Radiation Forecasting with WRF." Renewable Energy , 78, 203-211.

Enhanced Defect Detection in Photovoltaic Systems Using YOLOv10: A Deep Learning Approach

Mohamed Achit
Departement Of Electrical Engineerig
Laboratory of Materials and Sustainable Development.
University of bouira
bouira, algeria
m.achit@univ-bouira.dz

Nacera Yassa
Departement Of Electrical Engineerig
Laboratory of Materials and Sustainable Development.
University of bouira
bouira, algeria
n.yassa@univ-bouira.dz

Ahcene Bouzida
Departement Of Electrical Engineerig
Laboratory of Materials and Sustainable Development.
University of bouira
bouira, algeria
a.bouzida@univ-bouira.dz

Ali Mazari
Departement Of Mecanical Engineerig
Laboratory of The Applied And Automation And Industial Diagnostic
University Of Djelfa
Djelfa, Algeria
ali.mazari@univ-djelfa.dz

Okba Fergani
Departement Of Electrical Engineerig
Laboratory of Identification, Command, Control and Communication
university of biskra
Biskra, Algeria
okba.fergani@univ-biskra.dz

Billel Naceri
Departement Of Electrical Engineerig
Laboratory of Identification, Command, Control and Communication
university of biskra
Biskra, Algeria
billel.naceri@univ-biskra.dz

***Abstract*— The rising demand for renewable energy solutions has propelled advancements in photovoltaic (PV) technology, particularly in the area of solar cell quality control. Defect detection within solar cells is critical for ensuring their performance and longevity. This paper presents a novel approach to PV defect detection utilizing the YOLOv10n model, a lightweight version of YOLOv10, specifically optimized for mobile devices and environments with limited computational resources. Building upon previous YOLO iterations, YOLOv10n employs a non-maximum suppression-free post-processing strategy that enhances inference speed without compromising accuracy. Key innovations in this model include dual-label assignment mechanisms and the use of partial self-attention layers. The model was trained on the PVEL-AD dataset, containing 8 defect classes derived from over 3,800 high-resolution Electroluminescence (EL) images. Advanced data augmentation techniques, including HSV adjustment, random erasing, and mosaic augmentation, were applied to improve model robustness. Experimental results show that YOLOv10n achieves a mean Average Precision (mAP@50) of 0.89 and demonstrates superior performance across various defect types, such as short circuits and black cores. The model's scalability and real-time processing capabilities make it an ideal solution for mobile-based industrial-level defect detection in photovoltaic systems.**

Keywords—Solar cell defects, object detection, YOLOv10, Electroluminescence imaging, photovoltaic systems, deep learning

I. INTRODUCTION

The increasing demand for renewable energy has driven significant progress in photovoltaic (PV) technology. Solar cells, which are vital components of PV systems, require thorough quality control to ensure both efficiency and durability [1, 2]. Detecting defects in solar cells is a crucial aspect of the manufacturing process, as such defects can drastically affect the performance and reliability of solar panels [3]. Traditional inspection techniques are often manual and susceptible to errors [4], highlighting the necessity for automated, precise, and efficient detection methods [5].

Deep learning-based object detection models have shown considerable potential in various applications, including defect detection in industrial environments. These models can identify and localize objects within an image, which is essential for automating quality control processes. Notably, the You Only Look Once (YOLO) framework, introduced by Redmon et al. in 2016 [6], revolutionized object detection with its innovative methodology and remarkable performance. The original YOLOv1 model [7] brought about a paradigm shift by employing a single neural network to predict bounding boxes and class probabilities directly from full images in a single pass, contrasting with traditional two-stage approaches [8] that first generate region proposals and then classify those regions. This consolidated approach in YOLOv1 significantly improved processing speeds [9], making it highly suitable for real-time applications.

Subsequent iterations of YOLO have introduced several enhancements. YOLOv2 and YOLOv3 improved both detection accuracy and speed by adding features such as batch normalization, anchor boxes, and multi-scale predictions [10, 11]. YOLOv4 and YOLOv5 further optimized the architecture by refining backbone networks and incorporating advanced data augmentation techniques [12]. YOLOv6 and YOLOv7 focused on streamlining the network structure and reducing computational costs, making these models even more appropriate for real-time use. The latest versions, YOLOv8 and YOLOv9, have pushed performance boundaries further with advanced attention mechanisms and additional architectural improvements [13].

Electroluminescence (EL) imaging has emerged as a crucial technique in solar cell inspection [14-16], producing high-resolution images that reveal defects such as cracks, dislocations, and other anomalies. The literature has increasingly adopted EL imaging alongside various deep learning algorithms for defect detection. Early approaches employed convolutional neural networks (CNNs) to classify and localize defects in EL images. For instance, transfer learning using pre-trained CNNs like VGG and ResNet has been applied to improve defect detection accuracy [17]. More

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

recent studies have explored advanced architectures such as Faster R-CNN, SSD, and RetinaNet, which offer enhanced accuracy and speed for detecting and classifying defects in EL images [18]. Despite these advances, challenges remain in achieving real-time detection and addressing diverse defect types challenges our proposed YOLOv10 model aims to tackle.

Recent research has demonstrated the efficacy of deep learning techniques in detecting solar cell defects. For example, [19] employed a combination of CNNs and data augmentation methods to improve the detection of micro-cracks in EL images, achieving notable improvements over traditional approaches. Similarly, [20] used a hybrid model that combined Faster R-CNN with an attention mechanism to enhance the detection of fine-grained defects in solar cells. Despite these advances, many models still struggle to meet real-time processing demands while maintaining high accuracy across diverse defect types, underscoring the need for further innovation. Building on these advances, recent work has integrated more sophisticated techniques to address these challenges. For example, [21] introduced a multi-scale feature fusion approach in their deep learning model to better capture defects of varying sizes and shapes in EL images, resulting in improved detection rates for smaller, less prominent defects. Additionally, [22, 23] developed a novel ensemble model that combines the strengths of multiple neural network architectures, producing a robust system capable of detecting a broader range of defects with greater precision. These studies highlight ongoing efforts to improve model performance, but persistent challenges remain, particularly in achieving real-time processing and generalizing across diverse defect scenarios.

This work contributes to the field by presenting the YOLOv10n model, a lightweight and highly optimized version of YOLOv10 designed specifically for mobile devices and environments with limited computational resources. Building upon previous YOLO iterations, YOLOv10n introduces a non-maximum suppression-free post-processing strategy, dual-label assignment mechanisms, and partial self-attention layers to enhance inference speed without compromising accuracy. The model was trained on the PVEL-AD dataset, containing eight defect classes from over 3,800 high-resolution EL images, and was equipped with advanced data augmentation techniques like HSV adjustment, random erasing, and mosaic augmentation to improve robustness.

This paper is structured as follows: Section 2 presents the Methodology, detailing the architecture of the YOLOv10 model and the dataset used for training and evaluation. The augmentation techniques and their impacts on model performance are discussed in this section as well. Section 3 elaborates on the Experimental Results, providing an analysis of the model's performance on defect detection, followed by the validation results. Finally, Section 4 concludes the paper by summarizing the findings and suggesting directions for future research.

II. METHODOLOGY

A. *Yolov10 Model Architecture*

YOLOv10 represents a significant leap in object detection by setting new standards in inference speed, mainly by removing the need for non-maximum suppression (NMS) during post-processing. Instead, it employs an innovative NMS-free strategy that integrates dual-label assignment, combining detailed supervision through the one-to-many branch with fast inference via the one-to-one branch. To ensure alignment between these branches, the model adopts a standardized matching method, optimizing both training and inference processes effectively [24, 25].

Table 1 illustrates the YOLOv10 architecture, outlining the layer configurations, the filter types and sizes, and the output dimensions at each stage. This architecture leverages cutting-edge features like partial self-attention (PSA) and spatially constrained (SC) layers, which enhance the model's precision and efficiency. YOLOv10 also incorporates advanced feature integration techniques such as depthwise and pointwise convolutions. Additionally, a sophisticated dual-label assignment mechanism optimizes both training and inference phases, highlighting YOLOv10's superior capabilities in handling complex object detection tasks with enhanced speed and accuracy.

The model's architecture is designed to boost efficiency while maintaining high performance. A key element of this design is a lightweight classification head, which addresses the bottlenecks in regression tasks. The use of pointwise convolution increases the channel dimensions, while depthwise convolution, which reduces the resolution, has been shown to enhance average precision by 0.7%. Moreover, integrating a depthwise layer within a compact inverted block further improves precision by 0.3%. These enhancements are combined with a rank-guided block design, offering a more streamlined structure.

The architecture's precision-focused design suggests that large-kernel convolutions are especially beneficial for smaller models, as they increase both the size and receptive field. Additionally, the use of partial self-attention mechanisms effectively balances global modeling with reduced computational demands, lowering the complexity that typically comes with full self-attention mechanisms.

TABLE I. YOLOv10 ARCHITECTURE.

Layer	Filters	Size	Repeat	Output Size
Image	-	-	-	640 × 640
Conv	16	3 × 3/2	1	320 × 320
Conv	32	3 × 3/2	1	160 × 160
C2f	32	1 × 1/1	1	160 × 160
Conv	64	3 × 3/2	1	80 × 80
C2f	64	1 × 1/1	2	80 × 80
SCDown	128	3 × 3/2	1	40 × 40
C2f	128	1 × 1/1	2	40 × 40
SCDown	256	3 × 3/2	1	20 × 20
C2f	256	1 × 1/1	1	20 × 20
SPPF	256	5 × 5/1	1	20 × 20
PSA	256	-	1	20 × 20
Upsample	-	2×	1	40 × 40
Concat	-	-	1	40 × 40
C2f	128	1 × 1/1	1	40 × 40
Upsample	-	2×	1	80 × 80
Concat	-	-	1	80 × 80
C2f	64	1 × 1/1	1	80 × 80
Conv	64	3 × 3/2	1	40 × 40
Concat	-	-	1	40 × 40
C2f	128	1 × 1/1	1	40 × 40
SCDown	128	3 × 3/2	1	20 × 20
Concat	-	-	1	20 × 20
C2fCIB	256	1 × 1/1	1	20 × 20
Detect	64, 128, 256	-	1	80 × 80, 40 × 40, 20 × 20

B. Dataset

The PVEL-AD dataset [26], developed by Hebei University of Technology and Beijing University of Aeronautics and Astronautics, was utilized for training and evaluating the RT-DETR model. This dataset consists of 3,800 images representing 12 defect types and includes one non-defective class. However, due to the infrequency of certain defects, the number of classes was reduced to eight, focusing on the most prevalent defect types: Black Core, Crack, Finger, Horizontal Dislocation, Short Circuit, Star Crack, Thick Line, and Vertical Dislocation. To enhance model generalization, the less common defect types such as "Scratch," "Fragment," "Corner," and "Printing Error" were excluded. Figure 1 showcases the morphology of these eight selected defect types. For consistency, all images were resized to 640x640 pixels, and the dataset was divided into training (70%), validation (20%), and test (10%) subsets.

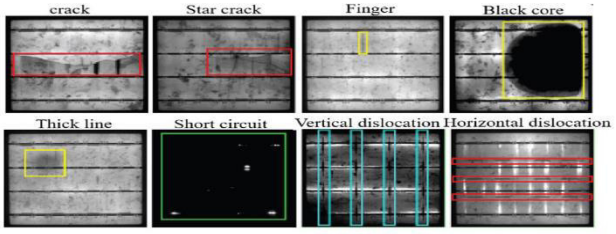


Fig. 1. The structure of 8 different types of abnormal defects in the PVEL-AD dataset.

C. Data Augmentation

IData augmentation plays a pivotal role in deep learning models, particularly in computer vision tasks such as defect detection in photovoltaic (PV) systems. Augmentation techniques artificially expand the dataset by generating new images through transformations applied to the original data, thereby enabling the model to generalize more effectively to unseen scenarios. This process mitigates overfitting by exposing the model to a broader range of data variations. Specifically, data augmentation helps address the problem of underrepresented defect types, ensuring balanced learning across all classes and reducing bias. By simulating various real-world conditions, augmentation improves model robustness, leading to enhanced accuracy.

HSV-Hue Augmentation (hsv_h: 0.015): This technique shifts the hue of the image by 1.5%, simulating lighting variations by altering the image's color spectrum, helping the model generalize better to different lighting environments.

HSV-Saturation Augmentation (hsv_s: 0.7): This increases or decreases the saturation by 70%, adjusting the intensity of colors. By simulating varying lighting environments, this helps the model handle different exposure levels and lighting conditions.

HSV-Value Augmentation (hsv_v: 0.4): This adjusts the brightness by 40%, allowing the model to train on images that range from dark to bright, further enhancing its robustness to different lighting and exposure scenarios.

Translation (translate: 0.1): This shifts the image horizontally or vertically by 10% of its size. Translation augmentation aids the model in detecting objects that are not centered in the image, improving its detection capabilities across the entire frame.

Scale (scale: 0.5): This technique scales the image by a factor of up to 50%. It helps the model handle objects of different sizes, ensuring that defects of varying scales are detected accurately.

Horizontal Flip (fliplr: 0.5): This randomly flips the image horizontally with a 50% probability. This augmentation helps the model detect symmetrical or mirrored defects, improving its generalization to defects that may appear in reverse orientations.

Mosaic Augmentation (mosaic: 1.0): This combines four images into one, creating a complex image with multiple objects. Mosaic augmentation significantly enhances the model's ability to detect defects in scenarios where multiple objects or defects are present, improving overall detection efficiency.

Random Erasing (erasing: 0.4): This randomly erases parts of the image with a probability of 40%, simulating occlusions or missing data. It helps the model become more robust by allowing it to learn from incomplete or partially obscured images.

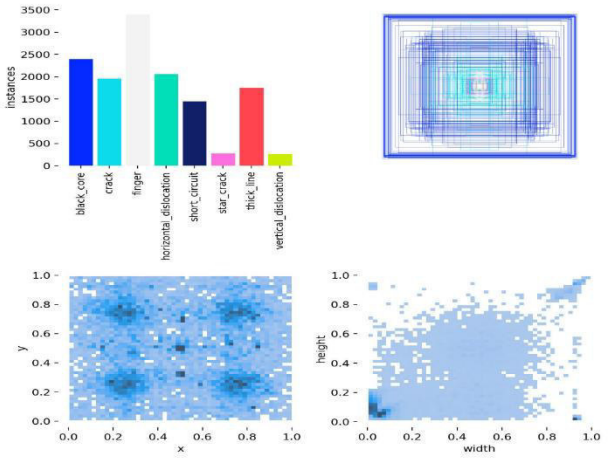


Fig. 2. The structure of 8 different types of abnormal defects in the PVEL-AD dataset.

Figure 2 highlights the distribution of defect instances across eight common defect types in the PVEL-AD dataset, as well as the spatial representation of anchor boxes used in the YOLOv10 model. The top left graph illustrates the number of instances for each defect type, with Black Core, Crack, and Thick Line being the most frequent. The top right image shows the layout of anchor boxes, providing insight into how bounding boxes are generated during training. The bottom row presents heatmaps of object positions and sizes in the dataset, demonstrating the diversity in defect scales and positions. These visualizations are integral to understanding the distribution of data and the model's anchoring mechanism, which enhances detection across varying defect types and scales.

III. Experimental Results

The experimental results for the YOLOv10 model, as shown in Figure 3, highlight the model's performance metrics throughout the training and validation phases. The model was trained using the PyTorch framework on a high-performance system equipped with NVIDIA GPUs, employing advanced deep learning techniques to optimize both speed and accuracy. Over 200 epochs, hyperparameters were fine-tuned to achieve optimal performance. The training process utilized the stochastic gradient descent (SGD) optimizer with an initial learning rate of 0.01 and a momentum value of 0.9. A learning rate decay strategy, combined with regularization techniques, helped the model avoid overfitting while maintaining high generalizability. Specifically, biases were subjected to L2 regularization with a decay factor of 0.0005, while some weights were exempt from decay (decay = 0.0), further improving model accuracy.

The batch size was set to 16, striking a balance between memory limitations and computational efficiency. Throughout the training process, the model demonstrated steady convergence, with both training and validation loss functions showing consistent improvement. The training losses (box loss, classification loss, and distribution focal loss) decreased sharply during the early epochs, stabilizing as training progressed, which indicates effective feature learning. The validation losses followed a similar pattern, further validating the model's generalization capabilities.

Precision and recall metrics improved steadily across the epochs, with both metrics reaching high levels towards the later stages of training. Notably, the mean Average Precision (mAP) at 50% IoU (mAP50) achieved a value of 0.89, demonstrating the model's strong detection performance. Additionally, the mAP across multiple IoU thresholds (mAP50-95) showed a positive trend, reflecting the model's ability to accurately detect objects at various scales and occlusions.

These results underscore the YOLOv10 model's robustness and efficiency, particularly in real-time object detection tasks, confirming that the model is well-suited for high-performance applications requiring precise and rapid detections.

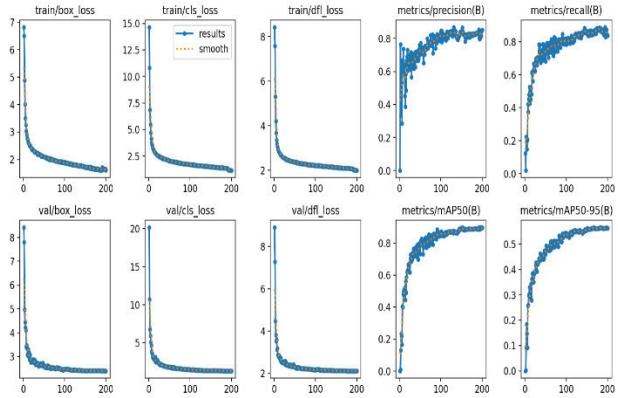


Fig. 3. YOLOv10 model's performance.

Figure 4, showcases the YOLOv10 validation results using the best-performing model weights (best.pt) across various defect classes. The model was evaluated on 685 images and 1,159 instances, achieving an overall mean Average Precision (mAP@50) of 0.892 and mAP@50-95 of 0.565. The figure details the precision (P), recall (R), and mAP metrics for each defect class, demonstrating the

model's strong performance across diverse defect types, particularly excelling in detecting short circuits (mAP50 = 0.992) and black cores (mAP50 = 0.981). This further validates the YOLOv10 model's effectiveness in real-world defect detection applications.

```
Validating runs/detect/train/weights/best.pt...
Ultralytics YOLOv8.2.91 🚀 Python-3.10.13 torch-2.1.2 CUDA:0 (Tesla T4, 15095MiB)
YOLOv10n summary (fused): 285 layers, 2,697,536 parameters, 0 gradients, 8.2 GFLOPs
```

Class	Images	Instances	Box(P	R	mAP50	mAP50-95)
<00:00, 3.41it/s]						
all	685	1159	0.825	0.883	0.892	0.565
black_core	146	155	0.974	0.952	0.981	0.919
crack	192	246	0.636	0.748	0.763	0.48
finger	178	306	0.77	0.866	0.891	0.511
horizontal_dislocation	52	156	0.937	0.968	0.992	0.398
short_circuit	73	73	0.98	0.986	0.992	0.983
star_crack	22	24	0.61	0.833	0.723	0.423
thick_line	136	171	0.773	0.901	0.902	0.487
vertical_dislocation	7	28	0.919	0.811	0.889	0.318

```
Speed: 0.2ms preprocess, 3.0ms inference, 0.0ms loss, 0.2ms postprocess per image
```

Fig. 4. Performance Metrics of YOLOv10 Model on Defect Detection.

Figure 5, the precision-confidence curve, shows how precision improves with higher confidence thresholds across the different defect classes. As expected, precision increases as the confidence threshold rises, indicating the model's improved accuracy when making confident predictions. Defects like black_core and short_circuit demonstrate strong performance, achieving near-perfect precision at higher confidence levels. In contrast, classes like finger and horizontal_dislocation exhibit more fluctuations in precision at lower confidence levels, stabilizing only at higher confidence thresholds. The star_crack class, however, shows significant variability, with lower precision across the confidence spectrum, particularly at lower thresholds, suggesting it is more challenging for the model to detect accurately.

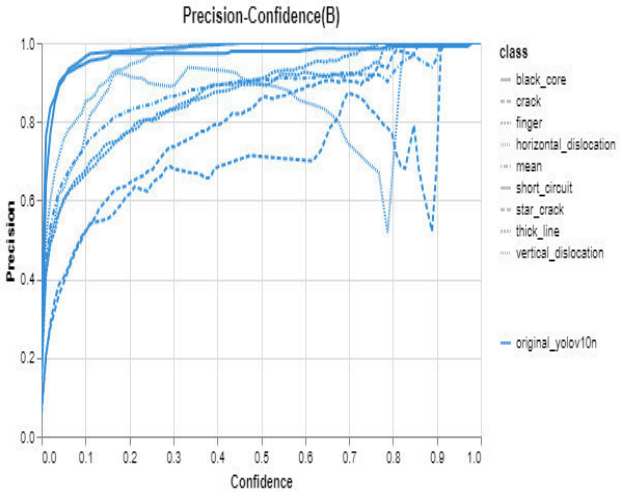


Fig. 5. Precision-confidence curve of YOLOv10.

In Figure 6, the recall-confidence curve further details the model's performance, showing how recall decreases as confidence levels rise. The model demonstrates strong recall across most defect types until the confidence threshold reaches approximately 0.9, after which recall drops sharply. Defects like black_core and short_circuit maintain high recall across a wide range of confidence thresholds, suggesting strong detection performance. However, defects like finger and horizontal_dislocation experience a drop in recall at higher confidence levels. The star_crack class shows significant variability, with recall diminishing rapidly, indicating that the model struggles to consistently detect this defect. Overall, the curves suggest that while the model maintains high precision and recall for many classes at moderate confidence levels, it may face challenges with more complex defect types at higher confidence thresholds.

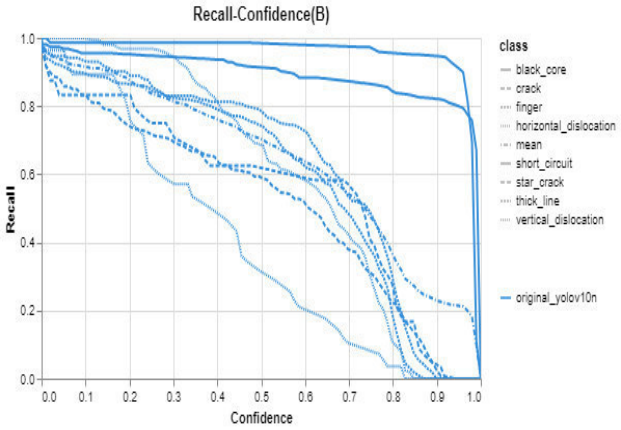


Fig. 6. recall-confidence curve of YOLOv10.

Figure 7 combines the insights from both precision and recall into an F1-confidence curve, which balances the two metrics. The F1 score generally increases with confidence until a point where it starts to plateau or decline. Defects such as black_core, short_circuit, and crack demonstrate strong overall performance, maintaining high F1 scores across a broad range of confidence values. However, star_crack exhibits the most variability in its F1 score, suggesting that this class remains challenging for the model to detect with both high precision and recall.

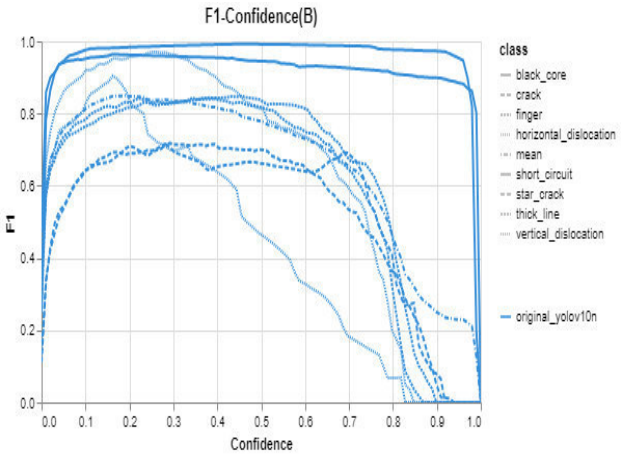


Fig. 7. F1-confidence curve of YOLOv10.

Figure 8 shows the normalized confusion matrix for the YOLOv10n model, highlighting its classification performance across various defect classes. The diagonal values indicate the percentage of correct predictions for each class, while off-diagonal values represent misclassifications. For instance, the model accurately predicts 94% of *black_core* defects, 85% of *finger* defects, and 99% of *short_circuit* defects. However, it struggles with classes like *star_crack* and *vertical_dislocation*, where only 2% and 64% of the instances were classified correctly, respectively. Misclassifications are evident in cases such as *crack* defects, where 37% are confused with *thick_line*, and

finger defects, which are misclassified as *crack* 38% of the time.

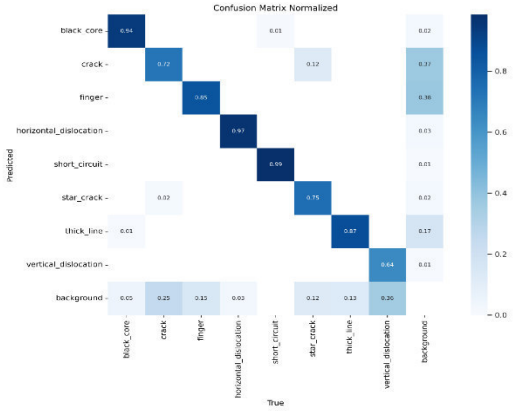


Fig. 8. confusion matrix of YOLOv10.

IV. Conclusion

IThe YOLOv10n model presents significant advancements in the detection of defects in photovoltaic cells, particularly through its lightweight architecture designed for mobile devices and environments with low computational power. The non-maximum suppression-free architecture and use of partial self-attention mechanisms contribute to its enhanced speed and efficiency. The model achieved excellent performance metrics, including a mean Average Precision of 0.89, across a diverse set of defect types. The integration of advanced data augmentation techniques further improved model generalization, while the dual-label assignment optimized both training and inference processes. The experimental results highlight YOLOv10n's suitability for real-time applications, making it an ideal candidate for mobile and embedded systems in industrial settings. Future work will focus on refining the architecture to improve detection for more subtle defects, such as star cracks, and further optimizing the model for various low-resource platforms. This research provides a practical, scalable solution for enhancing the reliability of photovoltaic systems through efficient, mobile-friendly defect detection.

References

[1] J. Kim, M. Rabelo, S. P. Padi, H. Yousuf, E.-C. Cho, and J. Yi, "A review of the degradation of photovoltaic modules for life expectancy," *Energies,* vol. 14, no. 14, p. 4278, 2021.

[2] A. Abubakar, C. F. M. Almeida, and M. Gemignani, "Review of artificial intelligence-based failure detection and diagnosis methods for solar photovoltaic systems," *Machines,* vol. 9, no. 12, p. 328, 2021.

[3] S. Jumaboev, D. Jurakuziev, and M. Lee, "Photovoltaics plant fault detection using deep learning techniques," *Remote Sensing,* vol. 14, no. 15, p. 3728, 2022.

[4] M.-H. Hwang, Y.-G. Kim, H.-S. Lee, Y.-D. Kim, and H.-R. Cha, "A study on the improvement of efficiency by detection solar module faults in deteriorated photovoltaic power plants," *Applied Sciences,* vol. 11, no. 2, p. 727, 2021.

[5] M. Hussain and R. Khanam, "In-depth review of yolov1 to yolov10 variants for enhanced photovoltaic defect detection," in *Solar*, 2024, vol. 4, no. 3: MDPI, pp. 351-386.

[6] J. Redmon, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[7] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Machine Learning and Knowledge Extraction,* vol. 5, no. 4, pp. 1680-1716, 2023.

[8] M. Hussain, "Yolov1 to v8: Unveiling each variant–a comprehensive review of yolo," *IEEE Access,* vol. 12, pp. 42816-42833, 2024.

[9] T. Ahmad, Y. Ma, M. Yahya, B. Ahmad, S. Nazir, and A. u. Haq, "Object detection through modified YOLO neural network," *Scientific Programming,* vol. 2020, no. 1, p. 8403262, 2020.

[10] J. Redmon, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767,* 2018.

[11] R. Sujee, D. Shanthosh, and L. Sudharsun, "Fabric defect detection using YOLOv2 and YOLO v3 tiny," in *Computational Intelligence in Data Science: Third IFIP TC 12 International Conference, ICCIDS 2020, Chennai, India, February 20–22, 2020, Revised Selected Papers 3*, 2020: Springer, pp. 196-204.

[12] U. Nepal and H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs," *Sensors,* vol. 22, no. 2, p. 464, 2022.

[13] R. Sapkota *et al.*, "Comprehensive Performance Evaluation of YOLOv10, YOLOv9 and YOLOv8 on Detecting and Counting Fruitlet in Complex Orchard Environments," *arXiv preprint arXiv:2407.12040,* 2024.

[14] M. Dhimish and Y. Hu, "Rapid testing on the effect of cracks on solar cells output power performance and thermal operation," *Scientific Reports,* vol. 12, no. 1, p. 12168, 2022.

[15] M. Dhimish and P. Mather, "Ultrafast high-resolution solar cell cracks detection process," *IEEE Transactions on Industrial Informatics,* vol. 16, no. 7, pp. 4769-4777, 2019.

[16] M. Rai, L. H. Wong, and L. Etgar, "Effect of perovskite thickness on electroluminescence and solar cell conversion efficiency," *The Journal of Physical Chemistry Letters,* vol. 11, no. 19, pp. 8189-8194, 2020.

[17] A. Bartler, L. Mauch, B. Yang, M. Reuter, and L. Stoicescu, "Automated detection of solar cell defects with deep learning," in *2018 26th European signal processing conference (EUSIPCO)*, 2018: IEEE, pp. 2035-2039.

[18] J. Ou *et al.*, "Infrared image target detection of substation electrical equipment using an improved faster R-CNN," *IEEE Transactions on Power Delivery,* vol. 38, no. 1, pp. 387-396, 2022.

[19] S. Hassan and M. Dhimish, "A Survey of CNN-Based Approaches for Crack Detection in Solar PV Modules: Current Trends and Future Directions," in *Solar*, 2023, vol. 3, no. 4: MDPI, pp. 663-683.

[20] A. Chen, X. Li, H. Jing, C. Hong, and M. Li, "Anomaly detection algorithm for photovoltaic cells based on lightweight Multi-Channel spatial attention mechanism," *Energies,* vol. 16, no. 4, p. 1619, 2023.

[21] J. Yang, S. Li, Z. Wang, H. Dong, J. Wang, and S. Tang, "Using deep learning to detect defects in manufacturing: a comprehensive survey and current challenges," *Materials,* vol. 13, no. 24, p. 5755, 2020.

[22] P. Monkam *et al.*, "Ensemble learning of multiple-view 3D-CNNs model for micro-nodules identification in CT images," *IEEE Access,* vol. 7, pp. 5564-5576, 2018.

[23] B. Zhang *et al.*, "Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images," *IEEE Access,* vol. 7, pp. 110358-110371, 2019.

[24] M. A. R. Alif and M. Hussain, "YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in the agricultural domain," *arXiv preprint arXiv:2406.10139,* 2024.

[25] A. Wang *et al.*, "Yolov10: Real-time end-to-end object detection," *arXiv preprint arXiv:2405.14458,* 2024.

[26] B. Su, Z. Zhou, and H. Chen, "PVEL-AD: A large-scale open-world dataset for photovoltaic cell anomaly detection," *IEEE Transactions on Industrial Informatics,* vol. 19, no. 1, pp. 404-413, 2022.

Environmental and Economic Impact of Photovoltaic Power Plant emplacement in the Electricity Network

a.Abdelatif GADOUM (gadoum.abdelatif@univ-ouargla.dz gdabdelatif@gmail.com)
b. Naima BENBOUZA (aintoutan@yahoo.fr)
b.Tarek SAHRAOUI (tarek.sahrawi030@gmail.com)

aUniversit¶ Kasdi Merbah Ouargla, Laboratoire LRPPS Facult¶ de Math¶matiques et des Sciences de la Mati¶re, Ouargla 30000, Algeria
bUniversit¶ Kasdi Merbah Ouargla, Facult¶ de Sciences Appliqu¶es, D¶partement G¶nie Electrique Ouargla Route de Ghardaia, BP.511,30000, Algeria

ABSTRACT

The present study focuses on the optimization of economic and environmental dispatching in an electricity network comprising two thermal power stations and one hydroelectric station. Economic and environmental dispatching involves the optimization of the production cost function and the function of toxic gas emissions in electrical networks. The research investigates two scenarios: one involving conventional production in the electricity network, and the other incorporating photovoltaic plant. The optimization process is conducted using the genetic algorithm method by Matlab environment. The results demonstrate the significant impact of integrating photovoltaic production, leading to reduce electricity production costs and minimize toxic emissions. Key terms relevant to this study include centralized and decentralized production, economic and environmental dispatching, optimization, and genetic algorithms.

Keywords-

Economic dispatching, environmental dispatching, optimization, genetic algorithms, Photovoltaic plant, Renewable energy, Genetic algorithm.

I. INTRODUCTION

The renewable energy technology has becoming the main concern of researchers and governments [1-10]. A simple example of the increasing integration of renewable energies into energy networks; the proliferation of solar panels on rooftops reflects a notable shift towards sustainable and economically viable solutions. This accelerated transition lead as also to understand the importance of controlling all associated processes, and going deeper in studies in the same time in order to benefit of the development taking place. Economic and environmental dispatching is one of interesting processes that we need to give importance.

Economic dispatch (ED) in electric power generation aims to optimize the allocation of power generation from available sources to satisfy demand, minimizing the overall cost of operation. This involves determining the optimal output levels for each committed generating unit while ensuring that various constraints, including operational limitations of individual units and system-wide constraints are met [11]. While the main goal of environmental dispatch (ED) in electric power generation is to reduce greenhouse gas emissions, air pollutants, and other environmental burdens associated with electricity production, with meeting demand and operational constraints.

The incorporation of renewable energies has redefined the conventional Economic Load Dispatch issue into a bi-objective Economic Emission Dispatch challenge, with the aim of simultaneously decreasing costs and emissions [12]. This explains the intense studies interested to this important branch of renewable energy technology [13-25].

The objective of reference [26] is to analyze the influence of power systems on operating costs and emissions, specifically focusing on the integration of wind energy generators. Trade-offs between economical environmental impacts of solar-wind-thermal power generation system by using multi-objective model is proposed in [27]. In [28], a stochastic dynamic economic dispatch for low-carbon power dispatching problem was proposed. Using non-dominated sorting genetic algorithm-II in [29], an economic environmental dispatch problem of solar-wind-hydro-thermal generation system with battery energy storage was solved. An economic environmental dispatch problem including solar, wind and small-hydro power is solved in Reference [30]. Ehab E. Elattar [31] proposed the modified harmony search (MHS) algorithm to solve the combined economic emission dispatch (CEED) problem of the microgrid taking into account the solar and wind power cost functions. Reference [32] introduces a framework for addressing the

economic environmental dispatch problem within a multi-objective problem context.

The objective of reference [33] is to utilize particle swarm optimization (PSO) algorithm to solve single objective optimization problems and multi-objective PSO algorithm for multi-objective optimization problems. The effectiveness of this approach is demonstrated through its application to a six-generator power system.

In the paper of Jie Cao et al [34], a novel low-carbon economic dispatch method, utilizing the multi-objective chaotic artificial hummingbird algorithm, is introduced to investigate Regional Integrated Energy Systems (RIES), emphasizing the connection of diverse energy supply subsystems to address varied user needs and enhance operational efficiency.

The study of [35] investigates the ability of bio-inspired metaheuristic approach named African Vulture Optimization Algorithm (AVOA) compared to other established and advanced metaheuristic algorithms for solving Economic-Emission Load Dispatch (EELD) problem of power systems under various practical constraints.

In our work, the aim to analyze the impact of the placement of a Photovoltaic (PV) Plant within the IEEE 9-bus network on cost and the environment, we employ a genetic algorithm. By optimizing both cost and environmental impact simultaneously, we can find effective compromises between these dimensions.

II. PROBLEM FORMULATION

II.1 ECONOMIC FUNCTION

A. Cost function of the thermal production

The production cost is determined from pre-defined curves experimentally derived for each unit as a function of the power it generates. The function associated with these curves is a polynomial of degree 'n'. In practice, it is most often presented in the form of a second-degree polynomial [36].

$$F_{thi}(P_{Gi}) = a_i P_{Gi}^2 + b_i P_{Gi} + c_i ...(\$/h) \text{ With } i = 1, 2, ..., n_g$$

$F_{thi}(P_{Gi})$ is the function of the fuel cost, P_{Gi} is the power generated, a_i, b_i and c_i are the cost coefficients specific to each unit of electrical energy production and n_g the total number of generators.

B. Cost function of the of the photovoltaic power plant production

The cost function of a photovoltaic power plant can be represented by the following linear function:

$$P_{pvj} = d_j P_{vj} ...(\$/h) \quad j = 1, 2, ..., n$$

(P_{pvj}) is the function of the cost of the photovoltaic power plant. P_{pvj} is the power generated by the photovoltaic plant at node j. d_j is the cost coefficient specific to the photovoltaic plant, and n is the total number of photovoltaic power plant. The cost of photovoltaic production can be obtained based on the following equations [37]:

$$C_{pv} = a l_p \sum (P_{pvj}) + G^E \sum (P_{pvj})$$

$$a = \frac{r}{1 - (1 + r)^{-N}}$$

Where P_{PV} is the production of the solar energy source, a is the annualization coefficient, r is the interest rate (0.09), N is the lifespan of the investment (20 years), l_p is the investment cost per unit of installed power and G^E is the Operation and Maintenance Cost. l_p and are taken as \$5000/kW and 1.6 cents/kW, respectively for solar energy sources. $F_{pvj}(P_{pvj}) = 547.7483 * P_{PV}$

C. The minimization of the thermal and photovoltaic production total cost functions:

The minimization of the function of the total cost of thermal production and photovoltaic production is presented as follows:

$$Min F = \sum_{i=1}^{ng} F_{thi}(P_{Gi}) + \sum_{j=1}^{nf} F_{pvj}(P_{pvj})$$

Under the constraints:

$$\sum_{i=1}^{ng} P_{Gi} + \sum_{j=1}^{nf} P_{pvj} - P_d - P_L = 0$$

$$P_{Gi} \text{ ¤ } i \text{ ¤ } P_{Gimax} \qquad i = 1, 2, ..., n_g$$

$$0 \text{ ¤ } P_{pvj} \text{ ¤ } P_{pvjmax} \qquad j = 1, 2, ..., n$$

偁: Total active load power.

偁: Total active losses in the network.

偁鍂Minimum active power of the generator.

偁j: Maximum active power of the generator

II.2 ENVIRONMENTAL FUNCTION:

The function of gas emissions from production plants can be described as follows:

$$E_i(P_{Gi}) = a_i P_{Gi}^2 + b_i P_{Gi} + d_i [Ton/h]$$

The environmental study consists of minimizing the function of emissions:

$$Min E = \sum_{i=1}^{ng} E_i(P_{Gi})$$

II.3 MONO-OBJECTIVE BI-OBJECTIVE AND OPTIMIZATION:

The economic-environmental study therefore consists of seeking the simultaneous minimization of the two functions described by the same object variables. The optimization problem represents a bi-objective or bi- criteria problem. The main difficulty of such an optimization problem is linked to the presence of conflicts between the two functions. For this, the bi-objective optimization problem can be transformed into a mono- objective optimization problem, introducing a price penalty factor Fp which represents the ratio between the maximum fuel cost and the maximum emissions of the corresponding generator [38].

$$F_{pi} = \frac{F_{thi}(P_{Gimax})}{E_i(P_{Gimax})} \ldots(\$/h)$$

The following steps are used to find the price penalty factor for a specific load demand.

1- Find the ratio between the maximum fuel cost and the maximum Emission $F_{thi}(P_{Gimax})$ of each generator $E_i(P_{Gimax})$

2- Arrange the values of the price penalty factor in ascending order.

3- Add the maximum generated power of each unit (P_{Gimax}) one by one, starting with the power of the plant with the smallest factor; ŧ P_{Gimax} ° P_{ch}, we stop the calculation.

4- At this stage, F{p} linked to the last unit in the summation process is the price penalty factor corresponding to the given charge.

The mono-objective optimization problem is presented as follows:

$$Min\ F = \sum_{i=1}^{ng} F_{thi}(P_{Gi}) + \sum_{j=1}^{nf} F_{pvj}(P_{pvj}) + \sum_{i=1}^{ng} E_i(P_{Gi}) \ldots(\$/$$

This equation can be rewritten according to the global coefficients and powers generated:

$$Min\ F = \sum_{j=1}^{nf} F_{pvj}(P_{pvj}) + \sum_{i=1}^{ng} A_i P_{Gi}^2 + B_i P_{Gi} + C_i \ldots(\$/h)$$

With 伴侵侵侵

expensive solutions. Stability after a certain number of generations indicates reaching the saturation point or the maximum possible improvement in the constraints and available resources, that is, it shows a process of improvement across generations where the total cost decreases slightly at first and then stabilizes.

The found quantities of productions and gas emissions are summarized in the table III.1:

	Without PV
P_{G1opt}(MW)	143.14
P_{G2opt}(MW)	78.93
P_{G3opt}(MW)	92.92
Production cost($/h)	5215.97
Gas emissions(ton /h)	0.0845
Total cost ($/h)	9882.61

Table.III.1 Results optimal cases without PV

III.2 OPTIMIZATION AFTER THE INSTALLATION OF PHOTOVOLTAIC PRODUCTION

We resume the simulation by considering the addition of photovoltaic production in the electricity production chain. The power of the mini photovoltaic plant is equal to 110 MW.

The results of active powers, emissions rate and total cost are given in the table

	With PV
PG1opt(MW)	77.35
PG2opt(MW)	60.18
PG3opt(MW)	67.46
Production cost($/h)	3135
Gas emissions(ton /h)	0.0757
Total cost ($/h)	6722.63

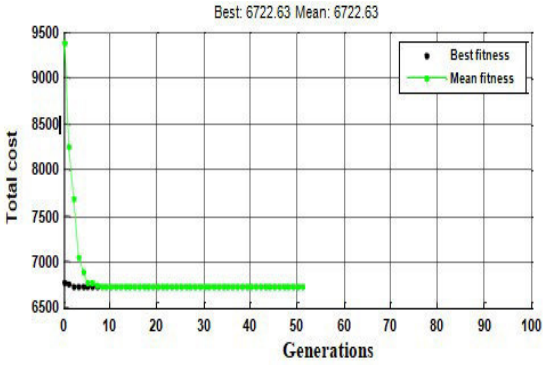
Table.III-2 Optimal results for the case with PV

The plots in the image display the results of an optimization process using a genetic algorithm

1. The first plot:

The horizontal axis (X) represents the generations, which is the number of iterations per formed by the genetic algorithm. The vertical axis (Y) represents the total cost.

The black dots represent the best fitness for each generation, and the green line represents the mean fitness for each generation.

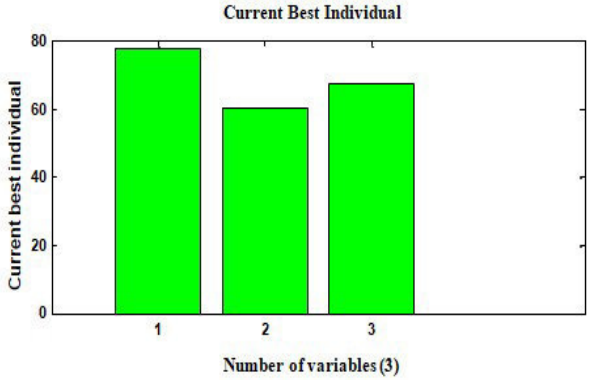


From the plot, we observe that the total cost drops sharply at the beginning during the first few generations quickly at first, but then the improvements become smaller and less frequent. At around generation 100, the total cost stabilizes around 6722.63, this means that the algorithm has approached the best possible solution

2. The second plot:

The first column represents the value of the first variable (generator) and shows it at around the level of 75. The second column represents the value of the second variable (generator) and shows it at around the level of 55.

The third column represents the value of the third variable (generator) and shows it at around the level of 65. This chart gives us an idea of the current distribution of the variable values in the best current individual, reflecting the optimal composition of the individual based on the three variables



The results after integrating PV into the electrical network showed a difference in terms of the total cost.

	Without PV	With PV
PG1opt(MW)	143.14	77.35
PG2opt(MW)	78.93	60.18
PG3opt(MW)	92.92	67.46
Production cost($/h)	5216.97	3135
Gas emissions(ton /h)	0.0845	0.0757
Total cost ($/h)	9882.61	6722.63

Table.III-3 Optimal results for two cases (without PV /with PV)

Based on the findings, it can be concluded that there is a notable disparity in the values of total cost and toxic emissions between those examined with and without photovoltaic (PV) integration:

The total cost associated with the specified load decreased from 9882.61$/h to 6722.63$/h, resulting in a reduction of 3259.95$/h, or 32.98%. Furthermore, gas emissions witnessed a decrease of 2028 Kg/Day, which is equivalent to 1816,8 Kg/day, or 13%.

These results underscore the significant impact of PV integration on both cost reduction and environmental sustainability, highlighting the potential benefits of incorporating renewable energy sources into energy systems.

IV. CONCLUSION

The economic and environmental analysis of the integration of photovoltaic (PV) production into the electricity network using a genetic algorithm gave very satisfactory results. The comparison of the results obtained before and after the integration of PV production into the electricity network highlighted considerable savings in electrical energy production and a notable reduction in greenhouse gas emissions into the atmosphere. In conclusion, the implementation of photovoltaic production demonstrates a promising step towards achieving sustainable energy goals and mitigating climate change impacts. This study underscores the potential of renewable energy integration to enhance both economic prosperity and environmental preservation.

REFERENCES

[1] Abdelhak, G., Abdellatif, N., & Youcef, D. Optimisation of a renewable energy system by hybridisation PSO algorithm and Artificial Neural Network. Przeglad Elektrotechniczny, 2023(10).

[2] KUMARAR, T. Praveen, GANAPATHY, S., et MANIKANDAN, M. Voltage Stability Analysis for Grid Connected PV System using Optimized Control on IOT based ANFIS. Przeglad Elektrotechniczny, 2024, vol. 2024, no 2.

[3] Taghizad-Tavana, K., Alizadeh, A. A., Ghanbari-Ghalehjoughi, M., & Nojavan, S. A comprehensive review of electric vehicles in energy systems: Integration with renewable energy sources, charging levels, different types, and standards. Energies, 16(2), 630. (2023).

[4] Smith, C., & Torrente-Murciano, L. The importance of dynamic operation and renewable energy source on the economic feasibility of green ammonia. Joule, 8(1), 157-174. (2024).

[5] Mamidala, S., Mohan, G., & Veeramani, C. (2023). Iterative modelling of convergent divergent passage to embed over the renewable energy resources to augment the power generation. Przeglad Elektrotechniczny, 99(1).

[6] Gayen, D., Chatterjee, R., & Roy, S. A review on environmental impacts of renewable energy for sustainable development. International Journal of Environmental Science and Technology, 21(5), 5285-5310. (2024).

[7] JAIN, D. K., et al. Battery Energy Storage System for Large Scale Penetration of Renewable Energy Sources. Przeglad Elektrotechniczny, 2023, vol. 2023, no 11.

[8] MANUSOV, V., KIRGIZOV, A., SAFARALIEV, M., ZICMANE, I., BERYOZKINA, S., & SULTONOV, S. Stochastic Method for Predicting the Output of Electrical Energy Received from a Solar Panel. Przeglad Elektrotechniczny, 2024(2).

[9] Alkhayyat, M. T., Khalaf, L. A., & Suliman, M. Y. Adaptive Control for Power Management Based on Renewable Energy. Przeglad Elektrotechniczny, 2023(12).

[10] Nadia, B., Nadia, B. S. A., & Nora, Z. Impact of Renewable Resources Penetration on Maximum Loading Point and dynamic voltage stability. Przeglad Elektrotechniczny, 2023(9).

[11] Abido, M. A. Environmental/economic power dispatch using multiobjective evolutionary algorithms. IEEE transactions on power systems, 18(4), 1529-1537. (2003).

[12] Mishra, S., & Shaik, A. G. Solving bi-objective economic-emission load dispatch of diesel-wind-solar microgrid using African vulture optimization algorithm. Heliyon, 10(3). (2024).

[13] Adetoro, S. A., Olatomiwa, L., Tsado, J., & Dauda, S. M. Techno-economic and environmental impact assessment of a hybrid renewable energy system employing an enhanced combined dispatch strategy. Green Energy and Resources, 1(4), 100044. (2023).

[14] Soni, J., & Bhattacharjee, K. Sine-Cosine Algorithm for the Dynamic Economic Dispatch Problem With the Valve-Point Loading Effect. International Journal of Swarm Intelligence Research, 14(1). (2023).

[15] TUAN-HO, L. E. et BON, Nguyen Nhan. Bi-objective robust design optimization for LED lens design. Przeglad Elektrotechniczny, 2024, vol. 2024, no 4.

[16] Dashtdar, M., Flah, A., Hosseinimoghadam, S. M. S., Reddy, C. R., Kotb, H., AboRas, K. M., ... & Jasiński, M. Solving the environmental/economic dispatch problem using the hybrid FA-GA multi-objective algorithm. Energy Reports, 8, 13766-13779. (2022).

[17] Abdollahzadeh Sangrodi, R., Samiei-Moghaddam, M., & Hajizadeh, A. Multi-area economic/emission dispatch considering the impact of the renewable energy resources and electric vehicles. International Journal of Nonlinear Analysis and Applications, 13(2), 3083-3094. (2022).

[18] Dey, B., Bhattacharyya, B., & Márquez, F. P. G. A hybrid optimization-based approach to solve environment constrained economic dispatch problem on microgrid system. Journal of Cleaner Production, 307, 127196. (2021).

[19] Habib, S., Kamarposhti, M. A., Shokouhandeh, H., & Colak, I. Economic dispatch optimization considering operation cost and environmental

constraints using the HBMO method. Energy Reports, 10, 1718-1725. (2023).

[20] Huang, C., Wang, J., Deng, S., & Yue, D. Real-time distributed economic dispatch scheme of grid-connected microgrid considering cyberattacks. IET Renewable Power Generation, 14(14), 2750-2758. (2020).

[21] Nagarajan, K., Rajagopalan, A., Bajaj, M., Sitharthan, R., Dost Mohammadi, S. A., & Blazek, V. Optimizing dynamic economic dispatch through an enhanced Cheetah-inspired algorithm for integrated renewable energy and demand-side management. Scientific Reports, 14(1), 3091. (2024).

[22] Liu, X., Li, X., Tian, J., Wang, Y., Xiao, G., & Wang, P. Day-ahead Economic Dispatch of Renewable Energy System considering Wind and Photovoltaic Predicted Output. International Transactions on Electrical Energy Systems, 2022(1), 6082642. (2022).

[23] Kaur, P., Chaturvedi, K. T., & Kolhe, M. L. Economic Dispatch of Combined Heat and Power Plant Units within Energy Network Integrated with Wind Power Plant. Processes, 11(4), 1232. (2023).

[24] Mazzoni, S., Ooi, S., Nastasi, B., & Romagnoli, A. Energy storage technologies as techno-economic parameters for master-planning and optimal dispatch in smart multi energy systems. Applied Energy, 254, 113682. (2019).

[25] Li, W., Die, X., Ma, Z., Zhang, J., & Dong, H. Two-Stage Optimal Dispatching of Wind Power-Photovoltaic-Solar Thermal Combined System Considering Economic Optimality and Fairness. Energy Engineering, 120(4), 1001-1022. (2023).

[26] Samal, R. K., & Tripathy, M. Cost and emission additionality of wind energy in power systems. Sustainable Energy, Grids and Networks, 17, 100179. (2019).

[27] Xu, J., Wang, F., Lv, C., Huang, Q., & Xie, H. Economic-environmental equilibrium based optimal scheduling strategy towards wind-solar-thermal power generation system under limited resources. Applied energy, 231, 355-371. (2018).

[28] Jin, J., Zhou, P., Li, C., Guo, X., & Zhang, M. Low-carbon power dispatch with wind power based on carbon trading mechanism. Energy, 170, 250-260. (2019).

[29] Basu, M. Economic environmental dispatch of solar-wind-hydro-thermal power system. Renewable Energy Focus, 30, 107-122. (2019).

[30] Biswas, P. P., Suganthan, P. N., Qu, B. Y., & Amaratunga, G. A. Multiobjective economic-environmental power dispatch with stochastic wind-solar-small hydro power. Energy, 150, 1039-1057. (2018).

[31] Elattar, E. E. Modified harmony search algorithm for combined economic emission dispatch of microgrid incorporating renewable sources. Energy, 159, 496-507. (2018).

[32] Altinoz, O. T. The distributed many-objective economic/emission load dispatch benchmark problem. Swarm and Evolutionary Computation, 49, 102-113. (2019).

[33] Salkuti, S. R. Multi-objective based economic environmental dispatch with stochastic solar-wind-thermal power system. International Journal of Electrical and Computer Engineering, 10(5), 4543. (2020).

[34] Cao, J., Yang, Y., Qu, N., Xi, Y., Guo, X., & Dong, Y. A low-carbon economic dispatch method for regional integrated energy system based on multi-objective chaotic artificial hummingbird algorithm. Scientific Reports, 14(1), 4129. (2024).

[35] Mishra, S., & Shaik, A. G. Solving bi-objective economic-emission load dispatch of diesel-wind-solar microgrid using African vulture optimization algorithm. Heliyon, 10(3). (2024).

[36] Wallach, Yehuda. Calculations and programs for power system/networks. Wayne State Univ., 1986.

[37] Elattar, E. Combined economic emission dispatch for microgrid considering solar and wind power cost functions using harmony search algorithm. Int. J. Appl. Eng. Res., 13(5), 3083-3092. (2018).

[38] Venkatesh, P., Gnanadass, R., & Padhy, N. P. Comparison and application of evolutionary programming techniques to combined economic emission dispatch with line flow constraints. IEEE Transactions on Power systems, 18(2), 688-697. (2003).

[39] Ma, R., & Yan, H.. Multi-objective optimal block transaction model based transient stability evaluation. In Proceedings of the 5th WSEAS International Conference on Applied Computer Science, Hangzhou, China. April , (2006).

Optimized MPPT for PV Systems With Grey Wolf Optimization Algorithm Under Partial Shading Condition

Alla Eddine Boukhedenna
Department of EEA
National Higher School of Technology and Engineering (ENSTI)
LTSE Laboratory
Annaba, Algeria
a.boukhdenna@etu.ensti-annaba.dz

Hamza Afghoul
Department of Electrotechnics
University of Ferhat Abbas Setif-1
LEPCI Laboratory, Faculty of Technology
Setif, Algeria
hamza.afghoul@univ-setif.dz

Djallal Eddine Zabia
Department of Electrical Engineering
University of Mohamed Khider Biskra
LI3CUB Laboratory, Faculty of Technology
Biskra, Algeria
djallaleddine.zabia@univ-biskra.dz

Belgacem Med Nassim Bouzidi
Department of Electrotechnics
University of Djelfa
LAADI Laboratory
Djelfa, Algeria
bouzidinassim23@gmail.com

Feriel Abdelmalek
Department of Electronics University of Ferhat Abbas Setif-1
LEPCI Laboratory
Setif, Algeria
feriel.abdelmalek@univ-setif.dz

Soumia Merah
Department of EEA
National Higher School of Technology and Engineering (ENSTI)
LTSE Laboratory
Annaba, Algeria
s.merah@etu.ensti-annaba.dz

***Abstract*— The efficacy of photovoltaic (PV) systems, which are a fundamental component of renewable energy solutions, is greatly impacted by natural fluctuations, particularly in solar irradiation. These fluctuations introduce complications in the maintenance of a consistent and optimal power output, necessitating the implementation of effective algorithm to optimize energy efficiency. Tracking the Maximum Power Point (MPP) is a critical component of PV system operation. This paper evaluates the efficacy of the Grey Wolf Optimization (GWO) algorithm in monitoring the MPP under various solar irradiation conditions, including uniform irradiation and partial shading. The MATLAB/SIMULINK simulation results indicate that the P&O algorithm performs well in scenarios of uniform solar irradiation, its limitations become evident when subjected to dynamic and complex conditions, such as partial shading. Specifically, the Perturb and Observe (P&O) algorithm exhibited significant power oscillations, which compromise system efficiency and reliability. Conversely, the GWO algorithm displayed exceptional tracking efficacy in partial shading conditions, illustrating its capacity to effectively adapt to these obstacles. The extraction of maximal power is facilitated by its rapid and precise response to dynamic changes in irradiation levels, making it a more reliable choice for optimizing PV systems.**

Keywords*— *Photovoltaic system, Grey wolf optimization, Perturb and Observe, Maximum power point, Partial shading.

I. INTRODUCTION

Rapid advances in technology and the rapid development of industrial sectors have resulted in a huge rise in the demand for electrical energy. In modern times, this energy is essential to digital technology, mechanical function, industrial production, and even everyday household tasks. This has led to several concerns about the availability of energy sources, especially with respect to the challenges related to their stability. Traditional energy sources like coal, oil, and natural gas are finite resources that need significant work and sophisticated technology to obtain, in addition to the high costs of production and transportation. These sources also have a negative impact on the natural world. Developing alternative energy sources that may satisfy the world's growing energy needs without harming the environment has become important in view of these problems. The best way to overcome these challenges is to use renewable energy sources. Because of its many benefits and low production costs, solar energy is a popular renewable energy source [1].

However, optimizing the potential of solar power presents with difficulties [2], main among which is the variability brought about by natural components like illumination and temperature. These factors may have a significant impact on photovoltaic (PV) cell efficiency, which in turn impacts the overall energy generation. To address this difficulty, researchers have developed a number of algorithms to maximize PV systems' power generation [3].

The Perturb and Observe (P&O) algorithm is a popular choice for Maximum Power Point Tracking (MPPT) because to its ease of use and efficiency Periodically altering the system's operational voltage or current while recording the consequent power change is how the P&O algorithm works [4]. When the power rises, the perturbation continues to move in the same direction, when it falls [5], the direction changes. With the help of this input mechanism, the system may respond rather quickly and converge towards the Maximum Power Point (MPP) [6]. Notwithstanding its benefits, the P&O algorithm sometimes has trouble with fluctuations around the MPP, particularly in situations when the environment is changing quickly. The grey wolf optimization (GWO) algorithm, on the other hand, provides a more reliable method as it is modelled after the social structure and hunting habits of grey wolves. Alpha, Beta, Delta, and Omega are the four social role categories into which GWO divides wolves [7]. The wolves' hunting and position-adjusting tactics are determined by these responsibilities, and the system is mathematically modelled to help it converge towards the best answer [8]. The GWO algorithm offers better search capabilities and tracking effectiveness by imitating the social structure and hunting behavior of wolves, especially in dynamic and complicated situations [9].

This paper evaluates the efficacy of the GWO algorithm and the P&O algorithm in photovoltaic systems for MPPT under a variety of environmental conditions. The two algorithms are compared using MATLAB/SIMULINK simulations in the research. Both P&O and GWO exhibit effective tracking of MPP under uniform solar radiation.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

However, substantial disparities emerge when non-uniform conditions, such as rapid fluctuations in solar radiation or partial shading, are present. The P&O algorithm is incapable of reaching the MPP. In contrast, the GWO algorithm, which is influenced by the hunting and social behavior of grey wolves, exhibits increased robustness in tracking the MPP under these conditions, providing more precise monitoring and faster response times than P&O. This makes GWO a more dependable option for the management of complex and variable environmental factors in photovoltaic systems.

II. PV SYSTEM MODELING

The equivalent circuit's most basic model in the context of photovoltaic cells comprises two primary components that are connected in parallel. The initial component is an ideal current source that produces a constant current, denoted by I_{pv}. The density of photons that impact the cell surface is directly correlated with this output. The second component, a real diode, represents the electrical properties of the p-n junction in the solar cell. Fig. 1 illustrates the electric model that is typically used to symbolize a PV panel [10].

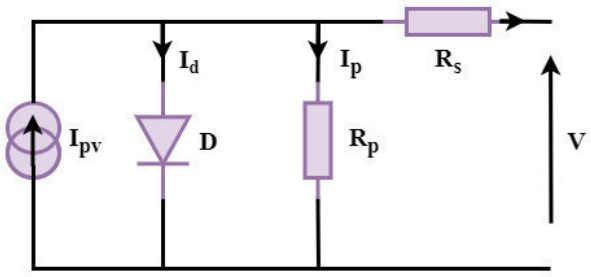


Fig. 1. The electric model of a PV Cell

The output current of the solar cell is determined by the following equation [12][11]:

$$I = I_{ph} - I_d - \frac{(V + R_s I)}{R_p} \quad (1)$$

With:

$$\mathrm{I_d} = \mathrm{I_o}[e^{q\frac{V+R_S I}{nKT_c}} - 1] \quad (2)$$

$$\mathrm{I_{ph}} = \frac{G}{G_{ref}}(I_{ph,ref} + \mu_{sc}.\Delta T) \quad (3)$$

The final PV system equation is calculated using [12]:

$$\mathrm{I_{pv}} = \frac{G}{G_{ref}}(I_{ph_ref} + \mu_{sc}.\Delta T) - \mathrm{I_o}[e^{q\frac{V+Rs.I}{nKT_c}} - 1] - \frac{V+Rs.I}{R_{sh}} \quad (4)$$

Where:

I_{ph}	Photocurrent.
I_d	Diode current.
G	Irradiance.
G_{ref}	Standard test condition.
I_{ph_ref}	photocurrent at STC.
μ_{sc}	coefficient temperature of short circuit current.
ΔT	Cell temperature at STC.
$\mathrm{I_o}$	Saturation current.
q	Electron charge.
n	Ideality criterion.
K	Boltzman constant.
T_c	actual cell temperature.
Rs	serial resistance.
R_{sh}	equivalent circuit shunt resistance.

TABLE I. SHOWS THE PV PANEL VALUES

Variable	Parameters	Values
P_{max}	Maximum-power	363.3 [W]
V_{max}	Maximum voltage	68.2 [V]
I_{max}	Maximum current	6.3 [A]
I_{sc}	Short-circuit current	7 [A]
V_{oc}	Open-circuit voltage	22.9 [V]

III. PARTIAL SHADING'S EFFECTS

PV system is arranged using PV panels connected in a series parallel configuration, as illustrated in in Fig. 2.

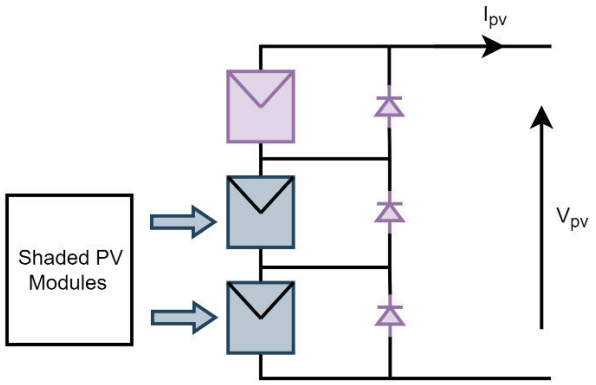


Fig. 2. Partially Shaded PV array.

Fig. 3 illustrates a typical power-voltage curve under uniform solar radiation, which clearly demonstrates the MPP. Nevertheless, the application of partial shade, which is induced by factors such as clouds or adjacent objects like structures or trees, results in the appearance of numerous peaks within the power-voltage curve, as illustrated in Fig. 4. The variation in shading leads to varying levels of power output all over the photovoltaic array, resulting in the formation of multiple peak locations in the curve. This complicates the process for determining the MPPT.

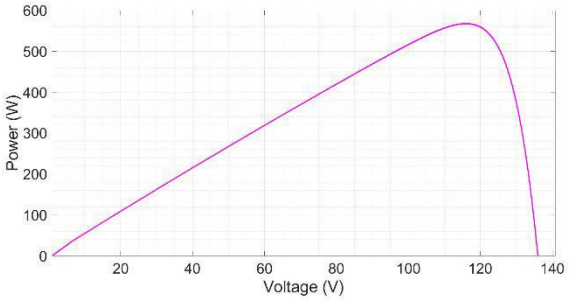


Fig. 3. Power-voltage (P-V) curve of PV array under standard test conditions (800W/m^2).

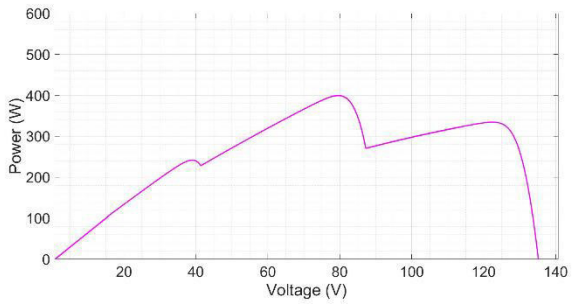


Fig. 4. Power-voltage (P-V) curve of PV array under partial shading (1000W/m² - 650W/m² - 300W/m²).

IV. Exploration of Grey Wolf Optimization and Perturb and Observe Algorithms

A. *Perturb and Observe algorithm*

In the realm of MPP tracking techniques, the P&O algorithm is known for its simplicity and ease of implementation [13]. This method utilises periodic perturbations of the PV system's operating voltage (V_{pv}) or current (I_{pv}) for tracking resulting variations in output power (P_{pv}). If the power increases as a result of the perturbation, the algorithm continues in the same direction. Otherwise, the direction of the perturbation is reversed to approach the MPP [14].

The change in power with respect to voltage can be expressed as:

$$\Delta P / \Delta V \tag{5}$$

When $\Delta P / \Delta V \geq 0$, the operating point moves closer to the MPP by increasing V_{pv}, and when $\Delta P / \Delta V \leq 0$ it moves closer by decreasing V_{pv}. The algorithm adjusts the duty cycle of the DC-DC converter as follows:

$$D_{new} = D_{old} + \Delta D \tag{6}$$

Where ΔD is the increment or decrement determined by the P&O logic.

The steps of the P&O algorithm are simplified in the flowchart in Fig. 5.

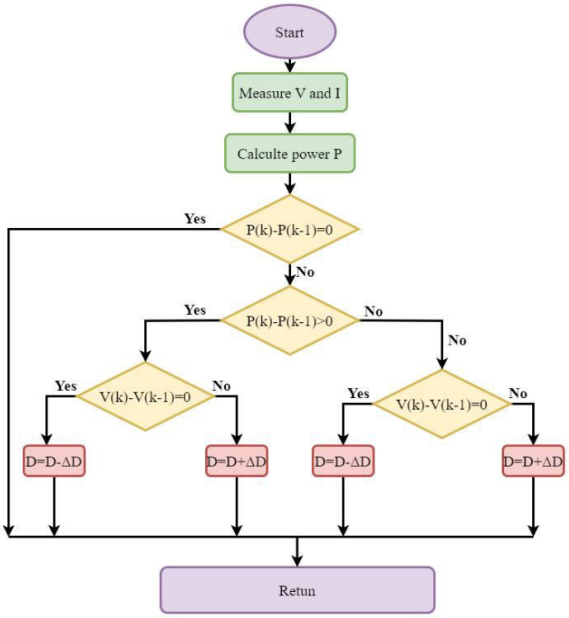


Fig. 5. flowchart of the P&O algorithm.

B. *Grey wolf optimization algorithem*

The grey wolf strategy is derived from the behaviors of the grey wolf during its quest for prey, during which they distribute responsibilities among the members of the pack. Fig. 6 illustrates the formation of a social hierarchy as a consequence of this division.

The alpha level refers to the wolves responsible for making decisions, while the second level Beta wolves are subordinate wolves who assist the Alphas in decision-making. Delta wolves comprise the third level of the social hierarchy, and the omega wolf serves as a "scapegoat" to capture prey [15].

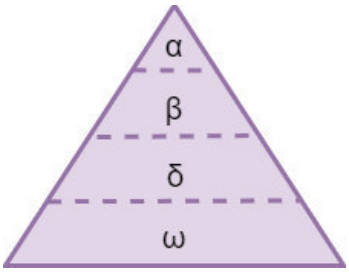


Fig. 6. Hierarchy of grey wolf.

The process of hunting prey goes through three steps: tracking, encircling, and attacking. These steps can be modeled with the following mathematical equations [16]:

$$\vec{D} = | \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) | \tag{7}$$

$$\vec{X}_p(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \tag{8}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{9}$$

With:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{10}$$

$$\vec{C} = 2 \cdot \vec{r}_2 \tag{11}$$

Where:

t	Iteration count.
$\vec{X}_p(t)$	Position vector of the prey.
$\vec{X}(t)$	Position vector of a grey.
$\vec{a}$	Variable whose value decreased linearly [2 to 0].
$\vec{r}_{1,2}$	Random number in [0-1].

The variability of radiation levels makes it challenging to determine MPP effectively[17]. Through the use of the GWO algorithm, PV array performance may be enhanced by varying the duty cycle 'D', as represented by the following equation:

$$\vec{D}_c(t+1) = \vec{D}_c(t) - \vec{A} \cdot \vec{D} \tag{12}$$

The fitness function of the GWO algorithm is expressed as follows:

$$P(d_i^t) > P(d_i^{t-1}) \tag{13}$$

Where:

p	Power generated.
d	Duty cycle.
i	Number of current grey wolves.

GWO algorithm follows basic steps,as depicted in the flowchart in Fig. 7. The first step involves initializing the initial number of agents. Secondly, following the evaluation of the MPP location, the duty cycle positions should be updated. Lastly, the standards are stopped, and the algorithm should be reset to begin the search for the MPP anew in the event of low power.

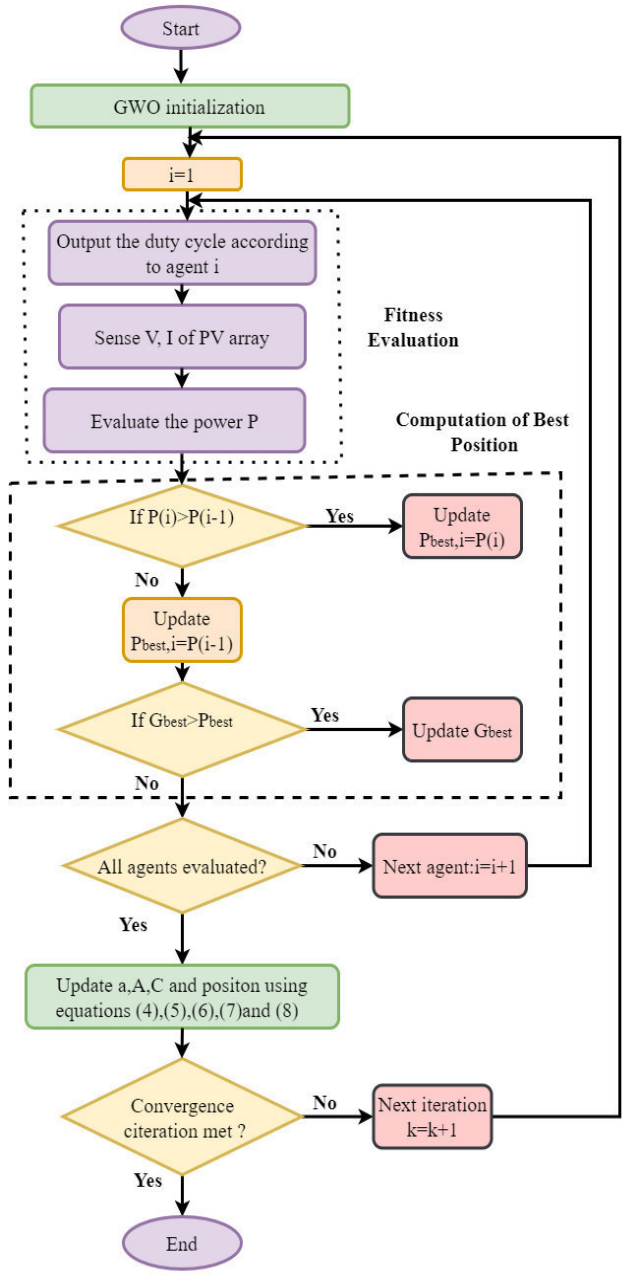


Fig. 7. flowchart of the GWO algorithm.

V. Simulation And Results

The simulation and system results were obtained using MATLAB/SIMULINK under standard test settings (STC), with uniform solar irradiation (800W/m²) in the first scenario and varying radiation levels (1000W/m² - 800W/m² - 450W/m²) in the next scenario, while maintaining a constant temperature of 25°C. The P&O and GWO algorithms were evaluated and contrasted for their Maximum Power Point (MPP) tracking efficiency.

A. *Uniform solar irradiation*

Fig. 8 illustrates the results of the P&O and GWO algorithms when a photovoltaic array is exposed to uniform shading with a radiation level of 800 W/m². In terms of tracking efficiency, the P&O algorithm achieved 98.91%, while the GWO algorithm attained 99.72%. The MPP was 567.17 W in both cases. However, a noticeable difference was observed in the system's stability and energy behavior. While the GWO algorithm demonstrated a smoother and more stable tracking process, the P&O algorithm exhibited energy fluctuations during its operation and achieved a quicker settling time of 0.13 s compared to the GWO algorithm's 0.88 s.

B. *Partial shading*

Fig. 9 presents the results for the second case, where the PV system is subjected to varying shading conditions with irradiation levels of 1000 W/m², 650 W/m², and 450 W/m². The MPP was ascertained to be 399.36 W. The P&O algorithm had a tracking effectiveness of 81.02%, markedly inferior than the 98.8% attained by the GWO method. Furthermore, the P&O method exhibited a settling time of 0.12 s, in contrast to the GWO algorithm's extended settling duration of 1.21 s.

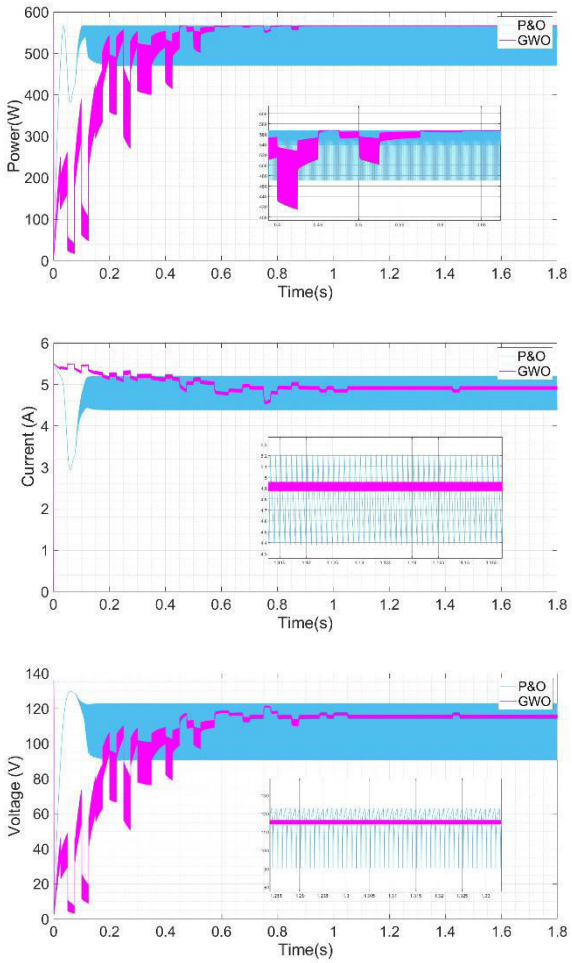


Fig. 8. Comparison results between P&O algorithm and GWO under a uniform shade (800W/m^2).

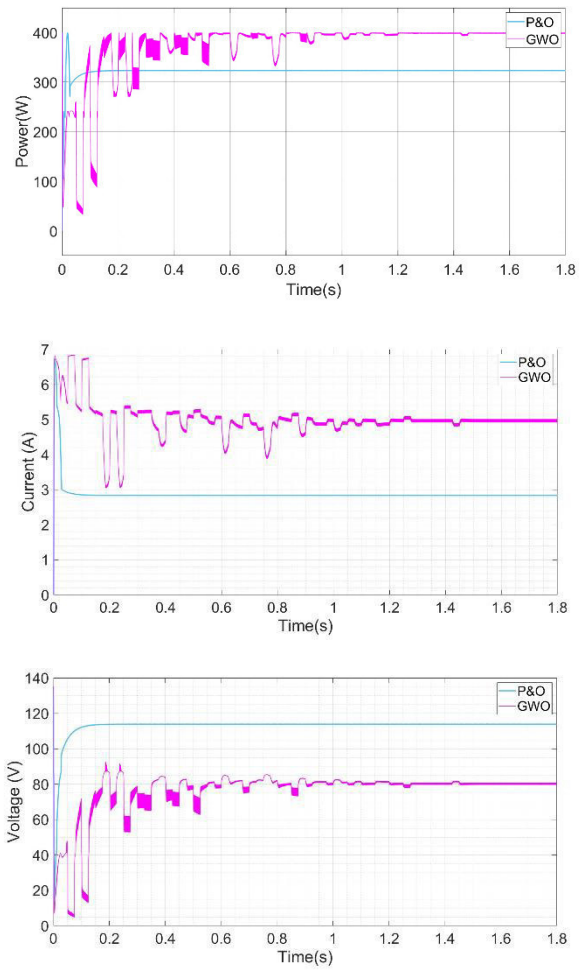


Fig. 9. Comparison results between P&O algorithm and GWO under different shading (1000W/m^2 - 800W/m^2 - 450W/m^2).

VI. CONCLUSION

This study provides an analytical model of a photovoltaic power system and performs a comparative analysis of maximum power point tracking, including the perturb and observe (P&O) and grey wolf optimisation (GWO) algorithms under scenarios of uniform solar radiation and partial shading. The simulation results obtained using MATLAB/SIMULINK indicate that the GWO approach is superior, with a tracking effectiveness of 97% under both uniform illumination and partial shading conditions. The P&O method exhibits commendable performance for the settling period. nonetheless, it is not resilient to partial darkening. It is deficient in tracking at the MPP, achieving a tracking effectiveness of just 75.09%.

REFERENCES

[1] M. Atef, T. Khatib, M. F. Abdullah, and M. F. Romlie, "Optimization of a Hybrid Solar PV and Gas Turbine Generator System Using the Loss of Load Probability Index," Clean Technol., vol. 2, no. 3, pp. 240–251, Jul. 2020, doi: 10.3390/cleantechnol2030016.

[2] M. V. Da Rocha, L. P. Sampaio, and S. A. Oliveira Da Silva, "Comparative Analysis of ABC, Bat, GWO and PSO Algorithms for MPPT in PV Systems," in 2019 8th International Conference on Renewable Energy Research and Applications (ICRERA), Brasov, Romania: IEEE, Nov. 2019, pp. 347–352. doi: 10.1109/ICRERA47325.2019.8996520.

[3] A. M. Eltamaly, "A novel musical chairs algorithm applied for MPPT of PV systems," Renew. Sustain. Energy Rev., vol. 146, p. 111135, Aug. 2021, doi: 10.1016/j.rser.2021.111135.

[4] J. Ahmed and Z. Salam, "An Enhanced Adaptive P&O MPPT for Fast and Efficient Tracking Under Varying Environmental Conditions," IEEE Trans. Sustain. Energy, vol. 9, no. 3, pp. 1487–1496, Jul. 2018, doi: 10.1109/TSTE.2018.2791968.

[5] K. Sundareswaran, V. Vigneshkumar, P. Sankar, S. P. Simon, P. Srinivasa Rao Nayak, and S. Palani, "Development of an Improved P&O Algorithm Assisted Through a Colony of Foraging Ants for MPPT in PV System," IEEE Trans. Ind. Inform., vol. 12, no. 1, pp. 187–200, Feb. 2016, doi: 10.1109/TII.2015.2502428.

[6] M. V. Da Rocha, L. P. Sampaio, and S. A. O. Da Silva, "Comparative analysis of MPPT algorithms based on Bat algorithm for PV systems under partial shading condition," Sustain. Energy Technol. Assess., vol. 40, p. 100761, Aug. 2020, doi: 10.1016/j.seta.2020.100761.

[7] S. Bhowmik, G. Shankar, S. Paul, K. Chatterjee, S. Kumar, and V. L. Srinivas, "Maximum Power Point Tracking of Grid Connected PV System Under Partial Shading Condition Using Grey Wolf Algorithm," in 2023 First International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV), Hyderabad, India: IEEE, Sep. 2023, pp. 1–6. doi: 10.1109/ICPEEV58650.2023.10391899.

[8] D. J. K. Kishore, M. R. Mohamed, K. Sudhakar, and K. Peddakapu, "An improved grey wolf optimization based MPPT algorithm for photovoltaic systems under diverse partial shading conditions," J. Phys. Conf. Ser., vol. 2312, no. 1, p. 012063, Aug. 2022, doi: 10.1088/1742-6596/2312/1/012063.

[9] J. Lu, F. Zhan, and C. Fan, "Composite PV MPPT algorithm based on improved grey wolf optimization algorithm with perturbation observation algorithm," in 5th International Conference on Information Science, Electrical, and Automation Engineering (ISEAE 2023), T. Lei, Ed., Wuhan, China: SPIE, Aug. 2023, p. 40. doi: 10.1117/12.2689402.

[10] D. E. Zabia et al., "Experimental validation of a novel hybrid Equilibrium Slime Mould Optimization for solar photovoltaic system," Heliyon, vol. 10, no. 20, p. e38943, Oct. 2024, doi: 10.1016/j.heliyon.2024.e38943.

[11] Laboratory of Identification, Command, Control and Communication (LI3CUB), University of Biskra, Algeria et al., "Improving Photovoltaic Grid Integration under Partial Shading by Equilibrium Slime Mould Optimization," Electroteh. Electron. Autom., vol. 71, no. 3, pp. 15–23, Nov. 2023, doi: 10.46904/eea.23.71.4.1108002.

[12] A. Kchaou, A. Naamane, Y. Koubaa, and N. M'sirdi, "Second order sliding mode-based MPPT control for photovoltaic applications," Sol. Energy, vol. 155, pp. 758–769, Oct. 2017, doi: 10.1016/j.solener.2017.07.007.

[13] C. Wu and H. Zhan, "Photovoltaic MPPT Strategy Based on Hybrid Particle Swarm Optimization Algorithm and Grey Wolf Algorithm," in 2023 5th International Conference on Smart Power & Internet Energy Systems (SPIES), Shenyang, China: IEEE, Dec. 2023, pp. 216–221. doi: 10.1109/SPIES60658.2023.10474932.

[14] J. Ahmed and Z. Salam, "An improved perturb and observe (P&O) maximum power point tracking (MPPT) algorithm for higher efficiency," Appl. Energy, vol. 150, pp. 97–108, Jul. 2015, doi: 10.1016/j.apenergy.2015.04.006.

[15] S. Mohanty, B. Subudhi, and P. K. Ray, "A grey wolf optimization based MPPT for PV system under changing insolation level," in 2016 IEEE Students' Technology Symposium (TechSym), Kharagpur, West Bengal, India: IEEE, Sep. 2016, pp. 175–179. doi: 10.1109/TechSym.2016.7872677.

[16] S. K. Cherukuri and S. R. Rayapudi, "Enhanced Grey Wolf Optimizer Based MPPT Algorithm of PV System Under Partial Shaded Condition," Int. J. Renew. Energy Dev., vol. 6, no. 3, pp. 203–212, Nov. 2017, doi: 10.14710/ijred.6.3.203-212.

[17] M. Y. Silaa, O. Barambones, A. Bencherif, and A. Rahmani, "A New MPPT-Based Extended Grey Wolf Optimizer for Stand-Alone PV System: A Performance Evaluation versus Four Smart MPPT Techniques in Diverse Scenarios," Inventions, vol. 8, no. 6, p. 142, Nov. 2023, doi: 10.3390/inventions8060142.

Performance Evaluation of Fisher Linear Discriminant Models Based Fault Detection and Diagnosis in PV Power Generation Systems

SAIDI Ahmed
Department of Electrical Engineering, Laboratory of Sustainable Development and Computer Science, Ahmed Draia University Adrar 01000, Algeria
saidi.ahmed@univ-adrar.edu.dz

ABDELOUAHED Touhami
Faculty of technologie , Departement of Electrical Engineering, University Mhamed Bougara of Boumerdes 35000, Algeria
a.touhami@univ-boumerdes.dz

DRAOUI Abdelghani
Laboratory of environmental and energy systems (LSEE)
University center Ali Kafi
Tindouf 37000, Algeria
draoui.abdelghani@gmail.com

MASMOUDI Ramadhan
Department of Electrical Engineering, Laboratory of Sustainable Development and Computer Science, Ahmed Draia University Adrar 01000, Algeria
ram.masmoudi@univ-adrar.edu.dz

***Abstract*— In photovoltaic (PV) power generation systems, ensuring robust and efficient fault detection and diagnosis (FDD) is essential for achieving optimal performance and long-term reliability. This study introduces a novel FDD methodology based on Fisher Random Matrix Theory (RMT), tailored to address the inherent complexities of fault detection in PV systems. The approach exploits multivariate statistical correlations within high-dimensional operational data, enabling real-time fault detection with exceptional accuracy while maintaining resilience against environmental noise—a common limitation of conventional methods. By incorporating Fisher RMT, the proposed technique significantly enhances fault isolation, allowing precise identification and diagnosis of faults even under varying irradiance and temperature conditions.**

The methodology is validated using real-world PV system data, including measurements of voltage, current, power, irradiance, and temperature. Results demonstrate notable improvements in detection accuracy and fault discrimination compared to traditional I-V curve-based techniques. Furthermore, the approach effectively classifies diverse fault types—such as open-circuit, shading, and short-circuit faults—with faster detection times and reduced false alarm rates. The findings highlight that the Fisher RMT-based methodology outperforms existing FDD strategies by offering higher precision, quicker response times, and fewer false positives, presenting a scalable and reliable solution for real-time monitoring of large-scale PV systems.

Keywords— Fault Detection and Diagnosis (FDD), Photovoltaic Systems, MPPT, Fisher Random Matrix Theory (RMT), Real-time Monitoring, Multidimensional Data Analysis.

I. INTRODUCTION

Photovoltaic (PV) power generation systems are integral to the transition toward renewable energy. As PV system deployment increases, so does the need for effective monitoring, fault detection, and diagnosis (FDD) techniques to ensure optimal performance and reduce downtime[1]. Faults such as shading, open-circuit, or short-circuit can lead to significant energy losses and pose safety risks, underscoring the importance of developing efficient FDD methods to maintain system reliability [2].

PV systems encounter various faults, including open-circuit, short-circuit, shading, and ground faults, each impacting performance uniquely [3]. For example, shading faults result from reduced sunlight exposure due to obstructions, while open-circuit and short-circuit faults often arise from wiring issues or component failures. Ground faults, which occur due to unintended electrical connections between conductors and the ground, present significant safety risks, including the potential for fires and other hazardous conditions [4]. To address these challenges, both traditional and advanced fault detection techniques have been developed and implemented [5]. Traditional methods often rely on basic electrical measurements and manual inspections, while advanced techniques leverage sophisticated algorithms, machine learning, and real-time monitoring systems to enhance detection accuracy and reliability. These advancements are critical for mitigating risks and ensuring the safe and efficient operation of electrical systems.

Early fault detection relied on manual inspections, including visual or infrared imaging, which remain useful for identifying visible surface defects like cracks or corrosion [6]. However, electrical-based methods, such as monitoring current-voltage (I-V) and power-voltage (P-V) characteristics, are more effective for comprehensive fault detection [7]. Deviations from standard I-V curves under test conditions indicate faults, and models like the one-diode model provide a baseline for identifying performance anomalies [8].

Statistical methods have gained traction for their ability to manage noisy and fluctuating data [9]. Techniques like t-tests, wavelet transformations, and Principal Component Analysis (PCA) compare real-time data against historical models to detect anomalies. Machine learning (ML) has further advanced FDD with models like Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Extreme Learning Machines (ELMs), which identify and classify

faults, including complex scenarios involving multiple simultaneous faults [10-13].

Deep learning models, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have demonstrated high accuracy in fault classification [14]. CNNs are particularly effective for image-based fault detection, analyzing electroluminescence images to detect defects such as micro-cracks, while LSTMs excel in analyzing sequential operational data [15].

Hybrid systems combining statistical and machine learning methods are increasingly popular [16-18]. For instance, integrating adaptive thresholding algorithms with PCA improves fault detection under variable irradiance and temperature conditions. Similarly, combining deep learning models with traditional electrical monitoring techniques enhances the scalability and adaptability of FDD systems. Intelligent diagnostic systems leveraging real-time monitoring, cloud-based analytics, and machine learning enable continuous learning from new data, improving detection accuracy over time [19].

Despite advancements, FDD systems face challenges such as handling data under fluctuating environmental conditions, noise, uncertainty, and missing values [20]. Many current methods focus on detecting single fault types, whereas real-world PV systems often experience multiple concurrent faults. Future research should explore hybrid diagnostic models that integrate diverse approaches to enhance accuracy in complex scenarios. Additionally, the adoption of IoT devices and cloud-based platforms could enable remote monitoring and predictive maintenance, improving reliability and efficiency [21].

Emerging techniques like Fisher Random Matrix Theory (RMT) offer promising advancements for FDD in PV systems. RMT utilizes statistical correlations in high-dimensional operational data, making it particularly effective for systems with continuously fluctuating environmental and operational parameters [22]. By leveraging this matrix-based approach, faults can be detected early, and specific fault types diagnosed with higher precision, significantly enhancing the speed and accuracy of FDD systems .

This combination of traditional, statistical, and advanced techniques underscores the potential of integrated approaches to address the growing demands of PV system monitoring and maintenance.

II. MATERIALS AND METHODS

Effective fault detection and diagnosis in photovoltaic systems require methods that can clearly distinguish between normal and faulty operations. A robust approach integrates the Fisher Random Matrix (FRM) with the Mahalanobis distance to establish reliable thresholds for fault identification.

a) Fisher Random Matrix (FRM)

The Fisher Random Matrix builds on Fisher's Linear Discriminant Analysis (LDA), a technique widely employed for dimensionality reduction and classification. LDA aims to project high-dimensional data onto a lower-dimensional space while maximizing class separation, such as between healthy and faulty states in a PV system [23].

The Fisher criterion seeks to optimize the separation between classes by maximizing the ratio of between-class variance (S_b) to within-class variance (S_W). Mathematically, these scatter matrices are defined as:

Between-Class Variance (S_b): Measures how distinct the different classes are.

Within-Class Variance (S_W): Measures how tightly grouped data points are within each class.

The optimal projection vector *w*, is calculated to maximize the Fisher criterion:

$$S_b = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

$$S_w = \sum_{i=1}^{n_1}(x_i - \mu_1)(x_i - \mu_1)^T + \sum_{j=1}^{n_2}(x_j - \mu_2)(x_j - \mu_2)^T$$

The mean vectors of the two classes are represented by μ1 and μ2, while xi and x_j refer to individual data points within each class. The Fisher Random Matrix (FRM) is employed to determine the optimal projection vector that enhances class separability by maximizing between-class variance and minimizing within-class variance. Between-class variance quantifies the distinction between different classes, while within-class variance gauges the concentration of data points within each class. The goal is to project the data onto a new axis where the ratio of between-class variance to within-class variance is maximized. This transformation is governed by the Fisher criterion [24-25].

$$J(w) = \frac{w^T S_b w}{w^T S_w w}$$

Where:

- S_b is the between-class scatter matrix,
- S_w is the within-class scatter matrix,
- w is the projection vector,
- J(w) is the Fisher criterion to be maximized.

This transformation projects the data onto a new axis that emphasizes class separability, enabling more effective classification of healthy and faulty states in PV systems.

The FRM offers several advantages in fault detection for PV systems:

- Improved Class Separation by maximizing the separation between healthy and faulty data, FRM facilitates more accurate anomaly detection.
- Dimensionality Reduction high-dimensional PV data, encompassing variables like voltage, current, irradiance, and temperature, is reduced to a lower-dimensional representation while retaining critical classification information.
- Noise Robustness: The random matrix structure enhances robustness against noise and data variations, ensuring reliability under real-world conditions.

- Scalability: FRM can handle large-scale PV systems, enabling simultaneous analysis of data from numerous PV modules in large arrays.

After transforming the data using the Fisher Random Matrix, the Mahalanobis distance is employed to classify new observations. This distance measures how far a data point deviates from the healthy class distribution, considering the covariance between variables—a common feature in PV system data.

b) Steps in Fault Detection

1. Projection: Data is projected into a space where healthy and faulty classes are well-separated using FRM.
2. Distance Calculation: The Mahalanobis distance of new observations from the healthy class distribution is computed.
3. Thresholding: A threshold is determined based on the healthy class distance distribution. If a new observation's Mahalanobis distance exceeds this threshold, it is classified as faulty.

This combined approach leverages the Fisher Random Matrix for class separation and the Mahalanobis distance for precise classification, offering an effective and scalable solution for real-time PV system fault detection.

III. PV SYSTEM DESCRIPTION

The proposed fault detection strategy was validated using real-world operational data from a large-scale 20 MWp grid-connected photovoltaic (PV) system located in Adrar, Algeria. Adrar, a southeastern Saharan region and the second-largest province in Algeria, spans an area of 427,386 km² and boasts substantial solar energy potential. The proposed fault detection strategy was rigorously validated through experimental testing utilizing real-world data obtained from a 20 MWp grid-connected photovoltaic (PV) system situated in Adrar, a southeastern Saharan region and the second-largest province in Algeria, spanning an area of 427,386 km². This region boasts substantial solar energy potential, with an average solar irradiance of 5.7 kWh/m²/day, which, if harnessed effectively, could fully meet the energy demands of the entire city [5]. The validation process underscores the practical applicability and robustness of the proposed strategy in real-world operational environments.

The PV system, operational since December 10, 2015, directly feeds generated power into the grid. It operates within a DC voltage range of 480–850 V and an AC voltage of 315 V. The facility is structured as a large-scale installation comprising 20 PV arrays, each rated at 1 MW capacity. Each array consists of 93 subarrays, with each subarray containing 44 YINGLI (YL245P-29B) PV modules. The combined power from these subarrays is processed through a 2×550 kW SUNGROW (5SG 500MX) grid-connected inverter and further stepped up using a SUNTEN ZBW10A-1250/30/0.315-0.315 transformer rated at 1250 kVA and 50 Hz for grid connection.

Meteorological parameters critical to the PV system's performance monitoring were measured with high-precision instruments:

- Solar Irradiance: Captured using a Kipp & Zonen CMP21 pyranometer with a daily uncertainty of <2%.
- Temperature: Monitored via a J-type thermocouple with an accuracy of ±1.1°C.
- Wind Speed: Measured using a WE-100 sonic anemometer with a resolution of 0.05 m/s.

Electrical parameters, including both DC and AC outputs from the inverters, and meteorological data were collected and transmitted via a Supervisory Control and Data Acquisition (SCADA) system. The SCADA system integrates advanced graphical user interfaces, computing platforms, and communication networks to enable real-time monitoring and performance analysis of the PV plant.

For fault detection and diagnosis, the study focused on a PV array comprising 44 YINGLI (YL245P-29B) modules, arranged in two parallel strings of 22 modules connected in series. These modules have well-defined electrical characteristics under standard test conditions (STC)—an irradiance of 1000 W/m² and a temperature of 25°C, as detailed in Table I.

The experimental data used for validation was collected between January and March 2017 at 10-minute intervals, encompassing electrical measurements from the inverter and meteorological data from sensors. The system's design and instrumentation ensured the availability of comprehensive and high-quality data for assessing the proposed fault detection strategy.

The Adrar PV plant exemplifies the scalability of modern grid-connected PV systems and serves as an ideal testing ground for advanced fault detection methodologies. The experimental results demonstrate the effectiveness of the proposed strategy in detecting and diagnosing faults, highlighting its applicability in real-world large-scale PV systems.

TABLE I
Key Electrical Specifications of the YINGLI (YL245P-29B) PV Module under Standard Test Conditions (STC)

Peak power (W)	Voltage at maximum power (V)	Current at maximum power (A)	Open circuit voltage (V)	Short circuit current (A)
245	29.6	8.28	37.5	8.83

This study employs a dataset comprising five critical variables essential for monitoring and diagnosing the performance of photovoltaic systems: voltage (V), current (A), power (W), temperature (°C), and irradiance (W/m²). These parameters are fundamental for assessing system functionality and identifying potential faults. To better understand the interrelationships among these variables, exploratory data analysis (EDA) techniques were applied, including the creation of correlation matrices, examination of variable distributions, and generation of box plots. These analyses provide valuable insights into system behavior and underlying data patterns.

The correlation matrix reveals the linear relationships between the various parameters. As shown in Figure 1, voltage and

power exhibit a very strong positive correlation (0.9986), indicating that power output increases substantially with higher voltage levels. Similarly, voltage and irradiance display a high correlation (0.9134), suggesting that irradiance significantly influences voltage output. Conversely, current and temperature exhibit weaker correlations with other parameters, indicating more independent behavior under varying operational conditions.

The distribution plots in Figure 2 provide an overview of the spread and frequency of values for each parameter in the dataset. The voltage distribution is concentrated around lower values, with a few outliers at higher voltage levels. Current distribution is skewed towards lower values, with occasional spikes. Power distribution reflects predominantly low power outputs, with a few high-power observations, which aligns with the typical variability of PV systems. Temperature data exhibits a relatively uniform spread, with most values ranging between 20°C and 50°C. Irradiance is primarily concentrated at lower levels, highlighting periods of reduced solar exposure.

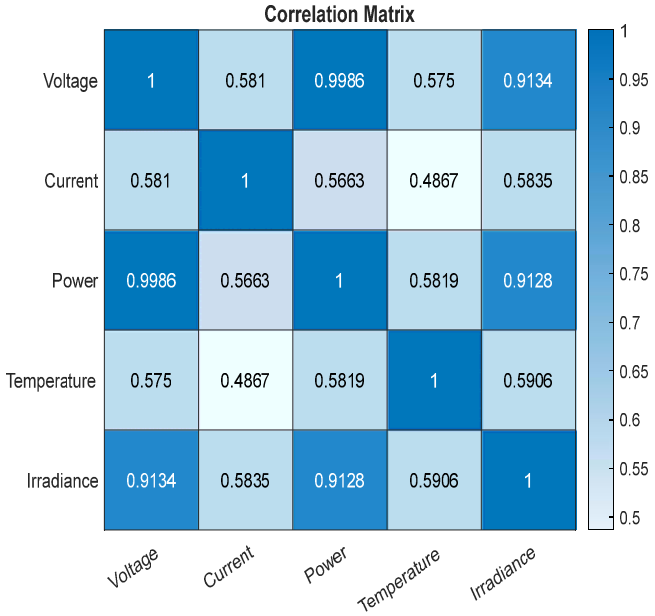


Fig. 1. Correlation matrix of training data.

The box plots in Figure 3 provide a visual summary of the central tendency, variability, and presence of outliers for each parameter. Voltage and power exhibit considerable variability, characterized by wide interquartile ranges and the presence of a few extreme values. In contrast, temperature shows a narrower range, though an extreme outlier is observed below 0°C. These box plots effectively highlight potential outliers, measurement variability, and deviations that may signal system anomalies or faults.

This initial descriptive analysis ensures a thorough understanding of the dataset, allowing for informed implementation of fault detection algorithms and accounting for potential trends and irregularities.

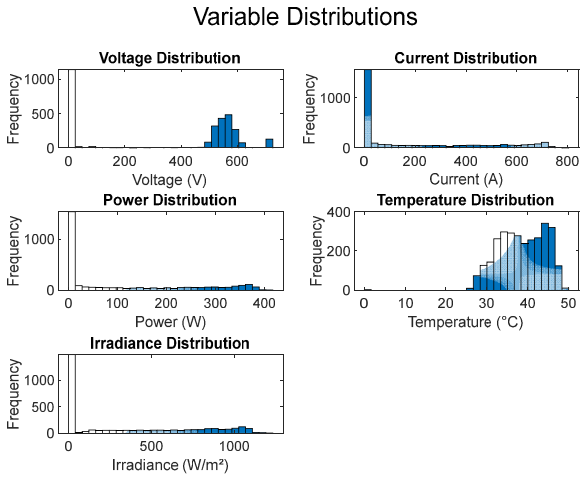


Fig. 2. Distribution plots of training and set data

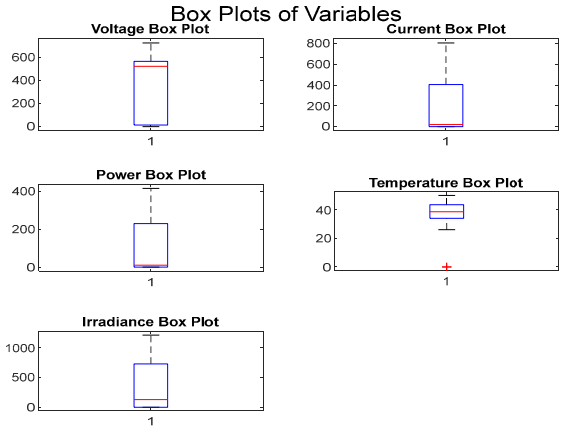


Fig. 3. Box plots of training and set data

IV. Results and Discussion

Fisher's linear discriminant (FLD) and the Mahalanobis distance method for fault detection and diagnosis. I'll examine the key aspects and offer a detailed discussion followed by a potential conclusion.

a) Histogram of Mahalanobis Distance for Training and Testing Data

The histogram of Mahalanobis distances for the training dataset provides valuable insights into the model's ability to differentiate between healthy and faulty operations, as shown in Figure 4.

- *Training Data:*

  For healthy system conditions, Mahalanobis distances are generally lower, reflecting operation within expected parameters. A compact distribution around these lower distance values indicates that most training data points lie close to the center of the distribution for normal operation. Faulty data points, on the other hand, exhibit significantly higher Mahalanobis distances, appearing as outliers in the histogram. This

demonstrates that the model has effectively learned to identify faults by recognizing significant deviations from normal behavior.

- Testing Data:

  The histogram for testing data serves as a validation of the model's generalization capability. Ideally, the healthy data in the testing set should display a distribution similar to that observed during training, confirming the model's ability to recognize normal operation in unseen data. Faulty data in the testing set should show a clear separation, with Mahalanobis distances clustering at higher values, mirroring the training pattern.

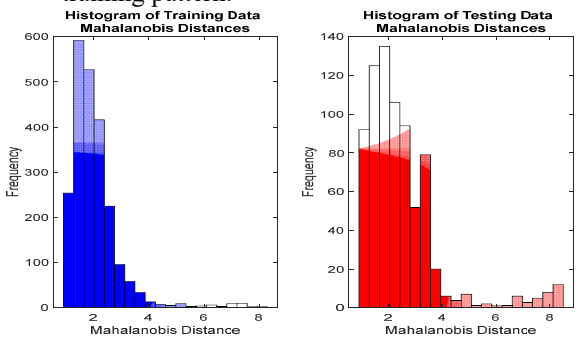


Fig. 4. Histogram of training and testing data mahalanobis distances

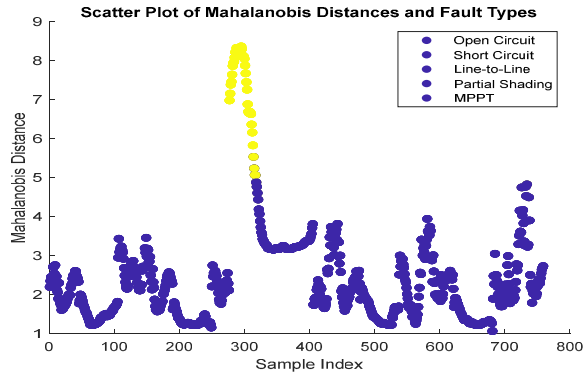


Fig. 5. Scatter plot of mahalanobis distances and fault type

The scatter plot in Figure 5 demonstrates the multivariate approach to fault classification using Mahalanobis distance. Points are distributed across the plot, with their locations indicating how each fault type deviates from normal operating conditions. Data points farther from the healthy cluster signify more severe faults. This visualization effectively illustrates the method's ability to detect faults and differentiate between fault types based on the magnitude of deviation. The scatter plot confirms that the Mahalanobis distance approach is not only suitable for identifying anomalies but also capable of categorizing fault types based on the extent of deviation.

b) Fisher's Linear Discriminant for Fault Detection

Fisher's Linear Discriminant is a robust dimensionality reduction technique that identifies an optimal hyperplane to maximize class separation while minimizing intra-class variance, ensuring that class distinctions remain intact during projection to a lower-dimensional space. This capability makes FLD an effective tool for fault detection in photovoltaic systems. By projecting data onto a single dimension, FLD enhances the separation between healthy and faulty states, significantly improving detection accuracy. As illustrated in Figure 6, FLD achieves a clear division between healthy and faulty data points, demonstrating its reliability and robustness in detecting anomalies, even subtle ones. The method's performance is heavily influenced by the degree of separability between fault and healthy clusters, and the results underscore its precision in classifying different fault types.

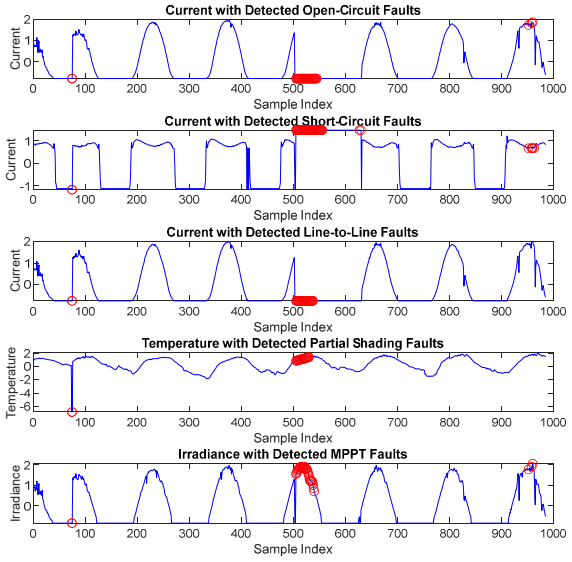


Fig. 6. Current, temperature, irradiance Time series with detected faults

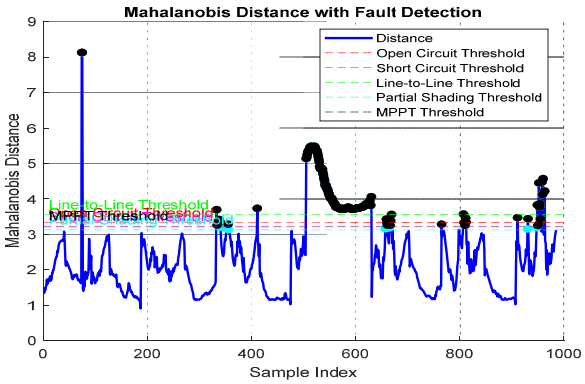


Fig. 7. Mahalanobis distance threshold categories fault detected

Figure 7 demonstrates the application of the Mahalanobis Distance for fault detection in a photovoltaic power generation system. The Mahalanobis distance, plotted on the y-axis, quantifies the multivariate distance of each data sample from a reference distribution, which typically represents the healthy operational state of the PV system. Fault detection is achieved by comparing these distances against predefined fault-specific thresholds. The figure effectively highlights how distinct fault

types can be identified by employing multiple thresholds based on the Mahalanobis distance. For instance, peaks exceeding higher thresholds may indicate severe faults, while smaller peaks suggest less critical anomalies.

Under normal operating conditions, the Mahalanobis distance fluctuates around a stable baseline. However, when the system experiences a fault, the distance increases significantly. When this distance surpasses a predefined threshold, the system triggers an alert, signaling the need for further investigation. This approach enables real-time monitoring and early fault detection, which is critical for maintaining system reliability and performance.

Figure 8 provides a clear and insightful visualization of fault detection using the mahalanobis distance in a PV system. By setting appropriate thresholds for various fault types, the method ensures timely identification and diagnosis of anomalies. The FLD approach further enhances the effectiveness of this method by leveraging multivariate distance metrics to distinguish between different fault types, making it a robust tool for anomaly detection in PV systems. This methodology not only improves fault detection accuracy but also supports proactive maintenance strategies, ultimately enhancing the overall efficiency and longevity of PV systems.

V. Conclusion

The integration of Fisher's Linear Discriminant with the Mahalanobis distance method for fault detection and diagnosis in photovoltaic systems demonstrates exceptional performance. FLD effectively reduces data dimensionality by projecting high-dimensional inputs onto a space that maximizes class separability, simplifying the detection of healthy and faulty states. Complementing this, the Mahalanobis distance method enhances fault diagnosis by leveraging variable correlations to accurately classify fault types, such as open circuits, shading, or short circuits, through its multivariate approach. Together, these techniques form a good strategy, achieving precision, while efficiently handling correlated variables and ensuring practical applicability in real-world PV system monitoring.

VI. References

[1] B. Gong et al., "Fault diagnosis of photovoltaic array with multi-module fusion under hyperparameter optimization," *Energy Convers. Manag.*, vol. 319, 2024.

[2] F. Harrou, A. Saidi, and Y. Sun, "Wind power prediction using bootstrap aggregating trees approach to enabling sustainable wind power integration in a smart grid," *Energy Convers. Manag.*, vol. 201, 2019, doi: 10.1016/j.enconman.2019.112077.

[3] S. Ahmed, C. Benoudjafer, and C. Benachaiba, "MPPT Technique for Standalone Hybrid PV-Wind Using Fuzzy Controller," in *Artificial Intelligence in Renewable Energetic Systems*, M. Hatti, Ed. Springer International Publishing, 2018, pp. 185–196.

[4] K. Ding et al., "Feature extraction and fault diagnosis of photovoltaic array based on current–voltage conversion," *Appl. Energy*, vol. 353, 2024.

[5] F. Harrou, Y. Sun, B. Taghezouit, A. Saidi, and M.-E. Hamlati, "Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches," *Renew. Energy*, vol. 116, 2018, doi: 10.1016/j.renene.2017.09.048.

[6] A. Saidi and B. Chellali, "Simulation and control of Solar Wind hybrid renewable power system," in *2017 6th International Conference on Systems and Control (ICSC)*, 2017, doi: 10.1109/ICoSC.2017.7958647.

[7] A. Saidi, "Solar-Wind hybrid renewable energy systems: Evolutionary Technique," *Electrotehnica, Electronica, Automatica (EEA)*, vol. 64, no. 4, pp. 24–27, 2016.

[8] A. Saidi and C. Benachaiba, "Comparison of IC and P&O algorithms in MPPT for grid connected PV module," in *2016 8th International Conference on Modelling, Identification and Control (ICMIC)*, 2016, doi: 10.1109/ICMIC.2016.7804300.

[9] A. Saidi, B. Cherif, and B. Chellali, "Fuzzy intelligent control for solar/wind hybrid renewable power system," *EEA - Electrotehnica, Electronica, Automatica*, vol. 65, no. 4, 2017.

[10] A. Saidi, A. Harrouz, I. Colak, K. Kayisli, and R. Bayindir, "Performance Enhancement of Hybrid Solar PV-Wind System Based on Fuzzy Power Management Strategy: A Case Study," in *2019 7th International Conference on Smart Grid (icSmartGrid)*, 2019, pp. 126–131, doi: 10.1109/icSmartGrid48354.2019.8990675.

[11] M. P. Suresh et al., "Automatic fault detection and stability management using intelligent hybrid controller," *Electr. Power Syst. Res.*, vol. 238, 2025.

[12] R. Tang et al., "Fault classification of photovoltaic module infrared images based on transfer learning and interpretable convolutional neural network," *Sol. Energy*, vol. 276, 2024.

[13] Z. Hu et al., "Improved multistep ahead photovoltaic power prediction model based on LSTM and self-attention with weather forecast data," *Appl. Energy*, vol. 359, 2024.

[14] M. Jalal et al., "Deep learning approaches for visual faults diagnosis of photovoltaic systems: State-of-the-Art review," *Results Eng.*, vol. 23, 2024.

[15] A. Keddouda et al., "Photovoltaic module temperature prediction using various machine learning algorithms: Performance evaluation," *Appl. Energy*, vol. 363, 2024.

[16] D. Li et al., "Sensing anomaly of photovoltaic systems with sequential conditional variational autoencoder," *Appl. Energy*, vol. 353, 2024.

[17] Q. Qi et al., "Development assessment of regional rooftop photovoltaics based on remote sensing and deep learning," *Appl. Energy*, vol. 375, 2024.

[18] R. Banik and A. Biswas, "Improving Solar PV Prediction Performance with RF-CatBoost Ensemble: A Robust and Complementary Approach," *Renew. Energy Focus*, 2023, doi: 10.1016/j.ref.2023.06.009.

[19] F. Harrou, A. Saidi, Y. Sun, and S. Khadraoui, "Monitoring of Photovoltaic Systems Using Improved Kernel-Based Learning Schemes," *IEEE J. Photovoltaics*, vol. 11, no. 3, pp. 806–818, May 2021, doi: 10.1109/JPHOTOV.2021.3057169.

[20] T. Peng et al., "An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting," *Energy*, vol. 221, 2021.

[21] M. Zhang et al., "Ultra-short-term photovoltaic power prediction based on similar day clustering and temporal convolutional network with bidirectional long short-term memory model: A case study using DKASC data," *Appl. Energy*, vol. 375, 2024.

[22] J. Zhang et al., "Fast object detection of anomaly photovoltaic (PV) cells using deep neural networks," *Appl. Energy*, vol. 372, 2024.

[23] L. Zaghba et al., "Investigating the theoretical and experimental effects of sand dust and sandstorms on photovoltaic power plants in arid environments," *Energy Sustain. Dev.*, vol. 81, 2024.

[24] Z. Mian et al., "A literature review of fault diagnosis based on ensemble learning," *Eng. Appl. Artif. Intell.*, vol. 127, 2024.

[25] A. Javaid et al., "Enhancing photovoltaic systems using Gaussian process regression for parameter identification and fault detection," *Energy Rep.*, vol. 11, pp. 4485–4499, 2024.

Enhancement of Sensorless Direct Torque Controlled Permanent Magnet Synchronous Motors (PMSM)

D. Kouchih[1*], Y. Diboune[1], R. Hachelaf [1,2], M. Tadjine [2], M.S Boucherit [2]

[1] *Automatic and Electrotechnic Department, University Saad Dahlab, Blida, Algeria*

[2] *Automatic Control Department, National Polytechnic School, Alger, Algeria*

**e-mail: djkouchih@yahoo.fr*

***Abstract*— The purpose of this work is the development of an adaptive observer in order to enhance the stability of the direct torque controlled permanent magnet synchronous motors. This observer makes it possible to reconstruct the mechanical speed and state variables using stator voltages and currents. Lyapunov theory has been used for the synthesis of this adaptive observer in order to guarantee the closed-loop stability. It is well that the main drawback of the direct torque control is its sensitivity to parametric variation, especially to stator resistance at low speed operation. For this reason, the sliding mode control has been used for the speed correction in order to guarantee control robustness against parametric variation. The undesirable chattering phenomena can be remedied using a smooth continuous function. The contribution of this work is that a new and simple algorithm has been obtained. It has the advantage to be easily implemented in calculators. Simulation tests are provided to evaluate the consistency of this observer at low and high speeds under disturbances. Simulation tests show that the control scheme introduces high performances of robustness, stability and precision under disturbances caused by parametric variation.**

Key Words*— Adaptive observer, direct torque control, permanent magnet synchronous motors, sensorless.

I. INTRODUCTION

Recently, the direct torque control (DTC) is one of the most widely used techniques for the speed control of Permanent Magnet Synchronous Motors (PMSM) [1-5]. This technique requires an accurate knowledge of the controlled quantities such as stator flux and electromagnetic torque. These quantities are conventionally calculated using an estimator operating in open loop. This estimation suffers from the well-known problem of the sensitivity to the parametric variation especially the stator resistance at low speed operation [6-9]. To remedy for this inconvenience and in order to avoid the use of the mechanical sensor for the speed, we have to elaborate an adaptive observer using Lyapunov theory to guarantee the stability of the DTC controlled PMSM. Several approaches have been developed for the sensorless control for PMSM drives. The first approach is based on the Model Reference Adaptive System (MRAS). In this context, Joshi, V. and al investigate on the Modeling and analysis of MRAS based speed sensorless control for PMSM [10]. Demir, R proposes an adaptive filtering based MRAS speed estimators for the speed-sensorless predictive current controlled PMSM drive [11]. In [12], the authors investigate on the sensorless real-time implementation of PMSM using MRAS and T-S fuzzy speed controller. Said, M.A.A. and al investigate on the effects of using PI Fuzzy logic controller in MRAS Model for sensorless control of PMSM [13]. The MRAS approach is very simple but its disadvantage is the sensitivity to parametric variation. On the other hand, Extended Kalman Filter combined to variable structure techniques are also used by many researchers. For this purpose, Ding, H. and al propose an adaptive robust Kalman Filter for sensorless control of PMSM [14]. In [15], the authors propose Extended Kalman Filter for high performance control of sensorless PMSM drive. Zhang, H.-W. and al investigate on dual Extended Kalman Filter for sensorless DTC of PMSM. Recently, a new approach, based on artificial intelligence techniques (fuzzy logic and neural networks), has been proposed in the literature. Khizhnyakov, Y.N. and al propose the use of fuzzy logic for sensorless control of a synchronous motor [17]. In [18], the authors use fuzzy sliding mode controller and fuzzy sliding mode observer for sensorless control strategy of PMSM. Mukherjee, D. and al investigate on the Application of Machine Learning for speed and torque prediction of PMSM in electric vehicles [19]. In [20], the authors propose artificial neural network for torque ripple minimization in speed sensorless control of direct torque controlled PMSM. Artificial intelligent methods have serious drawbacks that they require a high computation time and are complex to be implemented.

It should be noted that the main disadvantage of the DTC is its sensitivity to parametric variation, especially to stator resistance variation. For this reason, we use the sliding mode control (SMC) for the speed correction in order to guarantee control robustness. The SMC offers good properties of robustness under internal and external disturbances. It has been largely discussed in the literature [21-25]. The undesirable chattering phenomena can be remedied using a smooth continuous function [25]. The contribution of this work is that a new algorithm is developed to reconstruct the controlled quantities and the mechanical speed using Lyapunov theory. A simple algorithm has been obtained; it has the advantage to be easily implemented in calculators. Simulation tests show that the control scheme introduces high performances of stability and precision at low and high speeds.

II. DTC STRATEGY

The block diagram of the DTC is shown by Figure 1. We use a two-level hysteresis comparator to keep the end of the stator flux vector in a specified band. A logical variable $(\delta = 1)$ or $(\delta = 0)$ indicates whether the amplitude of the stator flux must be increased or decreased in order to maintain:

$$\left|\Phi_S^* - \Phi_s\right| \le \Delta\Phi_S \qquad (1)$$

Φ_S^* is the reference stator flux

$\Delta\Phi_S$ is the stator flux hysteresis band

Three-level comparator has been used for the correction of the electromagnetic torque in order to control the motor in both senses of rotation. For this a logical variable indicates if the torque must be increased $(\mu = 1)$, to maintain its constant $(\mu = 0)$ and $(\mu = -1)$ to reduce the torque. The optimal switching logic defines the best stator voltage which will be applied to the motor. The control sequences of the inverter are illustrated by Table 1.

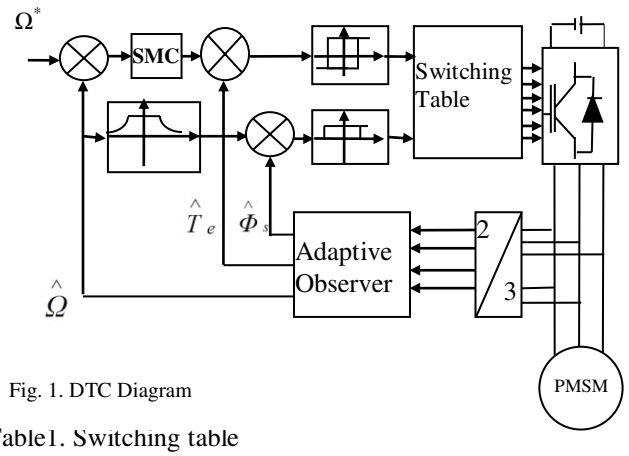


Fig. 1. DTC Diagram

Table1. Switching table

Sectors		1	2	3	4	5	6
$\delta = 1$	$\mu = +1$	110	010	011	001	101	100
	$\mu = 0$	111	000	111	000	111	000
	$\mu = -1$	101	100	110	010	011	001
$\delta = 0$	$\mu = +1$	010	011	001	101	100	110
	$\mu = 0$	000	111	000	111	000	111
	$\mu = -1$	001	101	100	110	010	011

It should be noted that in the direct and quadrature (d-q) reference frame referred to the rotor position, the state model of the PMSM is expressed by:

$$\begin{cases} \dfrac{di_{ds}}{dt} = -\dfrac{R_s}{L_d} i_{\alpha s} + \omega \dfrac{L_q}{L_d} i_{qs} + \dfrac{1}{L_d} v_{ds} \\ \dfrac{di_{qs}}{dt} = -\omega \dfrac{L_d}{L_q} i_{ds} - \dfrac{R_s}{L_q} i_{ds} - \dfrac{\omega}{L_d} \phi_0 + \dfrac{1}{L_q} v_{qs} \\ J \dfrac{d\Omega}{dt} = T_e - T_l - f_v \Omega \end{cases} \qquad (2)$$

Where v_{ds}, v_{qs} are the components of stator voltage vector in the d-q reference frame, i_{ds}, i_{qs} are the components of stator current vector, R_s is the stator and rotor resistance, L_d and L_q are respectively the direct and quadrature inductances, ω is the mechanical pulsation and ϕ_0 is the flux produced by the permanent magnet. J is the inertia of the rotor; T_l is the load torque and f_v is the viscose friction coefficient.

The electromagnetic torque is epressed:

$$T_e = \frac{3}{2} P\left[\Phi_0 i_{qs} + (L_d - L_q) i_{ds} i_{qs}\right] \qquad (3)$$

III. ADAPTATION MECHANISM

The adaptive observer has the following structure:

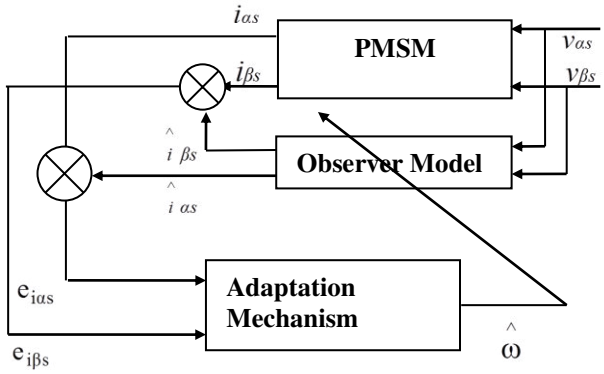


Fig. 2. Global adaptive observer

The objective is to develop the mechanism for adapting the mechanical speed. The PMSM parameters are considered to be perfectly known, the mechanical pulsation is unknown. The previous state model is expressed in matrix form:

$$\begin{cases} \dfrac{dX}{dt} = AX + BU \\ Y = CX \end{cases} \qquad (4)$$

With:

$$A = \begin{pmatrix} -\dfrac{R_s}{L_d} & \omega \dfrac{L_q}{L_d} \\ -\omega \dfrac{L_d}{L_q} & -\dfrac{R_s}{L_q} \end{pmatrix} \qquad (5)$$

$$B = \begin{pmatrix} \dfrac{1}{L_q} & 0 & 0 \\ 0 & \dfrac{1}{L_q} & -\dfrac{\omega}{L_d} \end{pmatrix}, \qquad (6)$$

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (7)$$

$X = \begin{pmatrix} i_{ds} \\ i_{qs} \end{pmatrix}$ is the state vector

$Y = \begin{pmatrix} i_{ds} \\ i_{qs} \end{pmatrix}$ is the output vector

$U = \begin{pmatrix} v_{ds} \\ v_{qs} \\ \phi_0 \end{pmatrix}$ is the input vector

Using Luenberger Observer, the sate model of the adaptive observer is expressed as follows [25]:

$$\begin{cases} \dfrac{d\hat{X}}{dt} = \hat{A}\hat{X} + BU + G\left(Y - \hat{Y}\right) \\ \hat{Y} = C\hat{X} \end{cases} \quad (8)$$

The symbol $\wedge$ denotes observed values and G is the observer gain matrix.

We define the error on unknown parameter as follows:

$$\delta\omega = \omega - \hat{\omega} \quad (9)$$

Using these differences, the state matrix of the adaptive observer is epressed:

$$\hat{A} = A + \delta A \quad (10)$$

With:

$$\delta A = \begin{pmatrix} 0 & -\dfrac{L_q}{L_d}\delta\omega \\ \dfrac{L_d}{L_q}\delta\omega & 0 \end{pmatrix} \quad (11)$$

Using the state error

$$e = X - \hat{X} \quad (12)$$

The dynamic of the observed variables becomes:

$$\frac{d\hat{X}}{dt} = \hat{A}\hat{X} + BU + GCe \quad (13)$$

From (4) and (13), the error dynamic is expressed:

$$\frac{de}{dt} = AX - \hat{A}\hat{X} + GCe \quad (14)$$

Thus:

$$\frac{de}{dt} = (A - GC)e - \delta A\hat{X} - \delta BU \quad (15)$$

We define the Lyapunov function:

$$V = e^T e + \frac{(\delta\omega)^2}{\lambda} \quad (16)$$

λ is a positive scalar. The stability is guaranteed for:

$$\frac{dV}{dt} < 0 \quad (17)$$

It can be verified that:

$$\frac{dV}{dt} = 2e^T\frac{de}{dt} + 2\frac{\delta\omega}{\lambda}\frac{d\delta\omega}{dt} \quad (18)$$

The first term of (18) is expressed:

$$2e^T\frac{de}{dt} = 2e^T(A - GC)e - 2e^T\delta A\hat{X} \quad (19)$$

Thus:

$$e^T\delta A\hat{X} = \delta\omega\left(\frac{L_q}{L_d}\hat{i}_{qs}e_{ids} - \left(\frac{L_d}{L_q}\hat{i}_{ds} + \frac{\phi_0}{L_q}\right)e_{iqs}\right) \quad (20)$$

For the second term of (18), we get:

$$2\frac{\delta\omega}{\lambda}\frac{d\delta\omega}{dt} = 2\frac{\delta\omega}{\lambda_1}\frac{d}{dt}\hat{\omega} - 2\frac{\delta\omega}{\lambda_1}\frac{d}{dt}\omega \quad (21)$$

Considering the fact that the machine speed changes slowly, we get:

$$\frac{d\hat{\omega}}{dt} = -\frac{d\delta\omega}{dt} \quad (22)$$

As a result, the derivative of Lyapunov function becomes:

$$\frac{dV}{dt} = 2e^T(A - GC)e - 2\delta\omega\left(\frac{L_q}{L_d}\hat{i}_{qs}e_{ids} - \left(\frac{L_d}{L_q}\hat{i}_{ds} + \frac{\phi_0}{L_q}\right)e_{iqs}\right) + 2\frac{\delta\omega}{\lambda_1}\frac{d}{dt}\hat{\omega} \quad (23)$$

If the derivative $\frac{dV}{dt} = 2e^T(A - GC)e$ is negative, the condition $\frac{dV}{dt} < 0$ is verified for:

$$-2\delta\omega\left(\frac{L_q}{L_d}\hat{i}_{qs}e_{ids} - \left(\frac{L_d}{L_q}\hat{i}_{ds} + \frac{\phi_0}{L_q}\right)e_{iqs}\right) + 2\frac{\delta\omega}{\lambda_1}\frac{d}{dt}\hat{\omega} = 0 \quad (24)$$

This condition is guaranteed for:

$$2\frac{\delta\omega}{\lambda}\frac{d}{dt}\hat{\omega} = .2\delta\omega\left(\frac{L_q}{L_d}\hat{i}_{qs}e_{ids} - \left(\frac{L_d}{L_q}\hat{i}_{ds} + \frac{\phi_0}{L_q}\right)e_{iqs}\right) \quad (25)$$

Finally, the adaptation mechanism is expressed by

$$\hat{\omega} = \int\lambda\left(\frac{L_q}{L_d}\hat{i}_{qs}e_{ids} - \left(\frac{L_d}{L_q}\hat{i}_{ds} + \frac{\phi_0}{L_q}\right)e_{iqs}\right)dt \quad (26)$$

The gain matrix G is calculated so that the eigenvalues of the matrix $A - GC$ are in the left of the complex plan and the real part of the eigenvalues is larger in absolute value than the real part of the eigenvalues of the state matrix A [25].

The electromagnetic torque is estimated using

$$\hat{T}_e = \frac{3}{2}P\left[\Phi_0\hat{i}_{qs} + (L_d - L_q)\hat{i}_{ds}\hat{i}_{qs}\right] \quad (27)$$

IV. SIMULATION RESULTS

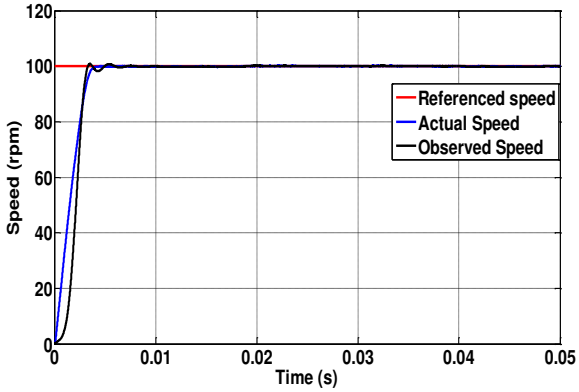
Simulation results have been obtained using MATLAB for low and high speeds. The PMSM parameters are: *3 kW*, $220/380\,V - 50\ Hz$, $\Phi_0 = 0.1546\,Wb$ $R_s = 1.4\,\Omega$

$L_d = L_q = 6.6\,mH$, $L_d = L_q = 5.8\,mH$, $J = 0.00176\ kgm^2$

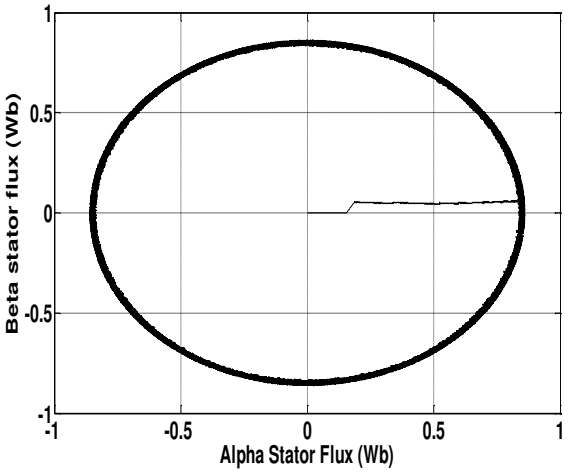
$f_v = 0.00038818\ Nms/rd$.

A. Low Speeds

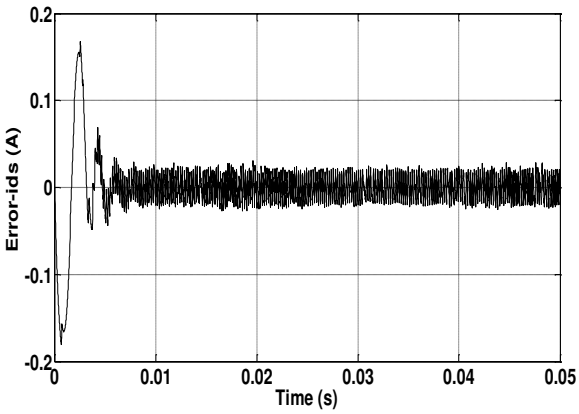
Simulation test has been provided for starting up with nominal load. The hysteresis bands chosen for torque and flux are **0.2 Nm** and **0.01 Wb** respectively. We considered the low speed operation **+100 rpm**. The simulation results are illustrated by Figure 3.



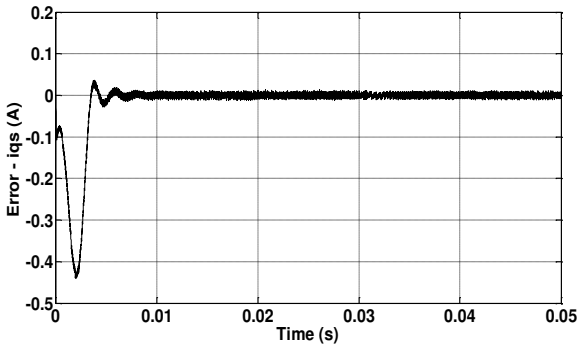
(a)



(b)



(c)

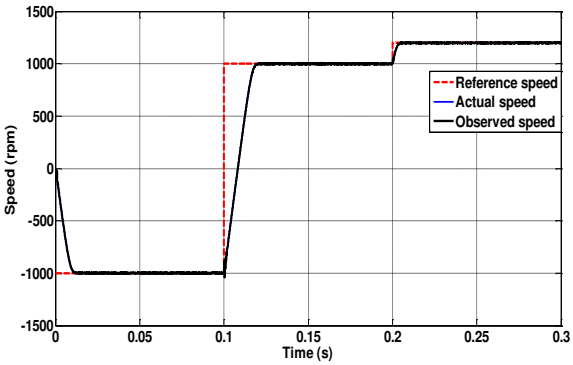


(d)

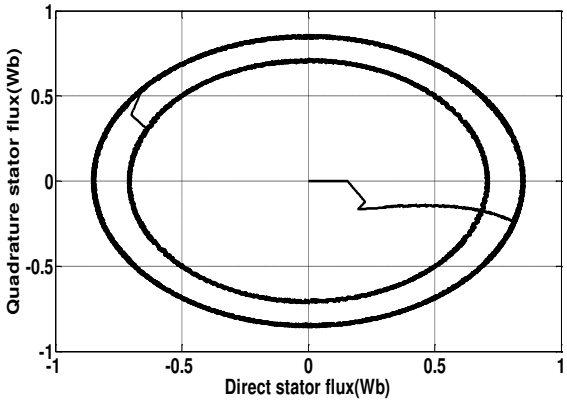
Fig. 3 Simulation results for low speed operation: (a) mechanical speed, (b) Hodograph of the stator flux, (c) error of the direct current i_{ds}, and (d) error of the quadrature current i_{qs}.

B. High Speeds

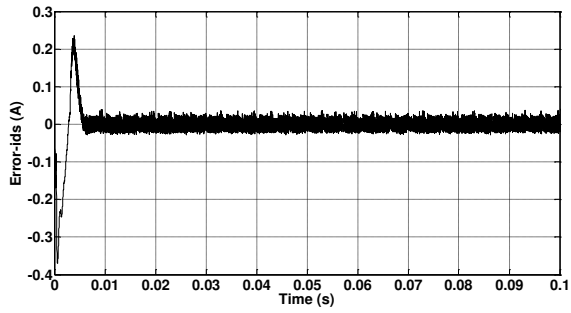
High speed operation is characterized by the weakening of the reference of stator flux. We consider a start up with a nominal load and a reference speed of 1000 rpm; then at t = 0.5 sec, we select the high speed of 1200 rpm. Simulation results are illustrated by Fig. 4.



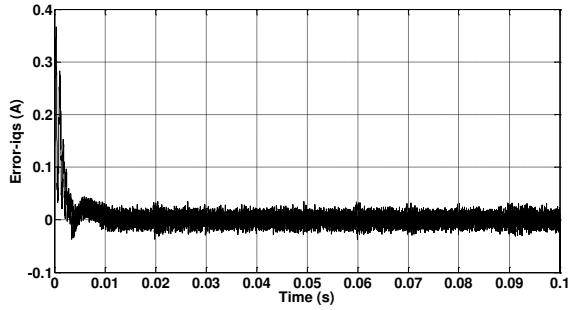
(a)



(b)



(c)



(d)

Fig. 4 Simulation results for high speed operation: (a) mechanical speed, (b) Hodograph of the stator flux, (c) error of the direct current i_{ds}, and (d) error of the quadrature current i_{qs}.

The state variables and mechanical speed are reconstructed for high speeds operation. The controlled quantities have not been disturbed by the stator resistance variation. Consequently, high performances of robustness, stability and precision are obtained for high speed operation under stator resistance variation.

V. CONCLUSIONS

New approach has been proposed for the design of an adaptive observer in order to enhance the stability of the direct torque controlled PMSM. The adaptive observer is based on Lyapunov theory to reconstruct the state variables and mechanical speed. The proposed observer is characterized by a simple algorithm which has the advantage to be easily implemented. High performances of robustness stability and precision are obtained for low and high speeds. This work will be performed by experiments in order to validate the theoretical study and to verify the consistency of this approach for the improvement of the direct torque controlled PMSMs.

REFERENCES

1. C. Youssef, Z. Said, D. Abdelkarim, "A Review of Non-linear Control Methods for Permanent Magnet Synchronous Machines (PMSMs) ," In: Farhaoui, Y., Hussain, A., Saba, T., Taherdoost, H., Verma, A. (eds) Artificial Intelligence, Data Science and Applications. ICAISE 2023. Lecture Notes in Networks and Systems, vol 837. Springer, Cham, (2024). https://doi.org/10.1007/978-3-031-48465-0_59.
2. O.E. Özçiflikçi, M. Koç, S. Bahçeci, *et al*, "Overview of PMSM control strategies in electric vehicles: a review," *Int. J. Dynam. Control,* (2023). https://doi.org/10.1007/s40435-023-01314-2.
3. Q. Wu, "Simulation Research on direct torque control of permanent magnet synchronous motor based on SVPWM," 2022 World Automation Congress, San Antonio, TX, USA, 11-15 October, (2022). https://doi.org/10.23919/WAC55640.2022.9934623.
4. R.E.K. Meesala, S. Athikkal, R. Aruldavid, "Improved Direct Torque Controlled PMSM Drive for Electric Vehicles," *J. Inst. Eng. India Ser. B* 103, 1177–1188 (2022). https://doi.org/10.1007/s40031-022-00716-8.
5. Z. Machhour, M.E. Mrabet, Z. Mekrini, M. Boulaala, "Electric Motors and Control Strategies for Electric Vehicles: A Review. In: Kacprzyk, J., Ezziyyani, M., Balas, V.E. (eds) International Conference on Advanced Intelligent Systems for Sustainable Development. AI2SD 2022. Lecture Notes in Networks and Systems, vol 714. Springer, Cham, (2023). https://doi.org/10.1007/978-3-031-35245-4_27.
6. Z. Wang, Y. Inoue, S. Morimoto, M. Sanada, T. Aizawa, T. Takagi, "A Comparative Study on Operating Characteristics of PMSM Based on Direct Torque Control in Low Speed Region, 23rd International Conference on Electrical Machines and Systems, Hamamatsu, Japan, 24-27 November, (2020). https://doi.org/10.23919/ICEMS50442.2020.9291117.
7. H.-Y. Chung Y.-L. Chung, Y.-H. Jhu, "Using Flux Compensator to Implement Direct Torque Control of Permanent Magnet Synchronous Motor at Low Speed Improvement," 2022 International Conf. on System Science and Eng. Taichung, Taiwan, 26-29 May, (2022). https://doi.org/10.1109/ICSSE55923.2022.9948228
8. R. Pothuraju, R. Tejavathu, D.K. Poondla, "Stator resistance and speed estimation for five level inverter fed DTFC-SVM of speed sensorless PMSM drive,"*Electr Eng* 104, 4415–4432 (2022). https://doi.org/10.1007/s00202-022-01626-0
9. D.H. Cho, K.B. Lee, "Sensorless DTC for Interior Permanent-Magnet Synchronous Motors Using Square-Wave-Type Stator Flux Injection at Low-Speed Regions,"*J. Electr. Eng. Technol.* 17, 329–337, (2022). https://doi.org/10.1007/s42835-021-00866-7

10. V. Joshi, S. Jeevanand, H. Mohan, "Modeling and Analysis of MRAS based Speed Sensorless Control for PMSM Drive," 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering, Dehradun, India, 11-13 November, (2021). https://doi.org/10.1109/UPCON52273.2021.9667623.

11. R. Demir, "Speed-sensorless Predictive Current Controlled PMSM Drive With Adaptive Filtering-based MRAS Speed Estimators," *Int. J. Control Autom. Syst.* 21, 2577–2586 (2023). https://doi.org/10.1007/s12555-022-0698-z
12. M. Benakcha, A. Benakcha, R. Abdessemed, *et al*, "Sensorless real-time implementation-based FS-MPCC and deadbeat predictive control with delay and dead-time compensation of PMSM using MRAS and T-S fuzzy speed controller," *Electr Eng* 105, 4139–4156, (2023). https://doi.org/10.1007/s00202-023-01939-8.
13. M.A.A. Said, M.Sreejeth, A.B.K. Bhattacharya, "Sensorless control of PMSM using Direct and Model References Adaptive system estimation techniques and observing the effect of using PI Fuzzy logic controller in MRAS Model," 2020 IEEE International Conference for Innovation in Technology. Bangluru, India, 06-08 November, (2020). https://doi.org/10.1109/INOCON50539.2020.9298309.
14. H. Ding, X. Qin, L. Wei, "Sensorless Control of Surface-Mounted Permanent Magnet Synchronous Motor Using Adaptive Robust UKF," *J. Electr. Eng. Technol.* 17, 2995–3013, (2022). https://doi.org/10.1007/s42835-022-01061-y.
15. G. R., G., Pudutha, M.B., "High Performance Control and Extended Kalman Filter Based Estimation of Sensorless Permanent Magnet Synchronous Motor Drive for Robotic Applications," In: Duffy, V.G., Krömker, H., A. Streitz, N., Konomi, S. (eds) HCI International 2023 – Late Breaking Papers. HCII 2023. Lecture Notes in Computer Science, vol 14057. Springer, Cham, (2023). https://doi.org/10.1007/978-3-031-48047-8_33.
16. H.-W. Zhang, D. Jiang, X.-H. Wang, M.-R. Wang, "Direct Torque Sensorless Control of PMSM Based on Dual Extended Kalman Filter," 33rd Chinese Control and Decision Conf., Kunming, China, May 22-24, (2021). https://doi.org/10.1109/CCDC52312.2021.9601516.
17. Khizhnyakov, Y.N., Yuzhakov, A.A., Storozhev, S.A. *et al*, "Sensorless Control of a Synchronous Motor Based on Fuzzy Logic," *Russ. Electr. Engin.* 94, 814–819, (2023). https://doi.org/10.3103/S1068371223110044.
18. X. Zou, H. Ding, J. Li, "Sensorless Control Strategy of Permanent Magnet Synchronous Motor Based on Fuzzy Sliding Mode Controller and Fuzzy Sliding Mode Observer," *J. Electr. Eng. Technol.* 18, 2355–2369,(2023). https://doi.org/10.1007/s42835-022-01352-4
19. D. Mukherjee, S. Chakraborty, P.K. Guchhait, J.Bhunia, "Application of Machine Learning for Speed and Torque Prediction of PMS Motor in Electric Vehicles," 2020 IEEE 1st International Conference for Convergence in Engineering, Kolkata, India, 05-06 September, (2020). https://doi.org/10.1109/ICCE50343.2020.9290632.

20. S.K Kakodia, D. Giribabu, R.K. Ravula, "Torque Ripple Minimization using an Artificial Neural Network based Speed Sensor less control of SVM-DTC fed PMSM Drive," 2022 IEEE Texas Power and Energy Conference,,College Station, TX, USA, 28 February - 01 March, (2022). https://doi.org/10.1109/TPEC54980.2022.9750850.

21. H. He, J. Gao, Q. Wang, *et al.*, "Improved Sliding Mode Observer for the Sensorless Control of Permanent Magnet Synchronous Motor," *J. Electr. Eng. Technol,* (2024). https://doi.org/10.1007/s42835-024-01785-z
22. X. Shizhou, J. Xinxin, F. Jingsheng, C. Jinhai, "Sensorless Control of Permanent Magnet Synchronous Motor Based on Adaptive Sliding Mode Observer," In: Cai, C., Qu, X., Mai, R., Zhang, P., Chai, W., Wu, S. (eds) The Proceedings of 2023 International Conference on Wireless Power Transfer (ICWPT2023). ICWPT 2023. Lecture Notes in Electrical Engineering, vol 1159. Springer, Singapore, (2024) https://doi.org/10.1007/978-981-97-0877-2_26.
23. S. Huang, Q. Zhang, "Stator Flux Sliding Mode Observer of Permanent Magnet Synchronous Motor Based on Effective Flux, 4th International Conference on Intelligent Control, Measurement and Signal Processing," Hangzhou, China, 08-10 July (2022). https://doi.org/10.1109/ICMSP55950.2022.9859031.
24. J.J.E. Slotine and W. Li, "Applied Nonlinear Control," Prentice Hall Inc., Englewood Cliffs NJ, 1991.
25. C. Edwards and S.K. Spurgeon, "Sliding mode control, theory and applications," The Taylor & Francis systems and control book series, 1998.

A Methodology for Solving VAr/Voltage Problems in distribution system connected wind farm based on IG model

ABDELOUAHED Touhami,
Department of Electrical Systems Engineering
University M'Hamed Bougara of Boumerdes (UMBB), Algeria
Smart Grids and Renewable Energies Laboratory(SGRE), University of, Bechar, Algeria
Email:a.touhami@univ-boumerdes.dz

BEGHDADI Azeddine
Smart Grids and Renewable Energies Laboratory (SGRE), University of Tahri Mohamed, Bechar, Algeria
Email:azeddine.beghdadi@gmail.com

SAIDI Ahmed
Department of Electrical Engineering
Sustainable Development and Computer Science Laboratory (LDDI), Ahmed Draia-University ADRAR
Email:ahmedsaidi@outlook.com

Bousserhane Ismail Khalil
Department of Electrical Engineering
Smart Grids and Renewable Energies Laboratory (SGRE), University of Tahri Mohamed, Bechar, Algeria
Email: bou_isma@yahoo.fr

BOUGHAZI Othmane
Department of Electrical Engineering
Smart Grids and Renewable Energies Laboratory (SGRE), University of Tahri Mohamed, Bechar, Algeria
Email: boughazi.othmane@univ-bechar.dz

DENNAI Mohammed Yassine
Smart Grids and Renewable Energies Laboratory (SGRE), University of Tahri Mohamed, Bechar, Algeria
Email: dennaimoh@yahoo.fr

***Abstract*— This paper describes the methodology adopted to deal with the VAr and voltage violation problems in the wind energy integration application. The methodology is based on the use of powers electronics: shunt, series or hybrid controllers as a powerful tool in solving the mentioned problems. Four (04) important devices controllers from FACTS technology are proposed for our study which are: The Static VAR Compensator (SVC), the Static series synchronous compensator (SSSC), the Synchronous Compensator (STATCOM), and the Unified Power Flow Controller (UPFC), the selected of these controllers is based on its efficiently and large utilities in the industry. For the validation capability of the proposed methodology for solving the power quality problems and improving the wind farm stability under faults conditions, a model of wind farm (9 MW) based on squirrel-cage induction generators (IG) integrated with distribution system (25-kV) is implemented in MATLAB/SIMULINK platform. Simulation results will be presented, which show that the relay system has different response between the four devices: SVC, SSSC, STATCOM and UPFC, which the UPFC can be considered more effective and efficient than: SSSC, STATCOM and SVC since it combines the features from both STATCOM and SSSC.**

Keywords*— *power quality, voltage violation, Wind farm, STATCOM, SVC, SSSC, UPFC.

I. INTRODUCTION

In recent years, wind energy has become one of the most important and promising sources of renewable energy, which demands additional transmission capacity and better means of maintaining system reliability. The need to integrate the renewable energy like wind energy into power system is to make it possible to minimize the environmental impacts[1-2]. However, wind power may cause problems to the existing grid in terms of power quality. One of the fundamental definitions to power quality problems is any power problem manifested in voltage, current or frequency deviations that results in failure or miss operation of customer equipment [3].On the other hand, the development of power electronics and microelectronics makes it possible to consider active power filters, which can provide flexible current harmonic compensation and contribute to reactive power control and load balancing [4]. The application of FACTS devices to power system security has been an attractive ongoing area of research. In most of the reported studies attention has been focused on the ability of these devices to improve the power system security by damping system oscillations and minimal attempts have been made to investigate the effect of these devices on power system reliability. The opportunities arise through the ability of FACTS controllers to control the interrelated parameters that governs the operation of transmission systems including series impedance and shunt impedance, current, phase angle and damping of oscillations at various frequencies below the rated frequency [1]. In this paper, it is suggested to use the FACT devices for grid connected wind farm system to improve the stability in wind farm connected to power system. In this regard, this paper proposes the use of either the Static series synchronous compensator SSSC or the static synchronous compensator STATCOM or the Unified Power Flow Controller (UPFC) to improve stability of wind farm that is connected to power system. So the scientific contributions of this paper are:

Firstly, stability analysis of IG based on wind turbine is explained. Furthermore, the wind farm model based on IG,

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

equipped with the controllers: SSSC, STATCOM and UPFC, connected to power system is developed using MATLAB-SIMULINK. Then the impact of all the mentioned controllers on power system during and after fault on the system recovery are investigated.

II. REVIEW OF WIND TURBINE SYSTEMS

The principle of kinetic transformation energy of the wind into electric power and the detailed description of the various types of aero-generators are presented in several references. The mechanical power which can be extracted from the wind determines by means of the following expression [3,4]:

$$P = \frac{1}{2}\rho S V^3 C_p(\lambda.\beta) \quad (1)$$

$$C_p(\lambda.\beta) = \frac{1}{2}(\Gamma - 0.022\beta^2 - 5.6)e^{-0.17\Gamma} \quad (2)$$

$$\Gamma = \frac{r.(3600)}{\lambda.(1609)} \quad (3)$$

Where, P is the extracted power from the wind, ρ is the air density, $S = \pi r^2$ is the surface swept by the turbine, the v wind speed, β is blade pitch angle and Cp the power coefficient. This coefficient corresponding to the aerodynamic efficiency of the turbine has a nonlinear evolution according to the tip speed ratio, λ as indicated in Table I [5].

TABLE I. VALUES OF THE COEFFICIENTS CP AND Λ OF WIND

λ	0	1.3	2.1	2.7	3.4	5	5.8	6.3
Cp	0	0.1	0.2	0.3	0.4	0.3	0.2	0.1

Where: $\lambda = \frac{w.r}{V}$ (4)

r: is the blade length and w is the angular velocity of the turbine.

III. INDUCTION GENERATOR MODEL

Wind turbines use squirrel cage induction generators are shown in Fig. 1. The stator winding is connected directly to the grid and the rotor driven by the wind turbine. The power captured by the wind turbine is converted into electrical power by the induction generator and is transmitted to the grid by the stator winding. The pitch angle is controlled in order to limit the generator output power to its nominal value for high wind speeds. In order to generate power the induction speed must be slightly above the synchronous speed but the speed variation is typically so small that the WTIG is considered to be affixed speed wind generator [3,4,6]:

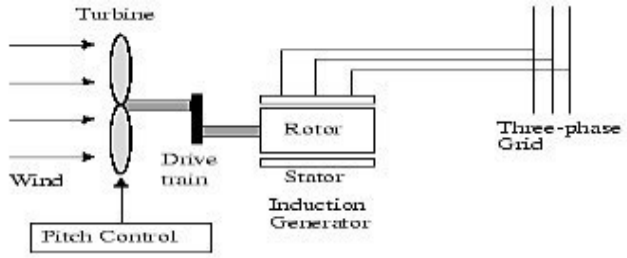


Fig. 1- Wind turbine and induction generator (IG)

Figure 2 shows wind turbine characteristic used for this study with the turbine input power plotted against the rotor speed of the turbine. The turbine mechanical power as function of turbine speed is displayed for wind speeds ranging from 4 m/s to 10 m/s.

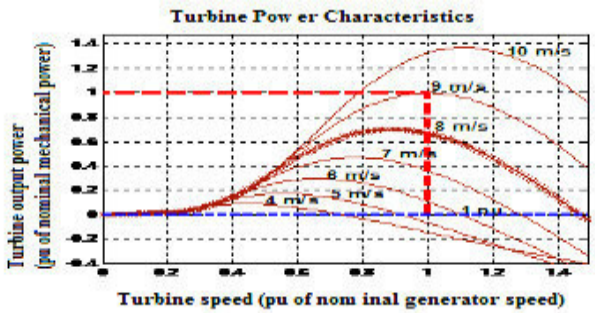


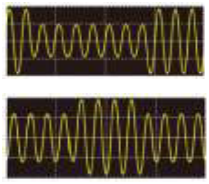
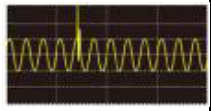
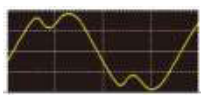
Fig. 2- Turbine characteristic with maximum power point tracking for IG

IV. POWER QUALITY PROBLEMES, ISSUES AND ITS CONSEQUENCES

As we mentioned in the abstract, the integration of wind power to an electric grid may cause problems important in terms of power quality such as: the active power, reactive power, variation of voltage, flicker, harmonics, and electrical behavior of switching operation and these are measured according to national/international guidelines.

In the following table, some of these phenomena like: voltage sag, swell and harmonic distortion will be summarized [6,7].

TABLE II. POWER QUALITY PROBLEMS

Power quality problems	**Definition, causes and consequences**
Power outages: 	***Power outages*** are total interruptions of electrical supply. ***Causes:*** lightning, wind, utility equipment failure. ***Effects:*** Complete disruption of operation
Voltage fluctuations: 	***Voltage fluctuations*** are changes or swings in the steady-state voltage above or below (sags and swells) the designated input range for a piece of equipment. ***Causes:*** Large equipment start-up or shut down; sudden change in load. ***Effects:*** equipment shut down, flickering lights; motors stalling/stopping.
Transients: 	***Transients***, are sub-cycle disturbances of very short duration that vary greatly in magnitude. ***Causes:*** Lighting; equipment start-up and shutdown; welding equipment
Harmonics: 	***Harmonics*** are the periodic steady-state distortions of the sine wave due to equipments generating frequency other than the standard 60 cycles per second. ***Causes:*** non-linear loads, variable frequency drives

V. THE PROPOSED SOLUTIONS: POWER ELECTRONIC DEVICES (FACTS SYSTEMS),

FACTS controller can be classified into four categories as seen in the figure 3:

a) Series Controller (e.g. Static Synchronous Series Controller (SSSC))

b) Shunt Controller (e.g. STATCOM, D-STATCOM)

c) Combined Series – Series Controller

d) Combined Series – Shunt Controller (e.g. Unified Power Flow Converter (UPFC))

In this paper, we are interest by the use of 4 controllers which are: SVC, STATCOM, SSSC and UPFC, theirs definition and principle operating its will be dealt in the following paragraphs:

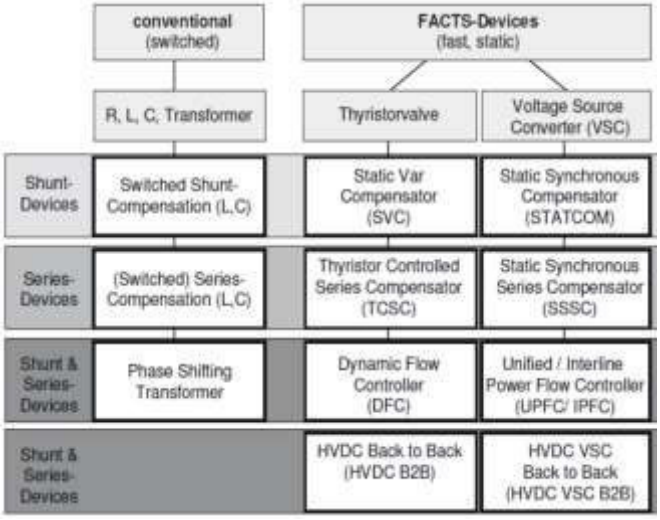


Fig. 3- Classification of FACTS devices

A. *Introduction of UPFC:*

The schematic diagram of the UPFC is shown in Figure 4. As shown this figure, the UPFC consists of 2 back-to-back, self-commutated VSCs connected through a common DC link. The first converter STATCOM (Static Synchronous Compensator) is coupled to the AC system through a shunt transformer, and the second converter SSSC (Static Series Synchronous Compensator) is coupled through a series transformer [8,9].

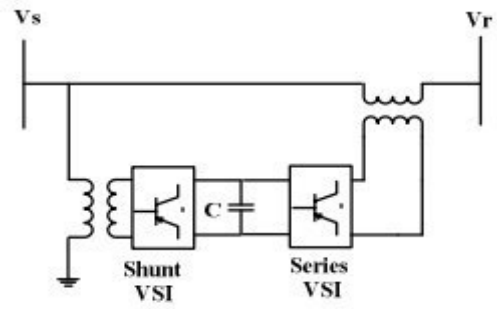


Fig. 4- UPFC schematic diagram

B. *The shunt part (STATCOM):*

The equivalent circuit of STATCOM is shown in Fig. 5(a). The STATCOM is a shunt FACTS controller based on voltage sourced converter (VSC) and a source of storage for the DC side [6]. The STATCOM can inject or absorb reactive power to or from the bus to which it is connected and thus regulate the bus voltage magnitude [10,11].

C. *The series part (SSSC):*

The equivalent circuit of SSSC is shown in Fig. 5(b). SSSC can be represented by a series voltage source [12]. it is capable of maintaining bus voltage, power flow, and line reactance . SSSC comprises of a VSC, a dc link capacitor and a coupling transformer. With proper control a voltage is injected in quadrature with the line current.

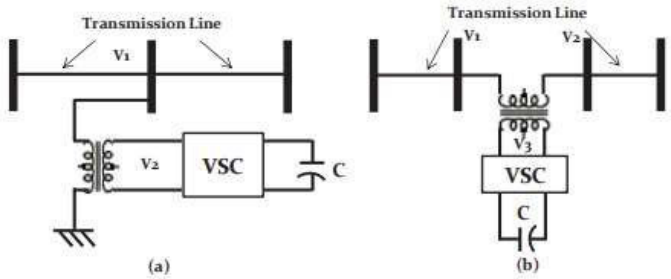


Fig. 5- Structure of (a) STATCOM and (b) SSSC

D. *Introduction of SVC:*

Fig. 7 shows a schematic diagram of a STATIC Var compensator (SVC) which is connected in the network in parallel. The compensator normally includes a thyristor controlled reactor (TCR), thyristor-switched capacitors (TSCs) and harmonic filters. With the presence of SVC, the transmission side voltage is controlled, and the Mvar ratings are referred to the transmission side[13].

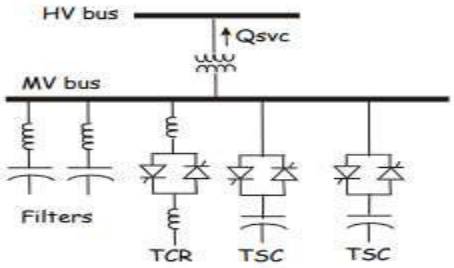
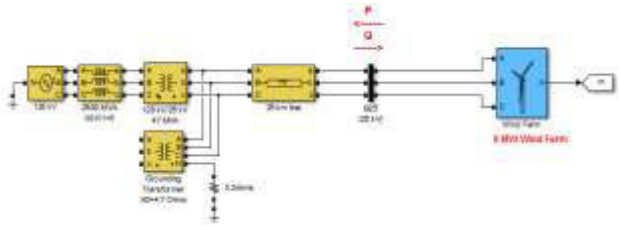


Fig. 6- Schematic diagram of an SVC.

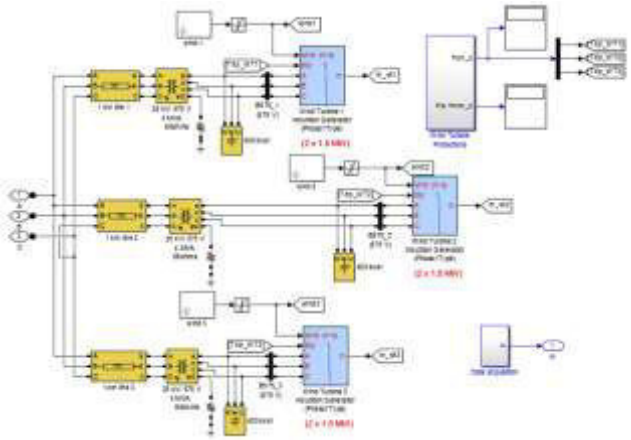
VI. SIMULATION RESULTS AND DISCUSSION

To verify the performance of the wind generation system under the control of the different proposed FACTS devices, we used in MATLAB SIMULINK software a six-1.5MW wind turbines connected to grid by a 25/120 KV transformer through 25 km transmission line as showed bellow in the figure 7 .

The Generators used in this model are squirrel cage Induction generators (IG) and stator windings are connected to the grid directly.



25-kV distribution system



9 MW wind farm (IG model)

Fig. 7- Simulation of a 9 MW wind farm based IG model connected to 25-kV distribution system

The simulations tests was performed for two cases, as follows:

• ***Case A:*** *System study with direct connection condition (no FACTS)*

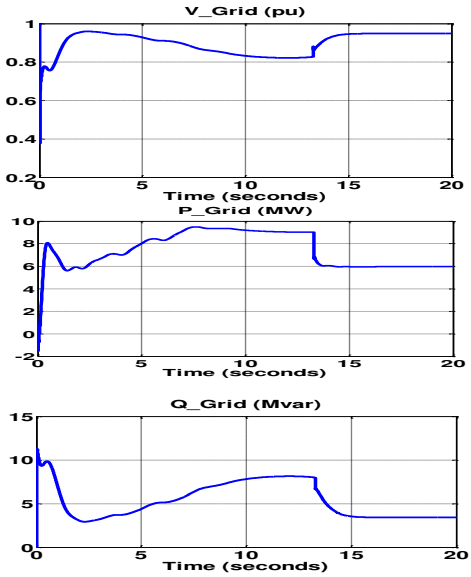


Fig. 8- The impact integration of wind turbine on the performance of the electrical network

The first test (Figure 8 without compensation) is done for 20s without intervention of any compensation. It is seen from Figure 9 that the grid voltage (V_ B25) is less than 1 pu, it is around 0.94 which is caused by fluctuations due by the variation in speed of Wind, This is infected the active power that presented in the form of an oscillatory signals and stabilizes at 6MW because the first wind turbine, lost its stability due to reduction of electrical torque, so the protection system is removed it from network and its active power become zero. We see also that, the reactive power dropped until the value of 4Mvar.

• ***Case b:*** *System study after applying FACTS devices (with SSSC, SVC, STATCOM and UPFC*

In the second test frame, and for investigating feasibility of the proposed devices: series (SSSC), shunt (SVC, STATCOM) and hybrid(UPFC) controllers a 03 phase (LLL) short circuit fault occurred at the level of one of the three wind turbines generators, we will randomly choose the wind generation system 2. The 03 phase (LLL) short circuit fault starts at t=15s, after 0.1 s, the circuit breaker isolated the wind generation system 2. The figure below gives the workspace in MATLAB/SIMULINK.

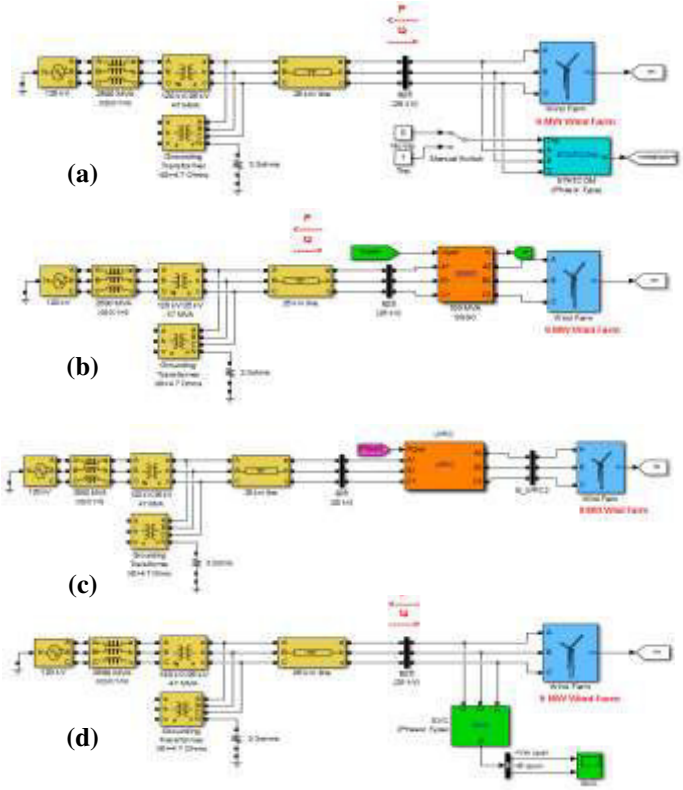


Fig. 9-Simulink model of the grid connected WECS based IG model: a-with STATCOM , b- with SVC, c- with UPFC

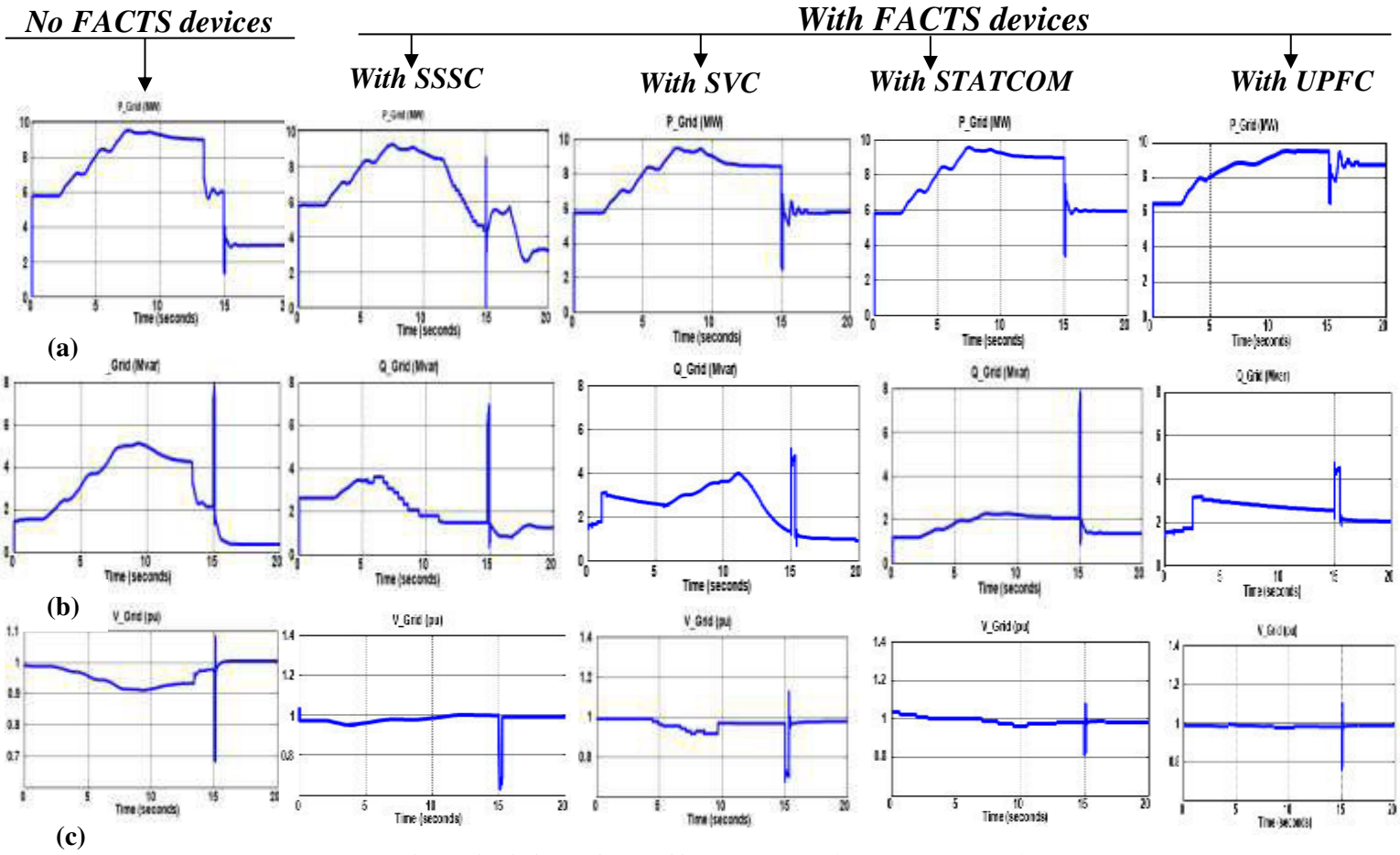


Fig. 10-Electrical network scope side: *a: active power, b: reactive power, c: Voltage*

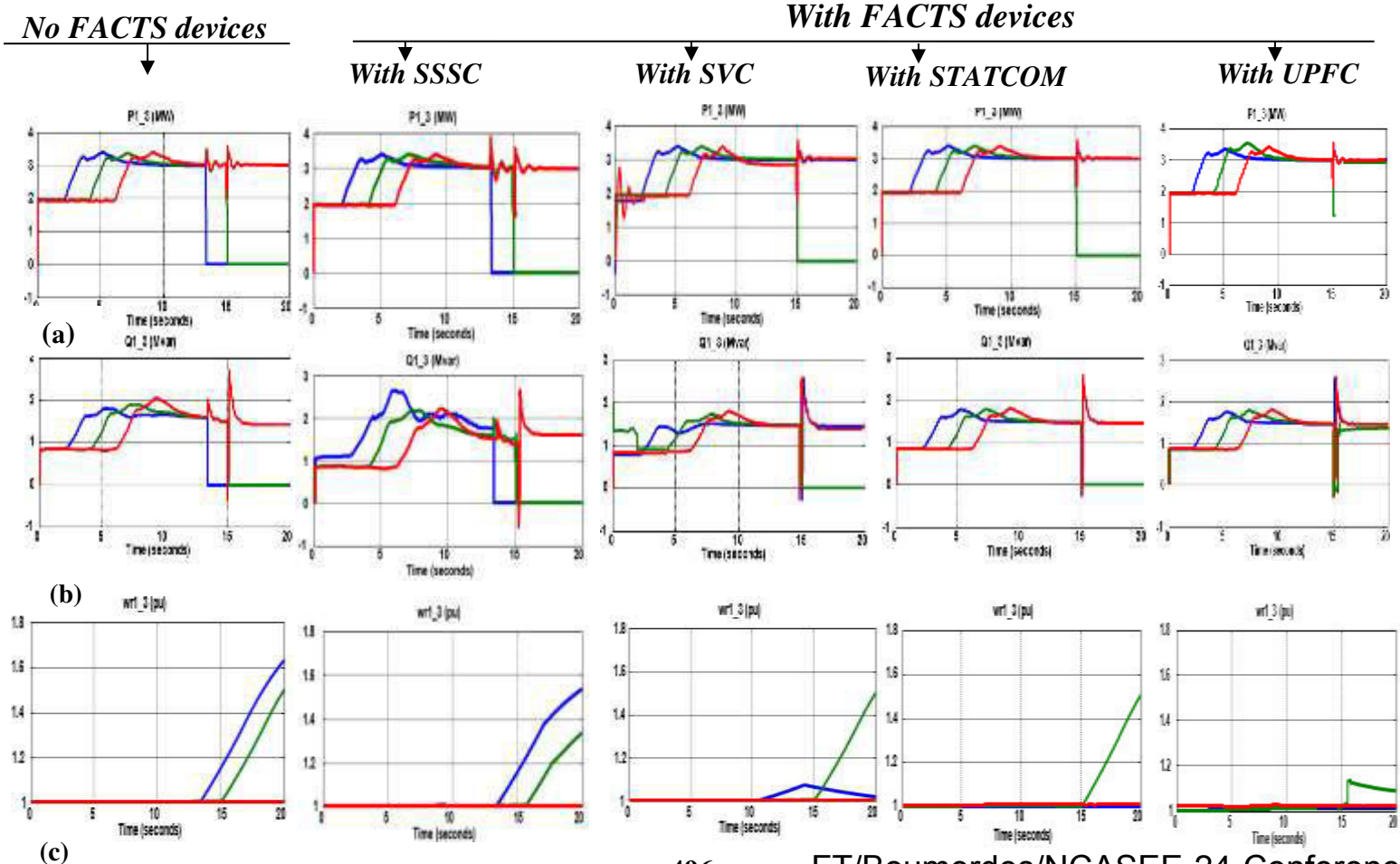


Fig.11- Wind turbine scope side: *d: active power, e: reactive power, f: Rotor speed*

The simulation results of the network configuration with the additional of: SSSC, SVC, STATCOM and UPFC are illustrated in Figs. 10, and 11 (*with FACTS devices*).

As indicated in the mentioned Figures, the blue color represent the wind 1, while the color green define the wind 2 and the rest (red color) represent the wind turbine 3. It is seen that because of voltage drop in bus 25kv at t = 13.43sec as clear in Fig (10-c) (*case: without FACTS devices and with 03 phases short circuit fault*) , the first wind turbine, cannot absorb enough reactive power and due to reduction of electrical torque, this turbine lost its stability and it cause to rise in rotor speed as Fig (11-c), so the protection system is removed it from network and its active and reactive power become zero. Due the occurrence of line to line to ground fault, on t = 15 sec. which is cleared on t=15.1 seconds, the second wind turbine will begin to accelerate as Fig (11-c). The second turbine is tripped by protection system as Fig (11-a,b,c), because of deficiency in reactive power. In this scenario, only the wind turbine 3 continues to work, it is responsible for supplying active and reactive power and delivering them to 25kV bus that after fault occurrence on t=15sec. The reactive power is injected to the 25kV bus via the: SVC, STATOCM and the UPFC devices, which are connected to terminal wind turbines at the point PCC. After fault clearance, the reactive power injection decreases (as presented in fig.11.b).

In the same figuration when we introduce the FACTS devices, we illustrate an improvement in the network gride and wind side.

At first, it is clear from these figures that: the first wind turbine which is removed by protection system on t=13.43 sec in the previous case (without compensation systems), it can keep its stability and produce active and reactive power because the presence of FACTS devices (the: SVC, STATCOM and UPFC) as shown in Figure (11-d) and (11-f). Second wind turbine cannot continue its service in the case of the compensation with: SSSC, SVC and STATCOM because of insufficiency capacity of SVC and STATCOM to supply necessary reactive power and it is tripped as presented in Figs.11:-d-e-f. Contrary, with UPFC the second wind turbine (wt_2) can continue its service normally. We note also that: the SSSC is not preferred for integrating wind energy into the grid due to its limitations compared to SVC and STATCOM.

Fig. 11.a,b show the active and reactive power exchanged between the IG and the grid for the two cases (with/without FACTS). After the fault, the active power is restored by using the UPFC while, with, STATCOM and SVC the wind 2 losses its fonctionnement in the same figuration we notice that the wind 1 and wind 2 where tripped and it's can not continues to work with SSSC.

Fig. 10.c shows the RMS value of the PCC voltage in the both cases (without and with FACTS devices). It is observed that in the case of without compensation the PCC voltage approximately decreases to 0.9 pu at t = 13.43sec when the first turbine isolated and also when the fault occurs at t=15s and cannot be restored to the normal level (1 pu). But using SVC and STATCOM the voltage at PCC can be restored quickly after the fault by injecting reactive power after fault clearance through shunt inverter. The UPFC not only increases the voltage sag to 1 pu, but also keep the stability of the wind turbine N'2 and keep it to produce active and reactive power.

As shown in Fig. 10,b, the absorbing reactive power from the grid is significantly reduced by using UPFC, which helps to avoid other problems such as voltage collapse (figure 10.c)

It is seen from Fig (11-a) that active power of 25 kv bus power decrease after fault occurrence on t =15, and this reduction continues until removal of fault on t = 15.1. However, the active power generation by the wind turbines is equal to 9MW with presence of UPFC is more than with presence of STATCOM (see Fig 11-a in case UPFC compared to Fig 8-11 in case of SVC and STATCOM).

Observing the Fig.11-a it can be seen that angle oscillations starts at the moment of fault (that is 15 s), then at the moment of tripping of wind turbine 2 starts to oscillate again and reaches steady state value approximately at t=15.2 sec. This means that from the time of the failure, and tripping wind farm 2 has passed around 0.2 sec, and change of voltage angle shows that system will remain stable.

As shown in Fig.11.c, the rotor speed gradually reduces to the pre-fault level and the system is stable in both cases with STATCOM, SVC and SSSC, but the UPFC can provide a better damping to post-fault oscillations.

VII. CONCLUSION

The cases investigated in this paper provide comparison between various FACTS Devices (i.e between UPFC, SVC, SSSC, and STATCOM) ant its role for energy renewable integration. After the different obtained resulted, we conclude:

1) The series controller (SSSC), suppresses the instantaneous voltage drop in the grid and it is able to improve transient behavior of WECS during fault, but the inclusion of SSSC does not change the WECS characteristics significantly.

2) The shunt controller (SVC) by injecting reactive power aims to restore the voltage at the terminals of the generator and improve the generator transient stability after fault clearing.

3) Compared to the SVC, the STATCOM can provide more reactive power under the same compensation position and capacity, as well as faster voltage recovery at the PCC in the fault period.

Finally, as a conclusion, the performance of UPFC is much batter then the performance of: SSSC, SVC and STATCOM in improving the stability of the wind energy conversion system (WECS) during disturbances.

REFERENCES

[1] M. Rama Sekhara Reddy1 , Dr. M. Vijaya Kumar Power Quality Improvement in DFIG based Wind Energy Conversion System using UPFC, published on: IOSR Journal of Engineering (IOSRJEN) , Vol. 3, Issue 1 (Jan. 2013), ||V2|| PP 46-54

[2] J. G. Slootweg and W. L. Kling, Wind power in power systems, John Wiley and Sons, Ltd., 2005.

[3] T. Abdelouahed , S.A Zidi and. S .Hadjeri , Control of a wind energy conversion system equipped by a STACOM for power quality improvement, in: 8th International Conference on Modelling, Identification and Control (ICMIC-2016) Algiers, Algeria- November 15-17, 2016.

[4] T. Abdelouahed , S.A Zidi and. S .Hadjeri and T. M.nasser , 'Apport des éoliennes dans les réseaux électriques- Utilisation d'un STATCOM ', in :

First International Conference on Smart Grids, CIREI'2019 March 17-18, 2019 at ENP- Oran – Algeria.

[5] M Maamri.,.T Abdelouahed, Tandjaoui. M. N. 'CONSEQUENCES D'INTEGRATION DES EOLIENNES DANS LES RESEAUX ELECTRIQUES', VIIth International Congress on Renewable Energy and the Environment,Sousse, Tunisi.2013

[6] M. N. Tandjaoui,C. Benachaiba, O. Abdelkhalek, B. Dennai, Renewable energy and its impact in power quality of distribution systems, in: College of Applied Sciences. Umm Al-Qura University Makkah, April 16-18, 2012

[7] M. N. Tandjaoui, M. Haidas & C. Benachaiba, Intégration des éoliennes pour la Production d'Energie Electrique en Algérie - Site Tindouf-, in :"SIPE'10, Journal of Scientific Research N° 0 vol. 1 (2010.)

[8] T. Abdelouahed , S.A Zidi and. S .Hadjeri, and Y.I. Djilani Kobibi, Control system design of UPFC for power flow management in transportation system using Fuzzy system and GA algorithm, First International Conference on Smart Grids, CIREI'2019, Oran – Algeria.

[9] Y.I. Djilani Kobibi, S. Hadjeri, S.A Zidi, M. Djehaf, Modelling a UPFC for the Study of Power System Steady state and Transient Characteristics, in International Conference on Electromechanical Engineering (ICEE'2012) Skikda, Algeria, 20-22 November 2012

[10] N.Nirmala, V.Suresh Kumar, 'A STATCOM Control Scheme for Wind Energy System to Improve Power Quality', Published in: International Conference on Information Communication and Embedded Systems (ICICES),2013

[11] M. S Alatshan , I AlHamrouni , T. Sutikno , A. Jusoh, 'Improvement of the performance of STATCOM in terms of voltage profile using ANN controller', Published in: International Journal of Power Electronics and Drive System (IJPEDS) Vol. 11, No. 4, December 2020, pp. 1966~1978

[12] T. F. Orchi, M. J. Hossain, 'Stability Improvement of Wind Farms using Shunt and Series Compensation ', in AUPEC 2012 Proceedings

[13] M.Noroozian; N.A.Petersson; B.Thorvaldson; A.B.Nilsson; C.W. Taylor, Benefits of SVC and STATCOM for Electric Utility Application, Published in: 2003 IEEE PES Transmission and Distribution Conference and Exposition (IEEE Cat. No.03CH37495)

Misalignment effect on efficiency of wireless power transfer

BENNIA Fatima[1], BOUDOUDA Aimad[1]

1 Laboratoire Ingénierie des Systèmes et Télécommunications (LIST), University M'Hamed Bougara, Avenue de l'indépendance, 35000, Boumerdes, Algeria
Corresponding author:
Email: f.bennia@univ-boumerdes.dz

Abstract – ***In the magnetic resonance-based wireless power transfer (WPT), any misalignment between resonators can reduce the strength of the magnetic coupling. Therefore, it is essential to analyze the effect of the degree of misalignment and receiver coil position on power transfer efficiency. This paper analyzes how misalignment affects WPT's key performance. An automated 3-axis platform was constructed for controlling coil motion. Test results indicate that efficiency can be maintained at 75% to 95% as long as the misalignment angle is less than 20 degrees and the air gap ratio is from 0 to 30 mm.***
Keywords: ***wireless power transfer (WPT), magnetic resonance, coil, power transfer efficiency, misalignment,***

I. INTRODUCTION

In recent years, wireless power transfer (WPT) has played a necessary role in various applications. The principle of this technology is represented by two electrically insulated coils, the first of which (the transmitter) is supplied with an alternating voltage; the electric current flowing through it generates a voltage and a current across the second coil (the receiver) by electromagnetic induction. This energy can be transferred via inductive coupling for short distances, resonant induction for medium distances, and electromagnetic waves for long distances [1]. Magnetic resonance inductive coupling transfer (MR-ICT) is A technology that is crucial to comfort and lifestyle in today's world. We rely on sophisticated electronic devices, such as implanted medical devices, electric cars, drones, and smartphones. These devices require a stable and reliable power source [2]. Traditionally, these devices are powered through wired connections or built-in batteries. However, using wires limits mobility and flexibility, while built-in batteries have a limited lifespan and require periodic replacements, leading to financial burdens, patient surgeries, and potential mission failures. Fortunately, WPT technology offers a more convenient and reliable solution to these challenges. Wireless power transfer (WPT) reduces associated problems and provides seamless charging for electronic devices by eliminating cables and batteries [3]. While WPT is the key to solving implant battery problems, coil misalignment can cause charging efficiency to drop. In this paper, an automated 3-axis platform was built to investigate the effects of misalignment on efficiency and power transfer capabilities under diverse configurations of misalignment and air gap. Traditionally, these devices are powered through wired connections or built-in batteries. However, using wires limits mobility and flexibility, while built-in batteries have a limited lifespan and require periodic replacements, leading to financial burdens, patient surgeries, and potential mission failures. Fortunately, WPT technology offers a more convenient and reliable solution to these challenges. Wireless power transfer (WPT) reduces associated problems and provides seamless charging for electronic devices by eliminating cables and batteries.

II. THEORETICAL BACKGROUND

Figure 1 shows a general schematic of a wireless power transfer system using inductive coupling with a series-parallel (SP) topology. This topology is an efficient way to send power wirelessly, as it achieves high power transfer efficiency. This system consists of a primary and secondary coil, which are inductively coupled and separated by air and skin. The primary and secondary coil self-inductances are called L1 and L2, respectively. The equivalent series resistances of the two coils are represented as R1 and R2, respectively. The equivalent resistance of the battery in the device is RL. The primary and secondary circuits are resonated with resonant capacitors C1 and C2 to improve transfer power and efficiency. An alternating current passing through the external coil induces an alternating magnetic field. The receiver coil picks up this field and induces AC voltage in the secondary coil. The resonant compensation circuit optimizes power transfer efficiency [4-7].

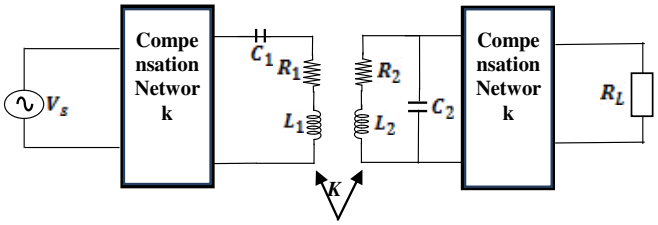


Fig. 1: Circuit model of the inductive power link.

Power transfer efficiency is a crucial parameter in wireless power transfer systems. It is defined as the ratio of the energy transmitted by the transmitting coil to the energy received by the receiving coil. In order to achieve high power transfer efficiency, it is necessary to maximize the coupling coefficient and the quality factor of the coils involved in the wireless power transfer. The coupling coefficient measures the degree of magnetic coupling between the transmitter and receiver coils. At the same time, the quality factor measures the energy losses in the coil due to resistance, capacitance, and inductance. Maximizing the coils' coupling coefficient and quality factor can significantly improve the overall wireless power transfer efficiency. The following equation gives the maximum power transfer efficiency:

$$\eta_{\text{total}} = \frac{K^2 Q_T Q_R}{(1+\sqrt{(1+K^2 Q_T Q_R)})^2} \quad (1)$$

The quality factor: The quality factor of the coil can be calculated using the parasitic resistance and inductance of the coil as [8]:

$$Q = \frac{\omega_0 L}{R_S} \quad (2)$$

Taking the skin effect into consideration, the total parasitic resistance, R_S, *is* computed as follows [9]:

$$R_S = R_{dc} \frac{t_c}{\delta(1-e^{\frac{t_c}{\delta}})} \quad (3)$$

Where: $R_{dc} = \rho_c \frac{l_c}{w t_c}$; $\delta = \sqrt{\frac{\rho_c}{\pi \mu f}}$; $\mu = \mu_r \mu_0$

$$\mathrm{l_c} = 4n[d_{out} - (n-1)s - nw] - s \quad (4)$$

t_c: Thickness of the conductor; f: Operating frequency;
ρ_c: Resistivity of the conductor; μ:Permeability constant;
μ: Relative permeability; δ: Depth of skin;
d_{out}: External diameter of the coil;
s: Space between coil's conductors;
n: Number of turns; w: Width of the conductor.

The angular frequency of resonance, ω_0, is calculated as follows:

$$\omega_0 = \frac{1}{\sqrt{LC}} \quad (5)$$

C: The resonant capacitors in the external.

Due to the resonant frequency f being constant, which equals 13.56 MHz, the angular frequency of resonance becomes $\omega_0 = 2\pi f$.

The self-inductances *L:* the following formula is used to determine the self-inductance of a spiral coil [10]:

$$L = \frac{\mu_0 n^2 d_{avg}}{2}\left[\ln\left(\frac{2.46}{\varphi}\right) + 0.2\varphi^2\right] \quad (6)$$

Where: $d_{avg} = \frac{(d_{in}+d_{out})}{2}$; $\varphi = \frac{(d_{in}-d_{out})}{(d_{in}+d_{out})}$ (7)

d_{avg}: the average diameter of the coil;
φ: fill factor of the coil

The mutual inductance M_{12}: In a wireless power transfer (WPT) system, mutual inductance is one of the most important factors affecting the efficiency and transmission of power. Mutual inductance refers to the physical phenomenon of an external circuit's magnetic field inducing an electromotive force (EMF) across an implanted circuit [11]. An equation can mathematically define the mutual inductance between the external and implanted circuits. Accurately measuring and controlling this characteristic is crucial for ensuring optimal performance and safety of the WPT system.

$$M_{12} = \theta \sum_{i=1}^{n_T} \sum_{j=1}^{n_R} M(x_i, y_i, d_{TR}) \quad (8)$$

Where: θ is a constant and changes with the shape of the coil. It is found empirically as 1.3 for square-shaped coils, n_T and n_R are the number of turns of the primary and secondary coil, x_i, y_i *is* the radius of the i^{th} turn in the primary coil, the radius of j^{th} turn in secondary coil [12].
$M(x_i, y_i, d_{TR})$ is the mutual inductance between the i^{th} turn of the primary coil and the j^{th} turn of the secondary coil and is defined as:

$$M(x_i, y_i, d_{TR}) = \mu\sqrt{x.y}\left[\left(\frac{2}{\gamma} - \gamma\right)K(\gamma) - \frac{2}{\gamma}E(\gamma)\right] \quad (9)$$

Where: $\gamma = \sqrt{\frac{4xy}{(x+y)^2 + {d_{TR}}^2}}$ (10)

The coupling coefficient k: indicates the proportion of the magnetic field of the primary coil L_1 that is captured by the secondary coil L_2 [13].

$$K = \frac{M_{12}}{\sqrt{L_1 L_2}}$$

III. MODEL SIMULATION

To determine the impact of misalignment on efficiency, this paper adopts a WPT model designed in 3D design Maxwell software with high misalignment tolerance, where D12 is the distance from the center point of the Rx to Tx coil, and X is the lateral misalignment distance of the Rx coil from the Tx coil. Shaft alignment can deviate in several directions. There can be offset misalignment in the vertical and horizontal directions, angular misalignments in these directions, and any varied combination of these misalignment cases. Figure 2 shows parallel offset misalignment, in which the shafts of the two machines are on two separate but parallel centerlines, and angular misalignment, in which the centerlines of the two shafts intersect at their coupling point and are at an angle to each other. These misalignments have a large impact on efficiency.

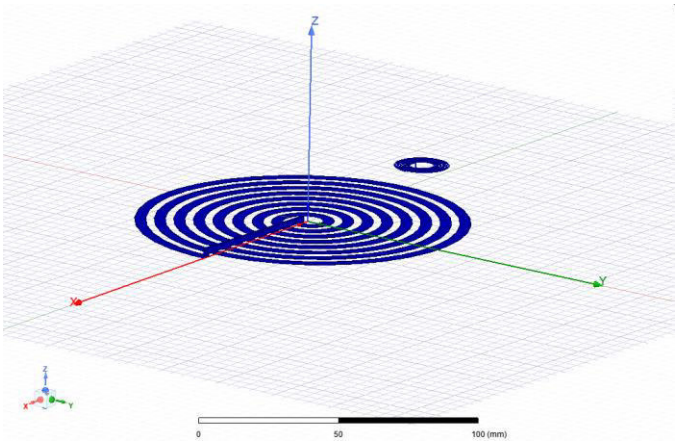


Figure.2. Coil modeling

TABLE I: Fixed parameters design parameters we will use in our analys.

Design parameters	Minimum value
Number of primary turns (N_T)	8
Number of secondary turns (N_R)	10
Width of conductor primary coil (W_T)	2.956 mm
Width of conductor secondary coil (W_R)	0.2313 mm
Spacing between conductors of primary coil (S_T)	4.083 mm
Spacing between conductors of secondary coil (S_R)	0.3792 mm
The outer diameter of the secondary coil ($dout_R$)	20 mm
The outer diameter of the primary coil ($dout_T$)	120.5 mm
The inner diameter of the secondary coil (din_R)	8 mm
The inner diameter of the primary coil (din_T)	10 mm
Frequency	13.56MHz

When the two coils are perfectly aligned (X=0), the coupling coefficient reaches its maximum value ($k=k_{max}$), allowing the energy transfer efficiency to reach its peak, typically around 90-95% for the given parameters. However, as X increases or decreases in negative values, the coils become increasingly misaligned, which reduces the magnetic coupling (k) and leads to a rapid decrease in efficiency. This drop becomes significant for large displacements (|X|>20 mm). The graph illustrating this relationship is symmetrical around X=0, as only the modulus of X influences the angle θ. A region of useful efficiency is observed in a range close to X=0, -5 mm≤ X ≤+5 mm, where the efficiency remains relatively high, above 70%.

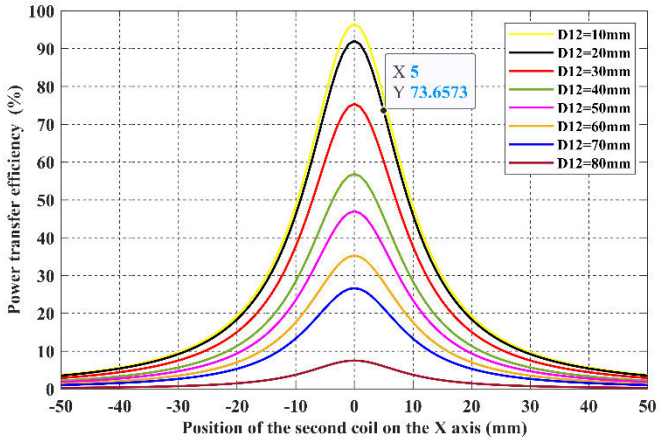


Figure.3. position of the second coil on X Axis

When the two coils are aligned (θ=0°), energy transfer is most efficient due to maximum magnetic coupling. As the angle increases, the coils become misaligned, gradually reducing efficiency. For angles greater than 30°, efficiency drops rapidly, and at 90°, the transfer becomes almost zero as the coils are perpendicular. The graph shows symmetry around θ=0°, with high efficiency maintained in a small range close to this alignment, highlighting the importance of minimizing angular misalignments.

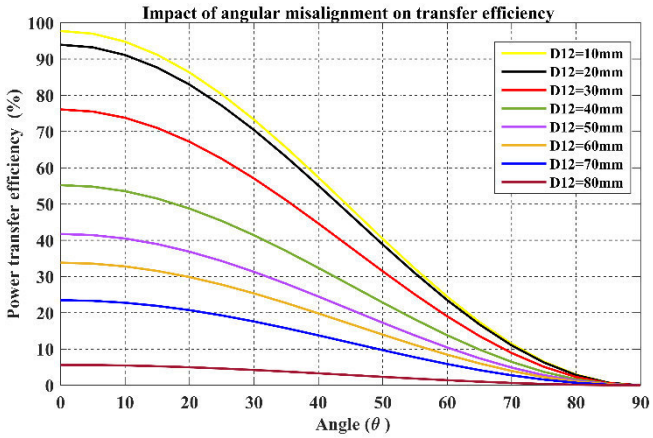


Figure.4. Impact of angular misalignment on power transfer efficiency

IV. CONCLUSION

This paper focuses on the analysis of the coils' misalignment with the magnetic resonance coupling wireless power transfer system (MRC-WPT), an important research area for powering implantable medical devices. We used a numerical model to analyze the influence of coil misalignment on transfer efficiency. The research presented in this paper highlights that The efficiency of wireless energy transfer depends on both the angular alignment (θ) and the horizontal position (X) of the secondary coil. Perfect alignment (X=0) provides maximum efficiency through optimal magnetic coupling. However, any misalignment, whether angular or horizontal, reduces magnetic coupling and, therefore, efficiency. As the gap in X or the angle θ increases, efficiency falls rapidly.

REFERENCES

[1] Xu D, Zhang Q, Li X. Implantable Magnetic Resonance Wireless Power Transfer System Based on 3D Flexible Coils. *Sustainability*. 2020; 12(10):4149. https://doi.org/10.3390/su12104149

[2] S. R.,Khan, Pavuluri, S. K., Cummins, G., & Desmulliez, M. P. (2020). Wireless power transfer techniques for implantable medical devices: A review. *Sensors*, *20*(12), 3487.doi: 10.3390/s20123487.

[3] C. Y. Kim, M. J. Ku, R. Qazi, H. J. Nam, J. W. ark, K. S.,Nam, ... & Jeong, J. W. (2021). The soft subdermal implant is capable of wireless battery charging and programmable controls for applications in optogenetics. Nature communications, 12(1), 535.doi: 10.1038/s41467-020-20803-y.

[4] S. Cheapanich, C. Anyapo and P. Intani, "Study of wireless power transfer using series-parallel topology," *2017 International Electrical Engineering Congress (iEECON)*, Pattaya, Thailand, 2017, pp. 1-4, doi: 10.1109/IEECON.2017.8075802.

[5] C. H. Lee, G. Jung, K. A. Hosani, B. Song, D. -k. Seo and D. Cho, "Wireless Power Transfer System for an Autonomous Electric Vehicle," *2020 IEEE Wireless Power Transfer Conference (WPTC)*, Seoul, Korea (South), 2020, pp. 467-470, doi: 10.1109/WPTC48563.2020.9295631.

[6] B. Luo, T. Long, L. Guo, R. Dai, R. Mai, and Z. He, "Analysis and Design of Inductive and Capacitive Hybrid Wireless Power Transfer System for Railway Application," in *IEEE Transactions on Industry Applications*, vol. 56, no. 3, pp. 3034-3042, May-June 2020, doi: 10.1109/TIA.2020.2979110.

[7] A. Lakhdari, N. E. M. Maaza, M. Dekmous, "Design and optimization of inductively coupled spiral square coils for the bio-implantable micro-system device, ". International Journal of Electrical and Computer Engineering, 2019. https://doi: 10.11591/ijece.v9i4.pp2637-2647

[8] M. Haerinia and R. Shadid, "Wireless Power Transfer Approaches for Medical Implants: A Review," *Signals*, vol. 1, no. 2, pp. 209–229, Dec. 2020, doi: 10.3390/signals1020012. [Online]. Available: http://dx.doi.org/10.3390/signals1020012

[9] S. Mehri, A. C. Ammari, J. Ben Hadj Slama, H. Rmili, "Geometry optimization approaches of inductively coupled printed spiral coils for remote powering of implantable biomedical sensors, " Journal of sensors, 2016. https://doi.org/10.1155/2016/4869571

[10] I. Hussain and D.-K. Woo, "Self-Inductance Calculation of the Archimedean Spiral Coil," *Energies*, vol. 15, no. 1, p. 253, Dec. 2021, doi: 10.3390/en15010253. [Online]. Available: http://dx.doi.org/10.3390/en15010253

[11] S. Raju, R. Wu, M. Chan, and C. P. Yue, "Modeling of Mutual Coupling Between Planar Inductors in Wireless Power Applications," in *IEEE Transactions on Power Electronics*, vol. 29, no. 1, pp. 481-490, Jan. 2014, doi: 10.1109/TPEL.2013.2253334.

[12] X. Zhang, C. Quan and Z. Li, "Mutual Inductance Calculation of Circular Coils for an Arbitrary Position With Electromagnetic Shielding in Wireless Power Transfer Systems," in *IEEE Transactions on Transportation Electrification*, vol. 7, no. 3, pp. 1196-1204, Sept. 2021, doi: 10.1109/TTE.2021.3054762.

[13] B. Fatima, B. Aimad and N. Fares, "Iterative Method Based Optimization of Wireless Power Transfer for Biomedical Implants," *2023 International Conference on Advances in Electronics, Control and Communication Systems (ICAECCS)*, BLIDA, Algeria, 2023, pp. 1-6, doi: 10.1109/ICAECCS56710.2023.10104669.

Synthesis and Application of Graphene Nanoribbon in Electronic Field

Benalia Kouini
Laboratory of caotings, Materials and Environment, M'Hamed Bougara University - Boumerdes 35000, Algeria
kouinib@univ-boumerdes.dz

Hossem Belhamdi
M'Hamed Bougara University – Boumerdes 35000, Algeria
h.belhamdi@univ-boumerdes.dz

Asma Nour El Houda Sid
Constantine University – Constantine 25000, Algeria
a.sid@univ-boumerdes.dz

Amina Hachaichi
Bourdj Bou Arreridj University – Bourdj Bou Arreridj 34000, Algeria
a.hachaichi@univ-boumerdes.dz

Abstract -Graphene has attracted the interest of both academia and industry in recent years, due to its isolation and the extensive research that followed have led, amongst others, to graphene nanoribbons (GNRs), a graphene-based structure having nano-scale dimensions and semiconducting or metallic electronic properties that depend on its geometry and dimensions.

These characteristics of GNRs are in stark contrast to those of graphene, which is a carbon sheet with semimetal, zero band gap characteristics. Due to their quasi-one-dimensional nature, GNRs differ significantly from the more widely known two-dimensional graphene sheets.

This review discusses the current development of graphene nanoribbons (GNRs), for application in electrochemical fields. This article reviews the progress that has been reported towards producing GNRs with predefined dimensions, by using bottom-up chemical synthesis approaches.

Keywords: Graphene, Nanoribbon, electronic properties, semimetal

1. Introduction to Graphene Nanoribbon and Carbon Nanotube

Carbon nanotube (CNT) and graphene nanoribbon (GNR) are new important carbon nanomaterials in the nanometer regime. Though CNT was discovered more than two decades ago, in 1991, by Japanese physicist Sumio Iijima, GNR was discovered more recently, in 2004, by Andre Geim and Konstantin Novoselov at the University of Manchester. Since their discovery, both of these carbon nanomaterials have gained a lot of importance due to their remarkable properties. Significant advances have been made in the discovery of the fundamental properties, exploring the possibility of engineering applications, and growth technologies. This chapter provides a brief description about CNT and GNR.

2. Graphene

Three-dimensional (3D) with a single layer graphite forms a material with two-dimensional (2D) called 2D graphite or a graphene layer. Graphene is an allotrope of carbon. Its structure is one-atom-thick planar sheets of sp2-bonded carbon atoms that are densely packed in a honeycomb crystal lattice [1]. The word *graphene* was originated as a combination of *graphite* and the suffix *-ene* [2]. Graphene is most easily described as an atomic-scale chicken net made of carbon atoms and their bonds (Figure 1.1).

The length of carbon/carbon bond in graphene is 0.142 nm [3]. Graphene sheets stack to form graphite with an interplanar spacing of 0.335 nm [4]. Graphene is the basic structural element of some carbon allotropes for example graphite, charcoal, CNTs, and fullerenes. The band theory of 2D graphite or a graphene layer was studied more than six decades ago, in 1947 by Wallace [5]. But until 2004, it was discovered that 2D crystals were thermodynamically unstable and could not exist. However, the experimental discovery of graphene in 2004 flaunted

common wisdom, and the Nobel Prize in physics for 2010 was awarded to Andre Geim and Konstantin Novoselov at the University of Manchester "for their groundbreaking experiments looking the two-dimensional material graphene" [6].

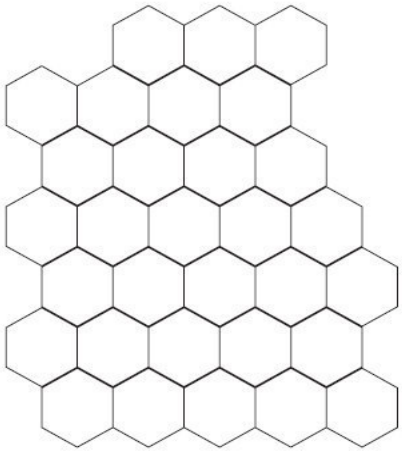


Figure 1.1 Graphene sheet

3. Graphene Nanoribbon

GNRs (also called nanographene ribbons) are strips of graphene with ultrathin width (<50 nm). Fujita and al mentioned in their studies the theoretical study on electronic states of graphene ribbons as a theoretical model [7].

In function of the orientation of carbon atoms on the edge of the graphene sheet, GNR is either armchair or zigzag (Figure 1.2). Zigzag GNR is always metallic; furthermore armchair GNR can be either semiconducting or metallic depending on geometry (chirality). For interconnect applications, zigzag GNR is the appropriate due to its metallic properties.

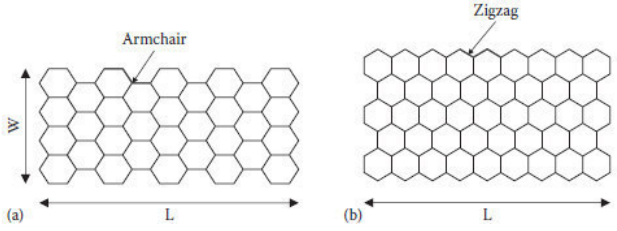


Figure 1.2 GNR with chirality (a) armchair and (b) zigzag.

4. Works Related to CNT- and GNR-Based Field-Effect Transistors

Javey et al. [8] fabricated Schottky barrier CNTFET (SB-CNTFET) using nanotube metal junctions. They used semiconducting SWCNTs that showed room-temperature conductance near the ballistic transport limit of 4e2/h and high current-carrying capability of ~25 mA per tube.

Durkop et al. [9] fabricated transistors using semiconducting CNTs with channel length exceeding 300 μm. They developed nanotubes in a tube furnace at 900°C.

Hoenlein et al. [10] discussed the applicability of CNTs in FETs. Zhou et al. [11] studied semiconducting SWCNTs in the diffusive transport regime.

Guo et al. [12] presented device physics of CNTFETs. They classified CNTFETs into two broad classes: (i) CNT MOSFET, which is similar to the traditional silicon MOSFET, and (ii) CNT MSDFETs, for metal source/drain FETs. For ballistic transport, they self-consistently solved the Poisson and Schrodinger equations using the nonequilibrium Green's function (NEGF) formalism.

The semiconducting CNTs are considered for the channel region of high-speed transistors due to their near ballistic electron transport. There are three different types of proposed CNTFET structures: SB-CNTFET, MOSFET-like CNTFET, and band-to-band tunneling CNTFET (BTBT-CNTFET). CNTFET has a structure that is similar to that of MOSFET. The ballistic transport operation and low OFF current makes CNTFET attractive for future nanoelectronic circuits.

In the work of Raychowdhury et al. [13], it has been shown that MOSFET-like CNTFET is superior as compared to SB-CNTFET. The proposed MOSFET-like CNTFET [14–16] is likely to be scaled down to 10 nm channel length. The simulation program with integrated circuit emphasis (SPICE) compatible equivalent circuit of MOSFET-like CNTFET is proposed by Deng and Wong [15,16].

Sinha et al. [17] developed a noniterative physics-based compact model for CNT transistors and interconnects. They showed that for a SB-CNTFET with the diameter range of 11.5 nm, the circuit can be more than 8× faster than that of 22-nm CMOS, with the tolerance to the variation in contact materials.
Lin et al. [18] designed the static random access memory (SRAM) cell using MOSFET-like CNTFET. They used dual-chirality in the design. It is shown that the CNTFET-based SRAM cell tolerates the process, power supply voltage, and temperature variations significantly better than its CMOS counterpart.
Zhang et al. [19] presented a design technique to overcome two major sources of CNTFET imperfections: metallic CNTs and CNT density variations.

5. Practical Circuits

Close et al. [20] reported the fabrication of the first stand-alone integrated circuit combining silicon transistors and individual CNT interconnect wires on the same chip. The operating frequency reported is above 1 GHz. They assembled MWCNT interconnects on top of a CMOS chip fabricated in a 0.25-µm silicon CMOS process containing about 11,000 transistors. They used MWCNTs as electrical wires interconnecting various stages of conventional ring oscillators.

6. Modeling of Graphene Nanoribbon

In graphene, the carbon atoms are arranged in a honeycomb structure, as shown in Figure 1.3. Though GNR and CNT are derived from the basic graphene structure, there are differences. The difference arises due to the boundary conditions. In CNT, the wave function is periodic along the circumference. In GNR, the wave function vanishes at the boundary of the two edges.
Depending on the orientation of carbon atoms, the edge of the graphene sheet is either armchair or zigzag. Zigzag GNR is always metallic, whereas armchair GNR can be either semiconducting or metallic depending on the number of carbon rings across the width. For interconnect applications, zigzag GNR is proposed due to its metallic property.

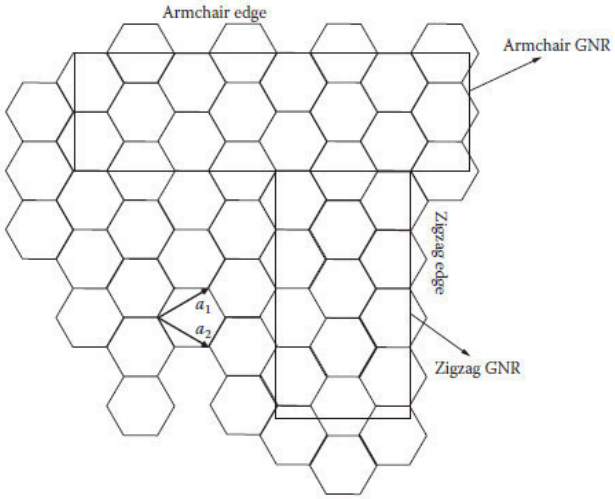


Figure 1.4 Lattice structure of graphene sheet.

Due to the high resistance of monolayer GNR, a multilayer GNR structure is proposed for the interconnects [21,22]. A multilayer GNR structure is modeled for VLSI interconnect as shown in Figure 1.5.
In Figure 3.10, the interconnect thickness is *t*, width is *w*, height from ground plane is *x*, and spacing between the interconnects is sp. The spacing between each graphene layer is δ (= 0.34 nm), which is the van der Waals gap [21]. The interconnect is assumed to have dimensions according to the specifications in International Technology Roadmap for Semiconductors (ITRS) [23] for 16-nm technology node.
Figure 1.6 shows the schematic diagram of an interconnect model [24] with a driver and a load at both ends of the interconnect. In Figure 3.9, the interconnect is modeled by a distributed RLC network, where *R*C is the resistance due to imperfect contacts, *R*Q is the quantum resistance, and *R*S (= *R*Q/λ) is the scattering resistance p.u.l., where λ is the MFP of electron in GNR. The quantum resistance is defined as [21].

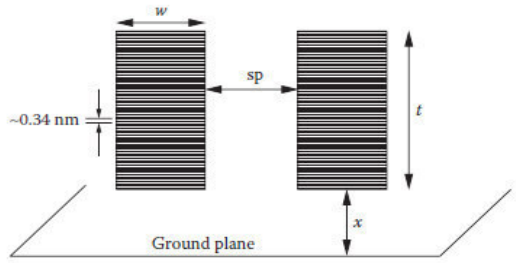


Figure 1.5 Multilayer GNR structure.

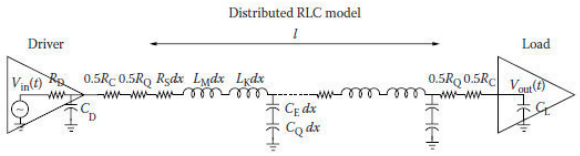


Figure 1.6 Schematic of the RLC interconnect model of GNR interconnect.

7. Application of GNRs as Semiconductors

Changing tool application can be the simplest process to use GNRs in electronics. In general, synthesized GNRs in solution do not only involve π-conjugation backbones but also consist of functional groups on the side to improve the solubility of substances and materials. Alkyl side chains are widely employed as the functional groups. Solubility and electrical conductivity are in trade-off, since insulating side chains disturb intermolecular electron hopping. Removal of these side chains is important for utilization of solvent-soluble GNRs for electronic devices and apparatus.

Mai and Feng developed poly(ethylene oxide) PEO-functionalized GNRs at the liquid−solid interface. The GNR thin-film-based field-effect transistors (FETs) were produced on SiO_2/Si substrates [25-29].

8. Application of GNRs in Dielectrics Field

The insulator is also an important part in electronics. Itami,Miyauchi, and Ito synthesized ~0.71 nm wide cove-type GNRs 109 with peripheral alkyl side chains that enables the GNRs soluble in various of organic solvents. A theory predicted that the cove-type GNRs also have partially flat bands at $2\pi/3 < |k| < \pi$ and two separated valleys in momentum space that were similar to those of zigzag-type GNRs. However, the cove-type GNRs 109 are supposed to have unclear flat bands because it is ultra narrow as was already mentioned in the example of the $N \leq 4$ zigzag-type GNR [30-32].

9. Application of GNRs to Electroluminescence Device

To an optical device application, red-colored electroluminescence from N = 7 armchair-type GNRs is proven. Prezzi and Schull suspended a N = 7 armchair-type GNR between the tip of STM and Au (111) substrate. The origin of the emission was attributed to intra-GNR excitonic transitions implicated electronic states located at the GNR terminal level. Yuuta Yano and al summarized the applications of structurally uniform GNRs [33,34].

10. Application of GNRs of electron transport

The theory to understand of electron transport in graphene and graphene nanoribbons is discussed,

focusing on the assistance provided by by atomic pseudopotentials (self-consistent and empiric) in finding out not only the band structure but also other essential transport factors like electron–phonon matrix elements and line-edge roughness scattering.

Electron–phonon scattering in suspended graphene sheets, impurity and remote-phonon scattering in supported and gated graphene, electron–phonon and line-edge roughness scattering in armchair-edge nanoribbons are summarized, take into account the potential utilization of graphene in devices of the future very large scale integration technology [35].

11. Application of GNRs of electrical properties utilizations

For to its interesting electronic properties, GNRs have become an active area of research for upcoming electronic applications. Theoretical calculations supplied direct band gap formation in ideal ribbons with controllable gap size. Also, remarkable magnetic properties were expected in zigzag edged GNRs, giving a new material system for the next generation spintronic device applications. More experimental studies developed different methods to reach GNR structures in the deep nano meter regime, for example top-down lithography, unzip of carbon nanotubes, exfoliation of chemical sono etc. And actually it was shown that a GNR in sub-10 nm regime can be made in a controllable manner with the nano wire masks. Principally, the latest confirmation of GNM structures indicates a potential pathway to produce large scale production of semiconducting graphene film by block copolymer lithography. The block copolymer lithography method used to make the nano mesh devices is permanently scalable and could permit for the rational design and manufacture of graphene-based devices and circuits with standard semiconductor processing.

On another side, the electrical transport studies of experimental manufactured GNRs presented more complicated transport behavior than that it was predicted in the theoretical studies of the ideal GNRs. This discrepancy was assigned to the disorders (edge roughness and charge impurities) induced Anderson localization and Coulomb blockade effect. By the way there is no doubt that the performance of current GNRs device can be well influenced by these parameters of disorder, and the device performance can be more improved with better control of graphene nanostructures [36].

12 Application of GNRs for logic applications devices

Juan and al presented the development in GNR devices. Simulations of new structures for nano devices and innovative fabrication techniques prove that GNRs can be suitable for digital circuit design in electronics applications.

The quantity and quality of scientific research works on graphene have motivated new researchers to progress in the domain. Recent works and developed articles present very optimistic predictions for the future of devices based on graphene and suitable suppositions of integration of graphene in to industry, where the processes will be always developed and scientist will have access to better materials for developing new devices with improved performance. However, the established industrial processes will only replace current materials for graphene even as the advantages of new profitable processes are sufficiently competitive to allow partial or complete changes [37].

References

1. Q. Ngo, D. Petranovic, S. Krishnan, A. M. Cassell, Q. Ye, J. Li, M. Meyyappan, and C. Y. Yang. Electron transport through metal-multiwall carbon nanotube interfaces. IEEE Transactions on Nanotechnology, 3(2):311–317, 2004.
2. S. Iijima. Helical microtubules of graphitic carbon. Nature, 354:56–58, 1991.
3. L. Delzeit, C. V. Nguyen, B. Chen, R. Stevens, A. Cassell, J. Han, and M. Meyyappan. Multiwalled carbon nanotubes by chemical vapor deposition using multilayered metal catalysts. Journal of Physical Chemistry B, 106:5629–5635, 2002.
4. Z. P. Huang, J. W. Xu, Z. F. Ren, J. H. Wang, M. P. Siegal, and P. N. Provencio. Growth of highly oriented carbon nanotubes by plasma-enhanced hot filament chemical vapor deposition. Applied Physics Letters, 73(26):3845–3847, 1998.
5. C. Bower, W. Zhu, S. Jin, and O. Zhou. Plasma-induced alignment of carbon nanotubes. Applied Physics Letters, 77(6):830–832, 2000.

6. Z. Yao, C. L. Kane, and C. Dekker. High-field electrical transport in single-wall carbon nanotubes. Physical Review Letters, 84(13):2941–2944, 2000.
7. L. Delzeit, I. McAninch, B. A. Cruden, D. Hash, B. Chen, J. Han, and M. Meyyappan. Growth of multiwall carbon nanotubes in an inductively coupled plasma reactor. Journal of Applied Physics, 91(9):6027–6033, 2002.
8. A. Javey, J. Guo, Q. Wang, M. Lundstrom, and H. Dai. Ballistic carbon nanotube field-effect transistors. Nature, 424:654–657, 2003.
9. T. Durkop, S. A. Getty, E. Cobas, and M. S. Fuhrer. Extraordinary mobility in semiconducting carbon nanotubes. Nano Letters, 4(1):35–39, 2004.
10 W. Hoenlein, F. Kreupl, G. S. Duesberg, A. P. Graham, M. Liebau, R. V. Seidel, and E. Unger. Carbon nanotube applications in microelectronics. IEEE Transactions on Components and Packaging Technologies, 27(4):629–634, 2004.
11. X. Zhou, J.-Y. Park, S. Huang, J. Liu, and P. L. McEuen. Band structure, phonon scattering, and the performance limit of single-walled carbon nanotube transistors. Physical Review Letters, 95:146805-1-146805-4, 2005.
12. J. Guo, S. O. Koswatta, N. Neophytou, and M. Lundstrom. Carbon nanotube field-effect transistors. International Journal of High Speed Electronics and Systems, 16(4):897–912, 2006.
13. A. Raychowdhury, A. Keshavarzi, J. Kurtin, V. De, and K. Roy. Carbon nanotube field-effect transistors for high-performance digital circuits-DC analysis and modeling toward optimum transistor structure. IEEE Transactions on Electron Devices, 53(11):2711–2717, 2006.
14. H.-S. P. Wong. Stanford University CNFET model. http://nano.stanford.edu/, 2014.
15. J. Deng and H.-S. P. Wong. A compact SPICE model for carbon-nanotube field-effect transistors including nonidealities and its application-Part I: Model of the intrinsic channel region. IEEE Transactions on Electron Devices, 54(12):3186–3194, 2007.
16. J. Deng and H.-S. P. Wong. A compact SPICE model for carbon-nanotube field-effect transistors including nonidealities and its application-Part II: Full device model and circuit performance benchmarking. IEEE Transactions on Electron Devices, 54(12):3195–3205, 2007.
17. S. Sinha, A. Balijepalli, and Y. Cao. Compact model of carbon nanotube transistor and interconnect. IEEE Transactions on Electron Devices, 56(10):2232–2242, 2009.
18. S. Lin, Y. Kim, and F. Lombardi. Design of a CNTFET-based SRAM cell by dual-chirality selection. IEEE Transactions on Nanotechnology, 9(1):30–37, 2010.
19. J. Zhang, N. Patil, A. Lin, H.-S. P. Wong, and S. Mitra. Carbon nanotube circuits: Living with imperfections and variations. Proceedings of the Design Automation and Test in Europe Conference and Exhibition, Dresden, March 8–12, 2010, pp. 1159–1164.
20. G. F. Close, S. Yasuda, B. Paul, S. Fujita, and H.-S. P. Wong. A 1 Ghz integrated circuit with carbon nanotube interconnects and silicon transistors. Nano Letters, 8(2):706–709, 2008.
21. C. Xu, H. Li, and K. Banerjee. Modeling, analysis, and design of graphene nano-ribbon interconnects, IEEE Transactions on Electron Devices, 56(8), 1567–1578, 2009.
22. Olga V. Zakharova, Elena E. Mastalygina, Kirill S. Golokhvast and Alexander A. Gusev, Nanomaterials 2021, 11, 2425.https://doi.org/10.3390/nano11092425
23. Arnab Hazra and Sukumar Basu,Journal of Carbon Research 2018, 4, 49; doi:10.3390/c4030049
24. Avanish Shukla , P. Dhanasekaran , N. Nagaraju , Santoshkumar D. Bhat ,Vijayamohanan K. Pillai , Electrochimica Acta 297 (2019) 267-280
25. Narita, A.; Feng, X.; Hernandez, Y.; Jensen, S. A.; Bonn, M.;Yang, H.; Verzhbitskiy, I. A.; Casiraghi, C.; Hansen, M. R.; Koch, A.H. R.; Fytas, G.; Ivasenko, O.; Li, B.; Mali, K. S.; Balandina, T.;Mahesh, S.; Feyter, S. D.; Müllen, K. Synthesis of structurally welldefined
and liquid-phase-processable graphene nanoribbons. Nat.Chem. 2014, 6, 126−133.
26. Mitoma, N.; Yano, Y.; Ito, H.; Miyauchi, Y.; Itami, K. GrapheneNanoribbon Dielectric Passivation Layers for Graphene Electronics.ACS Appl. Nano Mater. 2019, 2, 4825−4831.
27. Yano, Y.; Mitoma, N.; Matsushima, K.; Wang, F.; Matsui, K.;Takakura, A.;

Miyauchi, Y.; Ito, H.; Itami, K. Living annulative π-extension polymerization for graphene nanoribbon synthesis. Nature 2019, 571, 387−392
28. Abbas, A. N.; Liu, G.; Narita, A.; Orosco, M.; Feng, X.; Müllen,K.; Zhou, C. Deposition, Characterization, and Thin-Film-BasedChemical Sensingof Ultra-long Chemically Synthesized GrapheneNanoribbons. J. Am. Chem. Soc. 2014, 136, 7555−7558
29. Huang, Y.; Mai, Y.; Beser, U.; Teyssandier, J.; Belpula, G.;Gorp, H.; Straaso, L. A.; Hansen, M. R.; Rizzo, D.; Casiraghi, C.;Yang, R.; Zhang, G.; Wu, D.; Zhang, F.; Yan, D.; Feyter, S. D.;Müllen, K.; Feng, X. Poly(ethylene oxide) Functionalized Graphene Nanoribbons with Excellent Solution Processability. J. Am. Chem. Soc.2016, 138, 10136−10139
30. Nakada, K.; Fujita, M.; Dresselhaus, G.; Dresselhaus, M. S.Edge State in Graphene Ribbons: Nanometer Size Effect and Edge Shape Dependence. Phys. Rev. B: Condens. Matter Mater. Phys. 1996,54, 17954−17961.
31. Effect on Electronic Transport Properties of Graphene Nanoribbons and Presence of Perfectly Conducting Channel. Carbon 2009, 47,124−137.
32. Yamabe, T.; Tanaka, K.; Ohzeki, K.; Yata, S. Electronic Structure of Polyacenacene. A One-Dimensional Graphite. Solid State Commun. 1982, 44, 823−825
33. Chong, M. C.; Afshar-Imani, N.; Scheurer, F.; Cardoso, C.;Ferretti, A.; Prezzi, D.; Schull, G. Bright Electroluminescence from Single Graphene Nanoribbon Junctions. Nano Lett. 2018, 18, 175−181.
34. Yuuta Yano, Nobuhiko Mitoma, Hideto Ito, and Kenichiro Itami.A Quest for Structurally Uniform Graphene Nanoribbons: Synthesis,Properties, and Applications.This: J. Org. Chem. 2020, 85, 4−33
35. Massimo V Fischetti, Jiseok Kim, Sudarshan Narayanan, Zhun-Yong Ong, Catherine Sachs, David K Ferry and Shela J Aboud, J. Phys.: Condens. Matter 25 (2013) 473202 (37pp)
36. Jingwei Bai a, Yu Huang, Materials Science and Engineering R 70 (2010) 341–353
37. Juan M.Marmolejo-Tejada , JaimeVelasco-Medina, Microelectronics Journal48(2016)18–38

Benchmarking YOLOv10 Variants for Solar Panel Surface Defect Detection with Low Epoch Training

Younes Benazzouz [1], Djilalia Guendouz [2]
Department of instrumentation maintenance, Institute of Maintenance and Industrial Safety, University of Oran 2 Mohamed, Oran,Algeria
benazzouzyounes14@gmail.com [1]
lila.guen@yahoo.fr [2]

***Abstract*—Solar energy is becoming a good alternative energy source as the world shifts towards renewable energy sources. To ensure perfect energy production and minimize different losses, it is essential to maintain efficient and regular monitoring and maintenance. The surfaces of the panels should be clean to receive a good amount of solar energy to be transformed. Defects on solar panels, such as dust accumulation, cracks, and bird droppings, can reduce the efficiency of the transformation, so there is a need for an automated and a real-time monitoring systems that can accurately detect these defects. This study represents a computer vision approach using YOLOv10 (You Only Look Once) deep learning variants to detect defects and find anomalies through cameras. Three different algorithm variants were used to locate different defect classes, with the Convolutional Neural Network (CNN) trained on 3,272 images to achieve detectable classes with a good detection percentage level.**

Keywords—Photovoltaic Systems, Surface Defects, YOLOv10, Deep Learning, Computer Vision, Image Processing.

I. Introduction

The significant growth in the use of electrical and electronic devices correlates directly with the world demand for electric power [1], so the need for alternative energy sources is more important than before, especially renewable sources. Solar energy is one of the useful and crucial sources because of its availability throughout the day, making it a long-term energy solution. (PV) photovoltaic systems are made to convert this energy into (DC) direct current using diodes and solar cells. Monitoring the efficiency of this transfer depends on the good functioning of these cells, so regular and real-time maintenance is important to have a perfect DC energy output with a minimum of losses.

The idea of remote monitoring and defect detection is more useful than physical or on-the-ground human inspection. Utilizing cameras or drones to capture images of solar panels across different areas and applying image processing and computer vision techniques enables the analysis of visual defects such as dust, cracks, bird droppings, or any surface anomalies. This requires developing deep learning models, such as convolutional neural networks CNNs, to classify and detect the types of defects in real-time after processing by spotting unexpected patterns on the panel surface.

Three YOLOv10 model variants, YOLOv10-N, YOLOv10-S, and YOLOv10-M, are used as deep learning models to detect defects. Each model uses the same dataset for training, with the same number of images, to discuss the rapidity and efficiency of each model. Additionally, the F1 score, recall, and precision of each YOLOv10 variant are compared to highlight their respective strengths and weaknesses in the context of defect detection in PV systems. YOLOv10-N, designed for extremely resource-constrained environments, offers a balance of speed with a latency of 1.84 ms and an AP value of 38.5. YOLOv10-S, balancing speed and accuracy, provides a higher AP value of 46.3 with a latency of 2.49 ms. YOLOv10-M, intended for general-purpose use, achieves the highest AP value of 51.1, at the cost of increased latency at 4.74 ms [2].

The main contributions of this paper are as follows:

1- We present a comprehensive evaluation of three YOLOv10 variants YOLOv10-N, YOLOv10-S, and YOLOv10-M specifically applied to the task of defect detection in photovoltaic (PV) surfaces.

2- A demonstration of the practical application of deep learning models for real-time remote monitoring and defect detection in PV systems.

II. Literature Review

Recent studies developed methods to detect defects in photovoltaic systems. A research paper [3] improved the YOLOv5 algorithm with a weighted bidirectional feature pyramid mechanism, achieving a significant accuracy boost with a peak level of mean average precision (mAP), In [4], a study used YOLOv7 on RGB images with 75 epochs through grayscale conversion, data augmentation, and the Nadam optimizer, achieving a better level of accuracy in detecting solar panel faults, In [5], a system using YOLO was developed to locate planar solar panels in real-time using drone-captured images, In [6], Ameerdin et al. developed a deep learning model using YOLOv8 and thermal imaging to detect PV defects, with a 76% confidence level also in [7], Han et al. proposed a deep learning approach using an improved YOLOv3-tiny model for detecting faults in solar panels, employing UAVs equipped with thermal cameras, In [8], Özer and Türkmen developed an AI-based drone system using YOLOv5, YOLOv7, and YOLOv8 for solar panel detection, achieving an F1 score of up to 97% with YOLOv5l at 150 epochs and in [9], Yazdani et al. compared YOLO and Faster R-CNN for hotspot detection in photovoltaic panels, concluding that YOLO offers superior detection speed and quality, making it ideal for large-scale solar power plants, In [10], Greco et al. proposed a YOLO-based Convolutional Neural Network for detecting and segmenting PV panels from

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

aerial thermal images, In [11], Zhang and Yin proposed an improved YOLOv5 algorithm for detecting solar cell surface defects, achieving an mAP of 89.64%—a 7.85% improvement over the original algorithm—while maintaining real-time detection at 36.24 FPS.

III. Proposed Flowchart

The process begins with the collection of a dataset that should include a diversity of classes with different patterns such as Dust, Cracks, Bird Droppings, and Panels without Anomalies. After that, the data should be annotated and then divided into training, validation, and test sets. The preprocessed data is used to train three variants of the YOLOv10 model (YOLOv10s, YOLOv10n, YOLOv10m) on the Google Colab platform using Python. Then, the model variants are compared and analyzed to be deployed in the real world. The whole process can be represented as follows in a flowchart:

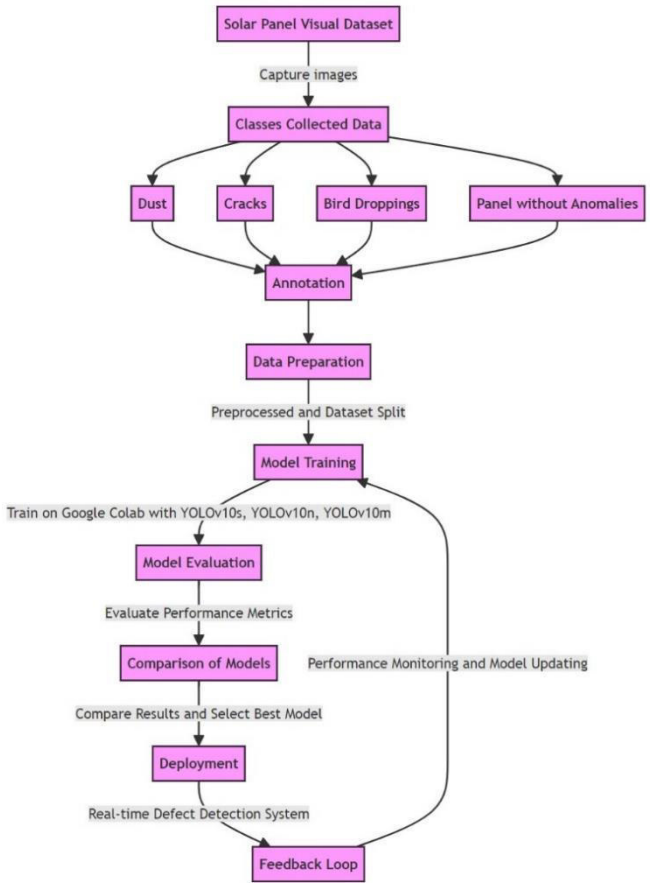


Fig. 1. Workflow Diagram of the Defect Detection Process Using YOLOv10 Variants

IV. Materials & Methods

A. YOLOv10 Architecture

Over recent years, YOLO models have become the leading choice for real-time object detection due to their balance between computational efficiency and detection performance. YOLOv10 is the latest version of them. Released in May 2024, this version represents a new generation in real-time object detection. This version eliminates Non-maximum suppression (NMS) in post-processing, which enhances inference latency, and it utilizes depthwise and pointwise convolutions. The Compact Inverted Block (CIB) reduces computational redundancy while maintaining high performance. The figure 2 below compares latency-accuracy using an official pre-trained models between YOLOv10 and previous versions like YOLOv7[12], YOLOv8[13], and YOLOv9[14], showing how much it outperforms the other versions.

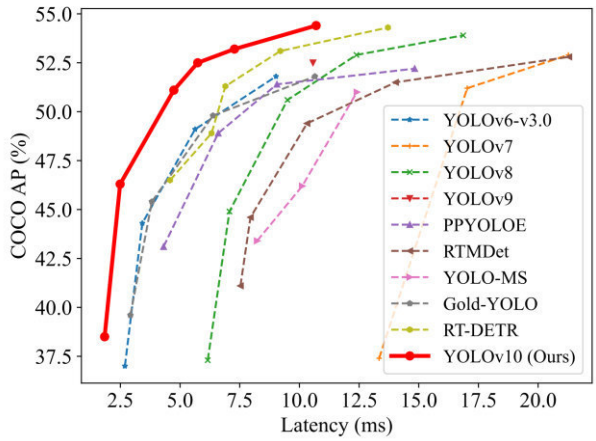


Fig. 2. Comparisons of YOLOv10 with others in terms of latency-accuracy[2].

B. Dataset

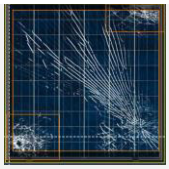
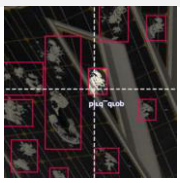
A well-prepared dataset is important for the training of deep learning models. We used in this study the clean-dirty solarPanel Dataset from Roboflow Universe [15]. Roboflow is a platform that offers a wide range of open-source datasets. A total of 3,272 images constituted the original dataset, containing various conditions of PV surfaces, both defected and non-defected, to be classified in the end by our YOLO model. We used only rotation as an augmentation technique to augment our data, such as 90° clockwise and counter-clockwise rotation, and random rotations of -20° and +20°. To standardize the input dimensions, we chose a size of 640x640 pixels for all images as a preprocessing step. After this process, the dataset size increased to 9,091 images. A total of 8,733 images were used as a training set, 257 images were reserved for model validation, and 101 images were set aside for final testing to evaluate the model's performance.

C. Dataset annotation and classes

Four classes were considered to be annotated as a bounding box. The first one is the panel itself, which should be classified and detected before detecting the defect itself. Then, three defects were considered: the first one is dust, which is common in dusty regions like deserts; the second one is cracks, which can be caused accidentally; and the last one is bird droppings, which can appear in places where birds live.

TABLE I. The Dataset Classses With Anotations

Class	Annotation	Description

Panel		the panel surface is the the relevant area of interest before detecting any defects, this annotation is important for isolating the panel from the background or other objects
Dust		This class includes any visible dust layers that can obstruct sunlight from reaching the solar cells and it refers to fine particles that accumulate on the surface.
Cracks		Cracks are fractures or breaks that appear on the solar panel's surface. These can result from physical impacts or even cyclic thermal stress [16]
bird droppings		This can be considered as an organic waste deposits left by birds and can cause localized shading so it reduce the panel's efficiency

D. Model variants Training

The environment used for model training is the cloud-based Google Colab coding platform. It allows for cloud training with intensive tasks without the need for local high-performance hardware. This platform provides an online based access to an NVIDIA Tesla T4 GPU, equipped with 2,560 CUDA cores. The GPU has 16 GB of GDDR6 memory, All model variants were trained with a batch size of 20 for a limited epochs training equal to 30, An epoch in machine learning refers to one complete pass through the entire training dataset by the learning algorithm and it indicate the number of times the model learns from the entire dataset, while the batch size determines how many samples are used in each forward and backward pass through the deep learning network The process of training included different image augmentation techniques: Blur, MedianBlur, ToGray, and Contrast Limited Adaptive Histogram Equalization (CLAHE) techniques. Each variant used a specific number of pretrained weights. The optimizer was chosen automatically, using AdamW with a learning rate of 0.00125 and momentum of 0.9.

```
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 535.104.05             Driver Version: 535.104.05   CUDA Version: 12.2     |
|-------------------------------+----------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |         Memory-Usage | GPU-Util  Compute M. |
|                                         |                      |               MIG M. |
|===============================+======================+======================|
|   0  Tesla T4                       Off | 00000000:00:04.0 Off |                    0 |
| N/A   55C    P8              10W /  70W |      0MiB / 15360MiB |      0%      Default |
|                                         |                      |                  N/A |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

Fig. 3. Specification of the GPU used on Google Colab

E. Model variants Performance Evaluation

To effectively evaluate the performance of any model, there are specific metrics made as indicators to judge the model's performance. In the end, Mean Average Precision (mAP) is the primary metric, and it represents the average of precision-recall curves considering all the classes. Also, the Precision metric indicates the proportion of the correctly predicted positive instances, and the Recall measures how much the model is able to detect all relevant instances of the target classes. And in the end, the F1 Score indicates the reliability of the model. These metrics are mathematically represented below:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negatives} \quad (2)$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \quad (4)$$

where N is the number of classes, and AP_i represents the average precision for class i.

F. Results and Discussion

We visually evaluated the detection capabilities of the YOLOv10 variants, showing the model's performance detection over all classes: bird droppings, dust accumulation, cracks, and intact panels, as seen in Figures 4, 5, and 6 below. We showed a sample of the same experimental pictures used to obtain visual computer vision accuracy results in the end.



Fig. 4. YOLOv10n Detection Results



Fig. 5. YOLOv10s Detection Results



Fig. 6. YOLOv10n Detection Results

For the represented picture samples Figures 4 to 6, YOLOv10n shows a consistent detection of intact panels, and YOLOv10s works better with dusty surfaces classe, showing a high confidence level. Both YOLOv10n and YOLOv10s show limitations in crack detection, and regarding YOLOv10m, it shows a balanced performance in detecting cracks and dust accumulation. The bird droppings class was detected with confidence ranging between 0.42 and 0.64, maintaining a moderate confidence level.

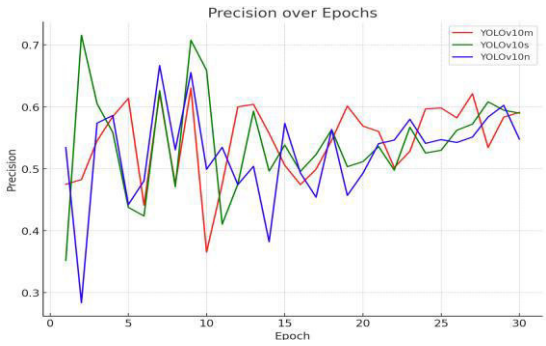


Fig. 7. Precision over Epochs for YOLOv10 Variants

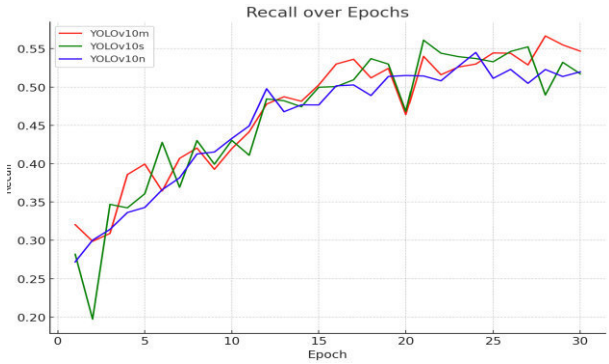


Fig. 8. Recall over Epochs for YOLOv10 Variants

Figure 4 and 5 show the recall and precision variation metrics over 30 epochs for the three variants used. Both metrics generally tend to stabilize after epoch 20 compared to the beginning of the training, which is a good indication that the models are converging. The tendency towards stability is a positive indicator that the models are learning effectively.

TABLE II. PRECISION, RECALL, AND F1 SCORES FOR DIFFERENT CLASSES ACROSS YOLOV1 VARIANTS

Model Variant	Class	Precision	Recall	F1 Score
YOLOv10n	Bird Drop	0.429	0.205	0.277
	Cracked	0.695	0.756	0.724
	Dusty	0.572	0.322	0.412
	Panel	0.709	0.769	0.738
YOLOv10s	Bird Drop	0.503	0.253	0.337
	Cracked	0.705	0.757	0.73
	Dusty	0.494	0.354	0.412
	Panel	0.685	0.776	0.728
YOLOv10m	Bird Drop	0.553	0.17	0.26
	Cracked	0.722	0.761	0.741

	Dusty	0.552	0.388	0.456
	Panel	0.662	0.792	0.721

The table presents the precision, recall, and F1 scores for different classes. Some values, such as those in the "Bird Drop" class, are relatively low, but the experimental results show that the three variants can detect this class. The "Panel" class consistently shows good detection results across all variants. This consistent performance could be due to the presence of the "Panel" class in every annotation scenario.

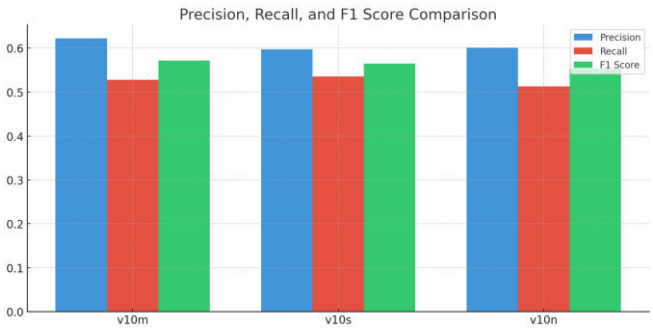


Fig. 9. Precision, Recall, and F1 Score Comparison

Figure 6 three variants (v10m, v10s, and v10n) are very similar in their overall performance with this 20 epochs number configuration. We can notice a difference in YOLOv10m, which has a slightly higher precision and that indicates better results at minimizing false positives. However, the recall and F1 scores are nearly the same across all three variants, indicating that their ability to detect true positives and balance precision with recall is quite similar. The slight edge in precision makes YOLOv10m better than the others overall, but the difference is not significant.

V. Conclusion

This research demonstrates how computer vision techniques can assist in detecting defects in photovoltaic (PV) systems. Using different variants of the YOLOv10 deep learning model for capturing surface defects, the proposed methodology resulted in an acceptable accuracy rating, as shown by the experimental images. These findings can improve remote monitoring and defect detection in PV systems and it contributes a more reliable and efficient solar energy production. An extension of this work could enhance model accuracy for specific defects, explore other PV systems, or train the model with more than 30 epochs, potentially showing greater stability in the performance metrics.

[1] Ahmad, T., & Zhang, D. (2020). A critical review of comparative global historical energy consumption and future demand: The story told so far. *Energy Reports*, *6*, 1973-1991.

[2] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*.

[3] Huang, J., Zeng, K., Zhang, Z., & Zhong, W. (2023). Solar panel defect detection design based on YOLO v5 algorithm. *Heliyon*, *9*(8).

[4] Wong, W. T., & Phang, S. K. (2024, January). Studies on Techniques to Improve YOLO in Fault Detection Using RGB Images of Solar Panels. In *2024 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)* (pp. 164-168). IEEE.

[5] Pérez, R. M., Arias, J. S., & Méndez-Porras, A. (2019, August). Solar panels recognition based on machine learning. In *2019 IV Jornadas Costarricenses de Investigacion en Computación e Informática (JoCICI)* (pp. 1-5). IEEE.

[6] Ameerdin, M. I., Jamaluddin, M. H., Shukor, A. Z., Kamaruzaman, L. A. H., & Mohamad, S. (2024, May). Towards Efficient Solar Panel Inspection: A YOLO-based Method for Hotspot Detection. In *2024 IEEE 14th Symposium on Computer Applications & Industrial Electronics (ISCAIE)* (pp. 367-372). IEEE.

[7] Han, S. H., Rahim, T., & Shin, S. Y. (2021, January). Detection of faults in solar panels using deep learning. In *2021 international conference on electronics, information, and communication (ICEIC)* (pp. 1-4). IEEE.

[8] Özer, T., & Türkmen, Ö. (2024). An approach based on deep learning methods to detect the condition of solar panels in solar power plants. *Computers and Electrical Engineering*, *116*, 109143.

[9] Yazdani, H., Radmehr, M., & Ghorbani, A. (2023). Innovative high-speed method for detecting hotspots in high-density solar panels by machine vision. *Energy Harvesting and Systems*, *10*(2), 277-286.

[10] Greco, A., Pironti, C., Saggese, A., Vento, M., & Vigilante, V. (2020, January). A deep learning based approach for detecting panels in photovoltaic plants. In Proceedings of the 3rd International Conference on Applications of Intelligent Systems (pp. 1-7).

[11] Zhang, M. and Yin, L., 2022. Solar cell surface defect detection based on improved YOLO v5. IEEE Access, 10, pp.80804-80815.

[12] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7464-7475).

[13] Sohan, M., Sai Ram, T., Reddy, R., & Venkata, C. (2024). A review on YOLOv8 and its advancements. In International Conference on Data Intelligence and Cognitive Informatics (pp. 529-545). Springer, Singapore.

[14] Wang, C. Y., Yeh, I. H., & Liao, H. Y. M. (2024). YOLOv9: Learning what you want to learn using programmable gradient information. arXiv preprint arXiv:2402.13616.

[15] S. Safa, "clean-dirty solarPanel Dataset," Roboflow Universe, Jun. 2024. [Online]. Available: https://universe.roboflow.com/safa-jl78c/clean-dirty-solarpanel.

[16] Ennemri, A., Logerais, P. O., Balistrou, M., Durastanti, J. F., & Belaidi, I. (2019). Cracks in silicon photovoltaic modules: a review. *J. Optoelectron. Adv. Mater*, *21*(1-2), 74-92.

Modeling and Simulation of Wind Power System: Comparative Study

Yaakoub Diboune

Electrical Systems and Remote-Control Laboratory (LabSET)
Automatic and Electrotechnics Department
University of Blida 1
Blida, Algeria.
yaakoubd479@gmail.com / diboune_yaakoub@univ-blida.dz

Djilali Kouchih

Electrical Systems and Remote-Control Laboratory (LabSET)
Automatic and Electrotechnics Department
University of Blida 1
Blida, Algeria.
djkouchih@yahoo.fr

***Abstract* – Wind energy has experienced a notable increase in its share of electricity generation, especially in transmission networks. To enhance the efficiency and reliability of power systems, precise and thorough mathematical modeling of wind turbine generators is crucial. Among different technologies, the doubly-fed induction generator is distinguished by its economic benefits, utilizing partial converters and induction machines for cost-effective performance. This paper presents a comparative study of modeling and simulation of a doubly-fed induction generator (DFIG) based wind turbine using two reference frames: the abc stationary frame and the d-q rotating frame. The investigation focuses on analyzing the impact of variable rotor speed on the electrical and mechanical quantities in both reference frames. Simulation results demonstrate that in the abc reference frame, currents and voltages exhibit dynamic changes corresponding to the rotor speed variations, while in the d-q frame, these quantities remain constant in their steady-state values. The findings highlight the advantages and limitations of each reference frame for dynamic system analysis and control design.**

Keywords – Doubly-fed induction generators, Wind turbine, Modeling, Simulation.

I. INTRODUCTION

The growing demand for renewable energy has led to significant advancements in wind turbine technology. Among various generator types, The doubly-fed induction generator (DFIG) is extensively utilized for its efficiency in changeable wind speed conditions and provide reactive power support. However, the accurate modeling and analysis of DFIG systems remain challenging, particularly when dealing with dynamic speed variations [1]-[5].

The principal benefit of employing DFIGs in wind turbine applications is their direct linkage to the grid through the three-phase stator windings, thereby preventing the need for full-scale power converters. Due to its cost-effectiveness, this design has gained significant popularity for wind turbines functioning under variable-speed situations [6]. Typically, the power electronic converter manages just a minor portion of the total power provided by the DFIG, usually between 20% and 30%, thereby considerably minimizing energy losses in the converter [7].

The DFIG model is defined in three reference frames: the stationary stator reference frame, the rotor-speed rotating reference frame, and the synchronous rotating reference frame. The synchronous reference frame is frequently favored for control design because of its simplicity, as all currents and voltages in this frame are expressed in direct current (DC) form [8, 9]. Alternative research has represented DFIG variables within their respective natural reference frames or utilized the "quadrature-phase slip-ring" model, which integrates stator and rotor dynamics [10].

DFIGs modeling in the three-phase (abc) reference frame is intrinsically intricate, as the system equations incorporate time-dependent coefficients associated with the angular position θ. The d-q reference frame, with the Park transformation, is frequently employed to simplify these equations and render them independent of θ. This mathematical method converts the equations into a two-phase orthogonal reference frame, yielding constant coefficients and enabling analysis and control design [11–17]. In the d-q frame, the variables are conventionally represented as d (direct) and q (quadrature), signifying the two orthogonal components.

According to the state model method [18], this study investigates the behavior of DFIG-based wind turbines under two distinct reference frames: the abc stationary frame, which captures instantaneous variations in phase quantities, and the d-q rotating frame, which simplifies analysis by transforming variables into a steady-state equivalent. By comparing simulation results, the study aims to highlight the strengths and weaknesses of each reference frame, providing insights into their suitability for dynamic system analysis and control design.

II. TURBINE MODELING

The blades of a wind turbine harness energy from the wind and transmit it to the rotor of an electric generator via a gearbox [19][20]. The power of the wind turbine is determined using equation (1).

$$P_v = \frac{1}{2}.\rho.S.v^3 \quad (1)$$

S : the surface swept by the blades of the turbine m^2
ρ : the air density
v : wind speed
The wind turbine's mechanical power is given by (2) [21]:

$$P_t = \frac{1}{2}.\rho.S.C_p(\lambda,\beta).v^3 \quad (2)$$

$C_p(\lambda,\beta)$, known as the power coefficient, quantifies the aerodynamic efficiency of the turbine. It is contingent upon the ratio λ, which represents the relationship between the blade tip speed and the wind speed, as well as the orientation angle β. The ratio λ can be articulated using the subsequent equation [22],[23]:

$$\lambda = \frac{R.\Omega_t}{V} \quad (3)$$

Ω_t : The turbine speed of rotation.
R : The length of a blade.
The model of the power coefficient is given by the expression

in (4) [24]-[26]:

$$C_p(\lambda,\beta) = Asin\left(\frac{\pi(\lambda+0.1)}{14.34-0.3(\beta-2)}\right) - B \quad (4)$$

With

$$\begin{cases} A = 0.35 - 0.0167(\beta - 2) \\ B = 0.00184(\lambda - 3)(\beta - 2) \end{cases} \quad (5)$$

Figure 1 illustrates the progression of C_p in relation to λ for different β values. This results in a maximum power coefficient of 0.35 for λ equal to 7.2, with β at 2°. The system will deliver optimal electrical power by configuring β and λ to their ideal levels.

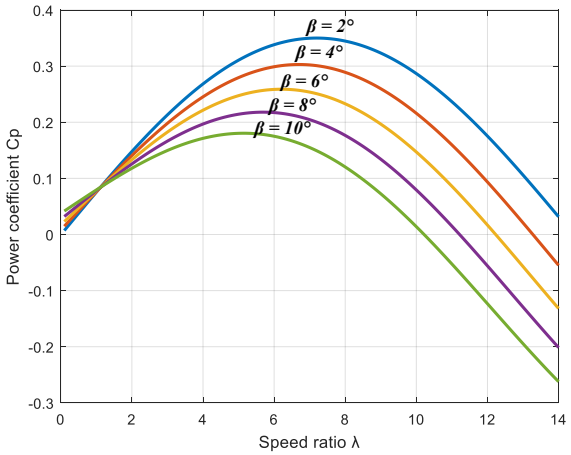


Fig 1. Power coefficient as a function of speed ratio

Aerodynamic torque is generated when aerodynamic power is converted into mechanical power, as illustrated in [27]:

$$T_t = \frac{P_t}{\Omega_t} = \frac{1}{2.\Omega_t}.\rho.S.C_p(\lambda,\beta).v^3 \quad (6)$$

The gearbox, positioned between the wind turbine and the generator, is designed to adjust the turbine speed Ω_t, which is relatively low, to the speed required by the generator Ω_m. This is given by the following two mathematical equations [28]:

$$\begin{cases} T_m = \frac{T_t}{G} \\ \Omega_m = G.\Omega_t \end{cases} \quad (7)$$

The shaft comprises a mass that reflects the inertia of the turbine rotor, which supports the blades and hub, together with a minor inertia associated with the generator rotor. The proposed mechanical model indicates that the overall inertia J comprises the generator's inertia J_g and the turbine's inertia J_t, which is applied to the generator's rotor [28]:

$$J = \frac{J_t}{G^2} + J_g \quad (8)$$

J_t: the turbine's inertia.

J_g : the DFIG's inertia.

The mechanical speed Ω_m is influenced by the mechanical torque applied to the generator rotor T_m, which is the result of an electromagnetic torque generated by the generator T_{em} and a torque of viscous friction ($f_v.\Omega_m$), as indicated in [26]:

$$J.\frac{d\Omega_m}{dt} = T_m - T_{em} - f_v.\Omega_m \quad (9)$$

The prior relations were used to construct the block diagram of the mechanical component of the wind system (Fig.2).

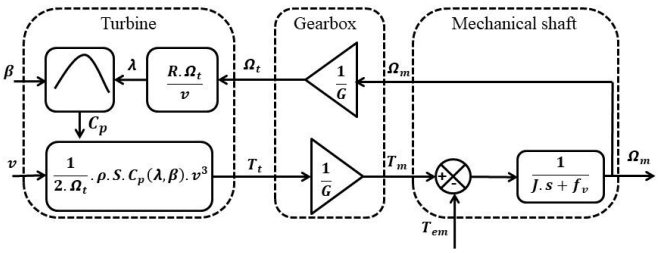


Fig 2. Block diagram of the turbine model

III. DFIG MODELING

A. *Modeling in abc reference frame*

To exemplify this method, a star-connected DFIG has been examined, as explained in [29],[30]. Which the stator voltages are unknown [31]. Where the stator and rotor voltages are given by:

$$\begin{cases} [v_s] = -[R_s].[i_s] + \frac{d[\Phi_{sn}]}{dt} \\ [v_r] = -[R_r].[i_r] + \frac{d[\Phi_r]}{dt} \end{cases} \quad (10)$$

With

$[R_s]$ is the matrix of stator resistances.
$[i_s]$ is the vector of the stator currents.
$[i_r]$ is the vector of the rotor currents.
$[v_s]$ is the vector of the stator voltages.
$[v_r]$ is the vector of the rotor voltages.
$[R_r]$ is the matrix of rotor resistances.

The stator and rotor flux are calculated using:

$$\begin{cases} [\Phi_s] = -[L_{ss}].[i_s] - [L_{sr}].[i_r] \\ [\Phi_r] = -[L_{rs}].[i_s] - [L_{rr}].[i_r] \end{cases} \quad (11)$$

With

$$[\Phi_{sn}] = [T_n].[\Phi_s] \quad (12)$$

$$[T_n] = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} \quad (13)$$

$$[L_{rs}] = [L_{sr}]^t \quad (14)$$

Where
$[L_{ss}]$ is the matrix of the stator inductances.
$[L_{rr}]$ is the matrix of the rotor inductances;
$[L_{rs}]$ is the rotor-stator mutual inductances;

We define stator currents and flux of two independent components(i_{as} , i_{bs}) and(Φ_{as} , Φ_{bs}) by (15).

$$\begin{cases} [i_s] = [B_n][i_{abs}] \\ [\Phi_{abs}] = [A_n][\Phi_s] \end{cases} \quad (15)$$

With

$$[A_n] = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \quad (16)$$

$$[B_n] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} \quad (17)$$

The stator and rotor fluxes are defined by (18),

$$\begin{cases} [\Phi_{abs}] = -[L_{sn}][i_{abs}] - [L_{srn}][i_r] \\ [\Phi_r] = -[L_{rsn}][i_{abs}] - [L_{rr}][i_r] \end{cases} \quad (18)$$

With

$$\begin{cases} [L_{sn}] = [A_n][L_{ss}][B_n] \\ [L_{srn}] = [A_n][L_{sr}] \\ [L_{rsn}] = [L_{rs}][B_n] \end{cases} \quad (19)$$

Thus, we obtain the equation of stator and rotor currents.

$$\begin{cases} [i_{abs}] = [C_{sn}]([\Phi_{abs}] - [L_{srn}][L_{rr}]^{-1}[\Phi_r]) \\ [i_r] = [C_{rn}]([\Phi_r] - [L_{rsh}][L_{sn}]^{-1}[\Phi_{abs}]) \end{cases} \quad (20)$$

With

$$\begin{cases} [C_{sn}] = -([L_{sn}] - [L_{srn}][L_{rr}]^{-1}[L_{rsn}])^{-1} \\ [C_{rn}] = -([L_{rr}] - [L_{rsn}][L_{sn}]^{-1}[L_{srh}])^{-1} \end{cases} \quad (21)$$

The mechanical speed and the electromagnetic torque are given by.

$$\begin{cases} \frac{d\Omega_m}{dt} = \frac{1}{J}(T_m - T_{em} - f_v\Omega_m) \\ T_{em} = p[i_s]^t \frac{\partial[L_{sr}]}{\partial\theta}[i_r] \end{cases} \quad (22)$$

p, θ are respectively the poles pairs number and mechanical angle.

B. Modeling in d-q reference frame

The electrical equations of DFIG in the (d-q) reference frame are written by:

$$\begin{cases} V_{ds} = -R_s.i_{ds} - \omega_s.\varphi_{qs} + \frac{d\varphi_{ds}}{dt} \\ V_{ds} = -R_s.i_{ds} - \omega_s.\varphi_{qs} + \frac{d\varphi_{ds}}{dt} \\ V_{dr} = -R_r.i_{dr} - \omega_r.\varphi_{qr} + \frac{d\varphi_{dr}}{dt} \\ V_{qr} = -R_r.i_{qr} + \omega_r.\varphi_{dr} + \frac{d\varphi_{qr}}{dt} \end{cases} \quad (23)$$

(V_{ds}, V_{qs}) and (V_{dr}, V_{qr}) are respectively the components of stator and rotor voltages vector, $(i_{\mathrm{ds}}, i_{qs})$ and (i_{dr}, i_{qr}) are respectively the components of stator and rotor currents vector.

The equations for the flux in (d-q) reference frame are expressed as follow:

$$\begin{cases} \varphi_{ds} = -L_s.i_{ds} - L_m.i_{dr} \\ \varphi_{qs} = -L_s.i_{qs} - L_m.i_{qr} \\ \varphi_{dr} = -L_m.i_{ds} - L_r.i_{dr} \\ \varphi_{qr} = -L_m.i_{qs} - L_r.i_{qr} \end{cases} \quad (24)$$

With

$L_s = L_{ms} + L_{ls}$ is the stator inductance.

$L_r = L_{mr} + L_{lr}$ is the rotor inductance.

$L_m = \frac{3}{2}L_{ms}$ is the mutual inductance.

L_{ls} , L_{ms} are respectively the leakage and magnetizing inductance of the stator windings.

L_{lr} , L_{mr} are respectively the leakage and magnetizing inductance of the rotor windings;

In the stationary (d-q) reference frame, the state model of the DFIG is represented by (25),

$$\begin{cases} \frac{di_{ds}}{dt} = -\frac{R_s}{L_s\sigma}i_{ds} + ai_{qs} + \frac{L_mR_r}{L_sL_r\sigma}i_{dr} + \omega\frac{L_m}{L_s\sigma}i_{qr} + U_{drs} \\ \frac{di_{qs}}{dt} = -ai_{ds} - \frac{R_s}{L_s\sigma}i_{qs} - \omega\frac{L_m}{L_s\sigma}i_{dr} + \frac{L_mR_r}{L_sL_r\sigma}i_{qr} + U_{qrs} \\ \frac{di_{dr}}{dt} = \frac{L_mR_s}{L_sL_r\sigma}i_{ds} - b\frac{L_m}{L_r}i_{qs} - \frac{\beta R_r}{L_r}i_{dr} - ci_{qr} + U_{dsr} \\ \frac{di_{qr}}{dt} = b\frac{L_m}{L_r}i_{ds} + \frac{L_mR_s}{L_sL_r\sigma}i_{qs} + ci_{dr} - \frac{\beta R_r}{L_r}i_{qr} + U_{qsr} \end{cases} \quad (25)$$

With

$\sigma = 1 - \frac{L_m^2}{L_sL_r}$; $\beta = \left(1 + \frac{L_m^2}{L_sL_r\sigma}\right)$;

$a = \left(\omega_s + \frac{L_m^2}{L_sL_r\sigma}\omega\right)$; $b = \left(\omega_s\left(\frac{1}{\sigma} - \beta\right) + \beta\omega\right)$;

$c = \left(\omega_s\left(\frac{L_m^2}{L_sL_r\sigma} - \beta\right) + \beta\omega\right)$;

$U_{drs} = \frac{L_m}{L_sL_r\sigma}V_{dr} - \frac{1}{L_s\sigma}V_{ds}$; $U_{qrs} = \frac{L_m}{L_sL_r\sigma}V_{qr} - \frac{1}{L_s\sigma}V_{qs}$;

$U_{dsr} = \frac{L_m}{L_sL_r\sigma}V_{ds} - \frac{\beta}{L_r}V_{dr}$; $U_{qsr} = \frac{L_m}{L_sL_r\sigma}V_{qs} - \frac{\beta}{L_r}V_{qr}$;

The electromagnetic torque is expressed by (26),

$$T_{em} = \frac{3}{2}p.L_m(i_{qr}.i_{ds} - i_{dr}.i_{qs}) \quad (26)$$

The active and reactive power of the stator and rotor are defined by:

$$\begin{cases} P_s = V_{ds}i_{ds} + V_{qs}i_{qs} \\ Q_s = V_{qs}i_{ds} + V_{ds}i_{qs} \end{cases} \quad (27)$$

IV. RESULTS AND DISCUSSION

The entire system simulations were conducted using MATLAB software. The wind-turbine and DFIG parameters are given in TABLE I and TABLE II respectively, while the wind turbine electromechanical values for abc and d-q reference frames in TABLE III and TABLE IV respectively.

The wind speed, speed ratio, power coefficient and mechanical speed are shown in Fig.3, Fig.4, Fig.5, and Fig.6 respectively. Fig.7, Fig.8 and Fig.9 illustrate the stator flux, currents, and voltages in the abc reference frame, respectively. Similarly, Fig.10, Fig.11 and Fig.12 depict the stator flux, currents, and voltages in the d-q reference frame.

TABLE I. DFIG Parameters

Rated power	Ps = 4000W
Mutual inductance	L_m = 0.2580 H
Stator inductance	L_s = 0.2740 H
Rotor inductance	L_r = 0.3030 H
Stator resistance	R_s = 4.85 Ω
Rotor resistance	R_r = 3.805 Ω
Number of pole pairs	P = 2
Moment of the inertia	J = 0.045 kg. m²
Viscous friction	f_v = 0.0038 kg.m².s^{-1}

TABLE II. Wind Turbine Parameters

Rated power	Ps = 4000W
Moment of the inertia	J = 0.31125 kg.m²
Wind turbine radius	R = 3m
Gear box ratio	G = 5.4
Number of blades	n_p = 3
Nominal wind speed	*v* = 12 m/s

TABLE III. Wind turbine electromechanical values for abc reference

Parameters	Maximum value	Minimum value
Wind speed (m/s)	14.7	12.4
Mechanical speed (rad/s)	164.7	157
Stator flux (wb)	0.67	0.40
Stator current (A)	10.80	9.01
Stator voltage (V)	261	163

TABLE IV. Wind turbine electromechanical values for d-q reference

Parameters	Maximum value	Minimum value
Wind speed (m/s)	14.7	12.4
Mechanical speed (rad/s)	164.7	157
Stator flux (wb)	0.18	0.17
Stator current (A)	10.41	10.07
Stator voltage (V)	252	241

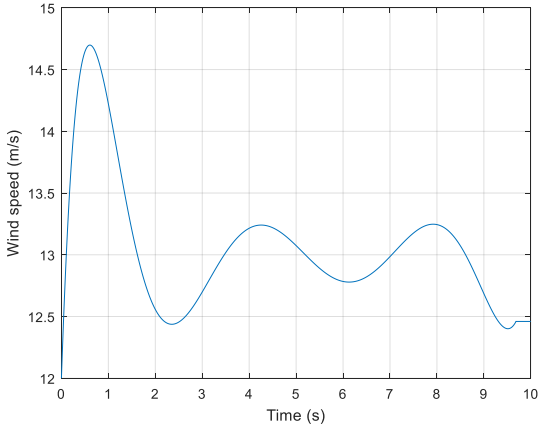


Fig 3. Wind speed

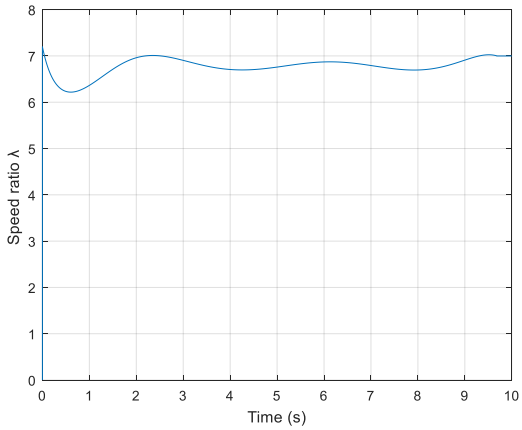


Fig 4. Speed ratio

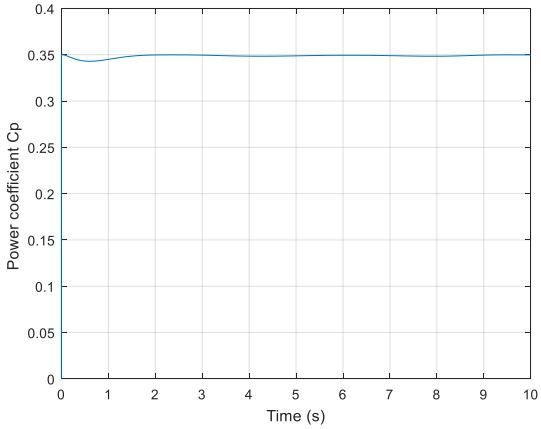


Fig 5. Power coefficient

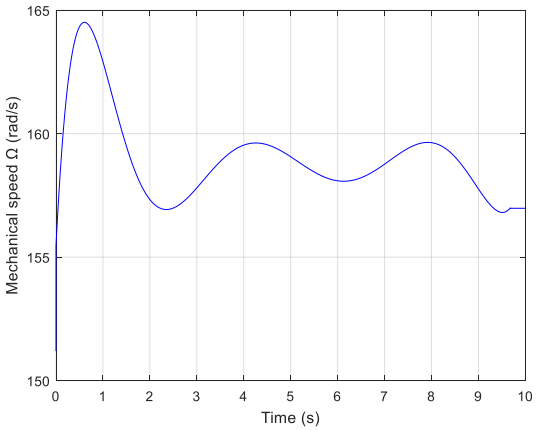


Fig 6. Mechanical speed

A. Simulation results: abc reference frame

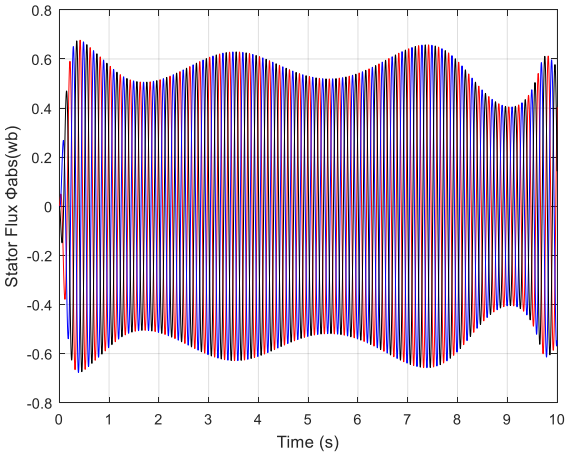


Fig 7. Stator flux

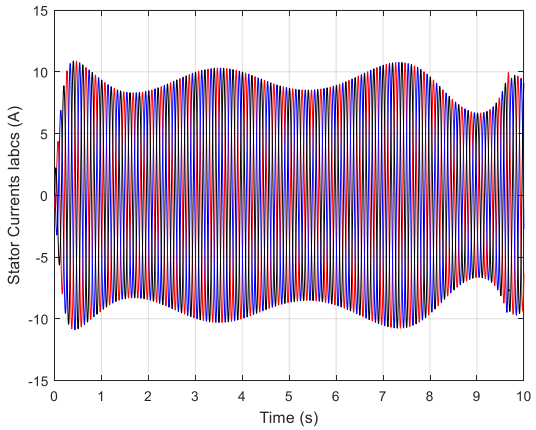


Fig 8. Stator currents

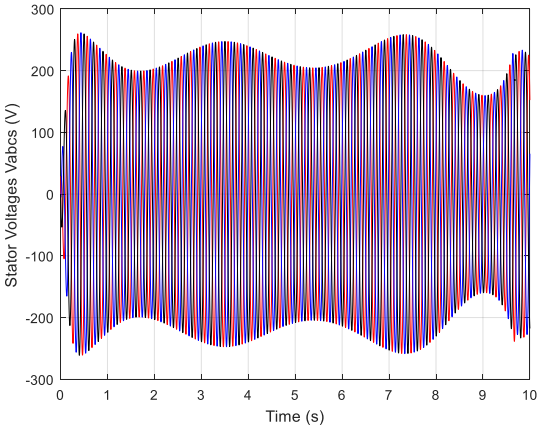


Fig 9. Stator voltages

B. Simulation results: d-q reference frame

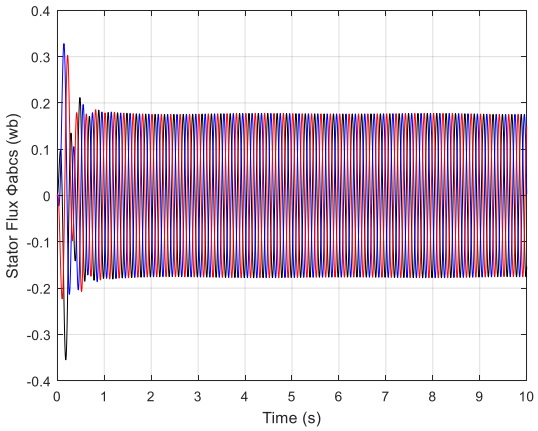


Fig 10. Stator flux

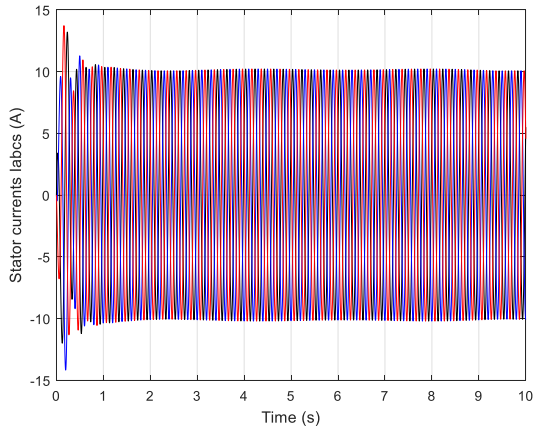


Fig 11. Stator currents

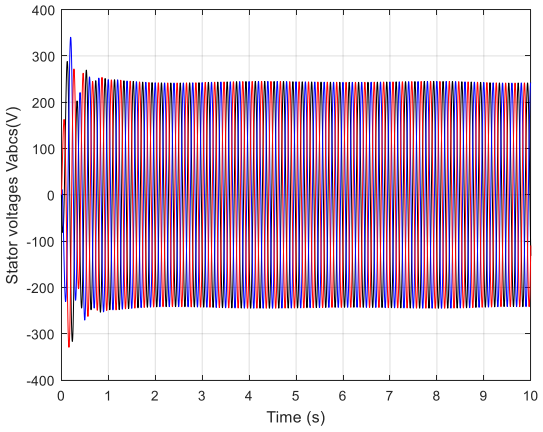


Fig 12. Stator voltages

The simulation results reveal distinct behaviors in the abc and d-q reference frames. In the abc frame, the currents and voltages dynamically follow variations in rotor speed, reflecting real-time system behavior. This makes the abc frame particularly useful for capturing detailed dynamics, analyzing transient phenomena, and studying harmonics. Conversely, the d-q frame simplifies the representation of the system by maintaining steady-state values for currents and voltages, even under varying rotor speeds. This provides an efficient framework for steady-state studies and facilitates the design of control systems. The comparison underscores the utility of the abc frame in capturing detailed dynamics and the efficiency of the d-q frame in steady-state analysis and controller development.

This distinction highlights the importance of selecting the appropriate reference frame based on the study's objectives. For dynamic performance evaluation under varying conditions, the abc frame is essential. Meanwhile, the d-q frame offers a computationally efficient alternative for steady-state analysis and control design.

V. CONCLUSION

The comparative study of the abc and d-q reference frames for modeling and simulating DFIG-based wind turbines demonstrates that each frame has unique advantages and limitations. The abc frame is ideal for capturing dynamic changes and transient phenomena, making it more suitable for detailed phase-by-phase analysis studies. On the other hand, the d-q frame simplifies the analysis by transforming the variables into a steady-state representation, which is advantageous for designing and analyzing control strategies. The choice of the reference frame depends on the study's specific objectives. For transient analysis and harmonic investigations, the abc frame is recommended, while for steady-state analysis and control design, the d-q frame offers significant computational and analytical benefits.

REFERENCES

[1] H. Benbouheni , "Application of DPC and DPC-GA to the dual-rotor wind turbine system with DFIG", International Journal of Robotics and Automation, vol.10, no .3, pp.224-234, Sep. 2021, doi: 10.11591/ijra.v10.i3.pp224-234.

[2] R. Mahroug, M. Matallah, and S. Abudura, "Modeling of wind turbine based on dual DFIG generators," International Journal of Power Electronics and Drive Systems, vol. 13, no. 2, pp. 1170–1185, Mar.2022, doi: 10.11591/ijpeds.v13.i2.pp1170-1185.

[3] M. Yessef, B. Bossoufi, M. Taoussi, A. Lagrioui, and H. Chojaa, "Improved Hybrid Control Strategy of the Doubly-Fed Induction Generator Under a Real Wind Profile, "Digital Technologies and Applications. ICDTA 2021. Lecture Notes, in Networks and Systems, vol 211, pp. 1279-1290, 2021, doi: 10.1007/978-3-030-73882-2_117 2_117.

[4] M. A. Soomro, Z. A. Memon, M. Kumar, M. H. Baloch,"Wind energy integration: Dynamic modeling and control of DFIG based on super twisting fractional order terminal sliding mode controller".Energy Reports,Volume 7,2021,Pages 6031-6043, doi:10.1016/j.egyr.2021.09.022.

[5] D. Maradin,"Advantages and disadvantages of renewable energy sources utilization". International Journal of Energy Economics and Policy, 2021, vol. 11, no. 3, pp. 176-183. doi: 10.32479/ijeep.11027.

[6] S. Sarika , S. A. J. Mary," High order sliding mode observer based fault-tolerant control scheme for DFIG in wind energy applications with model predictivecontroller", Automatika, 65,no.3, 793-802,2024, doi: 10.1080/00051144.2024.2308318.

[7] S.M.Alwash, O.Q.J. Al-Thahab, S.F. Alwash, "Modeling and control strategies for DFIG in wind turbines: A comparative analysis of SPWM, THIPWM, and SVPWM techniques". Journal Européen des Systèmes Automatisés, Vol. 56, No. 6, pp. 963-972. (2023). doi :10.18280/jesa.560607.

[8] S. Louarem, F.Z. Kebbab, H. Salhi, H. Nouri," A comparative study of maximum power point tracking techniques for a photovoltaic grid-connected system". Electrical Engineering & Electromechanics, no. 4, pp. 27-33. 2022, doi: 10.20998/2074-272X.2022.4.04.

[9] K. Manikandan, S. Sasikumar, R. Arulraj," A novelty approach to solve an economic dispatch problem for a renewable integrated micro-grid using optimization techniques". Electrical Engineering & Electromechanics, no. 4, pp. 83-89. 2023, doi: 10.20998/2074-272X.2023.4.12.

[10] M. Mazare, M. Taghizadeh, P. Ghaf-Ghanbari, "Fault tolerant control of wind turbines with simultaneous actuator and sensor faults using adaptive time delay control". Renew Energy. 174:86–101. 2021, doi:10.1016/j.renene.2021.04.077.

[11] I. Khan and al. "Dynamic Modeling and Robust Controllers Design for Doubly Fed Induction Generator-Based Wind Turbines under Unbalanced Grid Fault Conditions". Energies 2019, 12(3), 454, doi: 10.3390/en12030454.

[12] N. k. Mishra, Z. Husain, A. Iqbal:"Modeling and analysis of novel six-phase DFIG through asymmetrical winding structure for disperse generation". WILEY International Transactions on Electrical Energy Systems, 30(12), DOI:10.1002/2050-7038.12649.

[13] M. K addache, S. Drid, A. Khemis, D. Rahem, L. Chrifi-Alaoui, "Maximum power point tracking improvement using type-2 fuzzy controller for wind system based on the double fed induction generator". Electrical Engineering & Electromechanics, no. 2, pp. 61-66. 2024, doi: 10.20998/2074-272X.2024.2.09.

[14] L. Amira, B. Tahar and M. Abdelkrim, "Sliding Mode Control of Doubly-fed Induction Generator in Wind Energy Conversion System," 2020 8th International Conference on Smart Grid (icSmartGrid), Paris, France, 2020, pp. 96-100, doi: 10.1109/icSmartGrid49881.2020.9144778.

[15] D. R. Karthik, N. S. Manjarekar, S. M. Kotian, "Computation of steady-state operating conditions of a DFIG-based wind energy conversion system considering losses", Electrical Engineering, 105, 3, pp.1825-1838, 2023. doi:10.1007/s00202-023-01766-x

[16] B. Desalegn, D. Gebeyehu, B. Tamrat, "Smoothing electric power production with DFIG-based wind energy conversion technology by employing hybrid controller model", Energy Reports, 10.1016/j.egyr.2023.06.004, 10, pp.38-60, 2023.

[17] A. Khan, H. A. Khalid, M. Uzair Khalid and M. Khan, "Dynamic Modeling, Operations, and Steady-State Analysis of DFIG based

Nordex S77 Wind Turbine," 2022 International Conference on Electrical Engineering and Sustainable Technologies (ICEEST), Lahore, Pakistan, 2022, pp. 1-6, doi: 10.1109/ICEEST56292.2022.10077858.

[18] D. Kouchih and R. Hachelaf, "Theoretical and Experimental Analysis of the Squirrel Cage Induction Generators Operating under Unbalanced Conditions" , Research Square, vol. 1, no. 1,pp. 1–12, Dec. 2022, doi: 10.21203/rs.3.rs-2347428/v1.

[19] H. Alphan, "Modelling potential visibility of wind turbines: A geospatial approach for planning and impact mitigation," Renewable and Sustainable Energy Reviews, vol 152, no. 3, pp. 270-288, 2021, doi: 10.1016/j.rser.2021.111675.

[20] M. Allouche, et al, "A Novel fuzzy Control Strategy for Maximum Power Point Tracking of Wind Energy Conversion System," International Journal of Smart Grid, vol. 3, no 3, pp. 120-127, 2019.

[21] A. Nid, S. Sayah, A. Zebar, "Power fluctuation suppression for grid connected permanent magnet synchronous generator type wind power generation system". Electrical Engineering & Electromechanics, no. 5, pp. 70-76. 2024,doi: 10.20998/2074-272X.2024.5.10

[22] Y. Mousavi, *et al*., "Maximum power extraction from wind turbines using a fault-tolerant fractional-order nonsingular terminal sliding mode controller. Energies. vol.14, no.18, 5887. 2021 , doi:10.3390/en14185887

[23] O. Barambones, *et al*. "Real time observer and control scheme for a wind turbine system based on a high order sliding modes". J Franklin Inst. 358, no11,5795–5819. 2021, doi:10.1016/j.jfranklin.2021.05.022

[24] A. Fekih, S. Mobayen, C-C. Chen, "Adaptive robust fault-tolerant control design for wind turbines subject to pitch actuator faults". Energies. vol. 14, no. 6, 1791. 2021, doi:10.3390/en14061791

[25] H. Chojaa, A. Derouich, S. E. Chehaidia, O. Zamzoum, M. Taoussi, and H. Elouatouat, "Integral sliding mode control for DFIG based WECS with MPPT based on artificial neural network under a real wind profile," in Energy Reports, vol. 7, pp. 4809-4824, 2021,doi:10.1016/j.egyr.2021.07.066.

[26] N. Tidjani, A. Guessoum and D. Ounnas, "Permanent-Magnet Synchronous Generator Wind Turbine based on Takagi-Sugeno Fuzzy Models," 2019 7th International Renewable and Sustainable Energy Conference (IRSEC), Agadir, Morocco, 2019, pp. 1-6, doi:10.1109/IRSEC48032.2019.9078253.

[27] Y. Mousavi, G. Bevan, I. B. Kucukdemiral, and A. Fekih, "Sliding mode control of wind energy conversion systems: Trends and applications," Renewable and Sustainable Energy Reviews, vol. 167, 2022, doi: 10.1016/j.rser.2022.112734.

[28] X. Kong,L. Ma, X. Liu, M.A. Abdelbaky, and Q. Wu, "Wind turbine control using nonlinear economic model predictive control over all operating regions," Energies, vol. 13, no. 1, 2020, doi: 10.3390/en13010184.

[29] Y. Diboune and D. Kouchih, "Investigation on the modeling of doubly-fed induction generators for wind turbine", 4th *International Conference on Electronics and Electrical Engineering IC3E'2024*, Nov. 2024, Bouira, Algeria.

[30] Y. Diboune, R. Hachelaf and D. Kouchih, "Analysis of fault tolerant controlled doubly fed induction generators ", 6th *International Conference on Electrical Engineering and Control Applications ICEECA'2024*, Nov. 2024, Khenchela, Algeria.

[31] Y. Diboune, R. Hachelaf and D. Kouchih, "Theoretical and experimental analysis of unbalanced doubly fed induction generators", International Journal of Robotics and Automation IJRA, vol. 13, no. 4, pp. 476-484, Dec. 2024, doi: 10.11591/ijra.v13i4.pp476-484.

Improving Power Quality in Microgrids by Optimizing Energy Management Using Genetic Algorithm

Sabrine Morakeb
Electrical Engineering Department
Laboratory of Electrical Engineering and Renewable Energy, University of Souk Ahras
Souk Ahras, Algeria
s.morakeb@univ-soukahras.dz

Samira Boumous
Electrical Engineering Department
Laboratory of Electrical Engineering and Renewable Energy, University of Souk Ahras
Souk Ahras, Algeria
samira.boumous@univ-soukahras.dz

Zouhir Boumous
Electrical Engineering Department
Laboratory of Electrical Engineering and Renewable Energy, University of Souk Ahras
Souk Ahras, Algeria
zohir.boumous@univ-soukahras.dz

Mohamed Abderrazak Ghellouci
Electrical Engineering Department
Laboratory of Electrical Engineering and Renewable Energy, University of Souk Ahras
Souk Ahras, Algeria
m.ghellouci@univ-soukahras.dz

***Abstract*—Microgrids with diverse energy sources, including solar panels, wind turbines, battery storage, and diesel generators, present a promising solution for sustainable and reliable power generation. However, ensuring efficient energy sharing among these sources while maintaining power quality remains a critical challenge. This study focuses on optimizing energy management in hybrid microgrids through a genetic algorithm (GA). The primary objective is to achieve an optimal energy distribution among the available sources, considering their capacities, availability, and operational constraints. The proposed GA-based approach dynamically allocates power to balance supply and demand, minimize energy losses, and improve system efficiency. Simulation results demonstrate that the method effectively enhances energy sharing, reduces reliance on diesel generators, and maintains high power quality metrics. The findings highlight the potential of GA in managing complex energy interactions within microgrids, supporting their stability, sustainability, and cost-effectiveness.**

Keywords—microgrid optimization, energy management systems, genetic algorithm, renewable energy resources.

I. INTRODUCTION

Managing and securing enough energy has long been a major challenge for contemporary civilization [1]. Energy management is becoming even more crucial due to climate demands for the expanding use of renewable energy technologies, which frequently provide energy in a highly variable and unexpected style [2].

Microgrids (MGs), also known as micro-energy networks, have received substantial interest as a new paradigm in energy systems due to the swift expansion of renewable energy and the broad use of dispersed energy sources [3], [4]. The DERs must be operated to provide the maximum advantages, sustain a consistent supply, and assure the microgrid's economic viability. In grid-connected mode, the major purpose is to trade with the grid and optimally employ available DERs to optimize the microgrid's profitability. In islanded mode, however, the emphasis turns to maximizing the time of independent operation without relying on the external grid. All of this is performed using an energy management system that allocates the resources according to predefined goals utilizing various optimization approaches [5].

Due to their dispersed energy supplies and loads, MGs are essential to modern energy systems because they provide reliable, sustainable, and efficient energy distribution. This is especially relevant in regions like rural or distant areas where conventional electrical networks face issues. Effective resource distribution between supply and demand in MGs depends on efficient control and energy management, which lowers operating costs and has a positive environmental impact [6].

Optimization approaches are crucial to achieving the balance required for effective microgrid management. Genetic algorithms have been efficiently applied to enhance microgrid designs by progressively refining system features using an evolutionary strategy. Microgrid systems have also made substantial use of linear programming, a traditional optimization approach offering mathematical decision-making frameworks. Each optimization approach is chosen based on how effectively it satisfies particular microgrid optimization objectives, such as cost reduction, renewable energy integration, and resilience building [7]. In this paper we used the genetic algorithm; this optimization algorithm is applied to determine the optimal energy management system of hybrid microgrids with different components. The system components considered for capacity optimization include solar photovoltaics, wind turbines, diesel generators, and battery energy storage.

The subsequent sections of the paper are structured in the following manner. The paper begins with the modeling of the MG system. This is followed by an outline of the metaheuristic-based strategy used to optimize the MG components. The subsequent section presents the simulation results and analysis-based energy management systems (EMSs). Finally, the paper concludes with a discussion of the findings and potential avenues for future research.

II. MODELING OF THE MG SYSTEM

The suggested hybrid microgrid system configuration (HMS) explored in this research is presented in Fig. 1. In this configuration, the PV system, wind system, and battery storage system are the three DC power components. On the contrary, the AC power components contain the diesel generator and the residential load. All the elements have a bidirectional communication link with the energy management system (EMS), so the EMS continually gets information on the condition of each component. In exchange, the EMS gives control signals to each component to ensure the HMS's best performance.

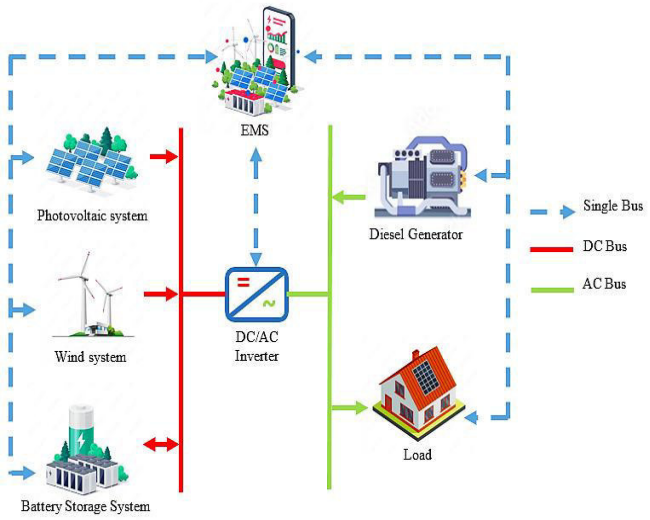


Fig. 1. The Microgrid System Configuration.

A. Photovoltaic system

PV's output power can be computed using equations (1) through (3). This model has taken into account how the temperature of the surrounding environment and solar radiation affect the output power [8]. The following are the Maximum Power Point (MPP) equations:

$$P_{pv} = V_{mmp} \times I_{mmp} \quad (1)$$

$$V_{mmp} = V_{mmp,ref} \times P_{v,oc}(T_c - T_{c,ref}) \quad (2)$$

$$I_{mmp} = I_{mmp,ref} \times I_{sc,ref}(T_c - T_{c,ref}) \quad (3)$$

Where P_{pv} is the panel's power, V_{mmp} is the potential voltage, $V_{mmp,ref}$ is V_{mmp} at standard condition, I_{mmp}, and $I_{sc,ref}$ is the panel's current and short circuit current at standard condition respectively, $P_{v,oc}$ is the open circuit (V/°c) temperature coefficients. $T_{c,ref}$ is the temperature of the panel at standard operational conditions that is equal to 250°c and T_c is the operational temperature of the panel that is calculated as follows:

$$T_{C(t)} = T_{a(t)} + \frac{NOCT-20}{800} \cdot G_T \quad (4)$$

Where, $T_{a(t)}$ is the ambient temperature (°C), NOCT (Normal Operational Cellular Temperature), for 500 W/m^2 of solar irradiation and temperature of 20°c is in the range of 40 to 46°c and G_T is the average daily solar irradiation (W/m^2).

B. Wind turbine

Wind energy is a readily available and sustainable energy source, and its cost has been steadily declining due to ongoing technological advancements [9]. The following is the formula for converting wind energy into fan power output for wind turbines:

$$P_W = \frac{1}{2}\rho\pi R^2 V^3 C_P \quad (5)$$

In this formula, ρ is the air density; R is the radius of the fan blade; V is the wind speed from the blade tip; C_P is the wind energy conversion efficiency, which is a function of the blade tip speed ratio λ and the blade pitch angle θ, as:

$$C_P = f(\theta, \lambda) \quad (6)$$

In this formula, the blade tip speed ratio λ is defined as:

$$\lambda = \frac{W_W R}{V} \quad (7)$$

Where W_W is the mechanical angular velocity of the fan (rad/s).

C. Diesel generator

The following formula can be used to determine how much is required to meet demand [10].

$$F_{DG}(t) = \left(a_{DG} \times P_{DG,gen}(t) + b_{DG} \times P_{DG,rat}\right) l/h \quad (8)$$

Where b_{DG} and a_{DG} represent coefficients of DG consumption curve and their respective values are b_{DG} = 0.08414 (l/KWh) and a_{DG}= 0.246(l/KWh).

DG total annual fuel consumption in liters (TAFCIL) is calculated using the following equation:

$$TAFCIL = \sum_{t=1}^{8760} F_{DG}(t) \quad (9)$$

D. Battery energy storage system

The excess electrical energy generated by the RES can be stored in storage devices such as batteries in the absence of a grid-connected network. The battery is charged if the energy generated (E_T) exceeds the load demand (E_L) [11]. The available battery bank's capacity at any given moment (t) throughout this process may be expressed as follows:

$$E_B(t) = E_B(t-1) \times (1-\sigma_i) + [E_T(t) - \frac{E_L(t)}{\eta_i}] \times \eta_B \quad (10)$$

If the load requirement exceeds the energy manufactured from RES, The electrical energy stored in the battery is useful to fulfill the load demand. In the battery discharge phenomenon, the battery bank capacity at time t is expressed as follows:

$$E_B(t) = E_B(t-1) \times (1-\sigma_i) - [\frac{E_L(t)}{\eta_i} - E_T(t)] \quad (11)$$

Where $E_B(t)$ and $E_B(t-1)$ are the available energy (kWh) at a time (t) and (t-1) respectively, σ_i is battery self-discharging rate, η_i is battery bank efficiency and η_B is the inverter efficiency

III. OPTIMIZATION ALGORITHM

This research applies GA optimization techniques to reduce microgrid power losses and voltage deviations by ensuring the best efficient use of solar, wind, battery storage, and diesel resources in a hybrid renewable power generating system.

A. *Genetic Algorithm*

John Holland invented genetic optimization in 1960. It is a kind of optimization algorithm that implements techniques inspired by natural and evolutionary processes to solve optimization problems. GA tackles both unconstrained and constrained nonlinear optimization problems by emulating the natural selection process evident in biological evolution. It is categorized as a collective intelligence algorithm and is distinguished by the presence of individuals. Each individual in the population signifies a potential solution to the existing situation. The population is constituted of a collective of these individuals. By applying a precise update approach, the collective will progressively undergo evolution toward the aim, ultimately converging to the best solution. Fig. 2 depicts how the GA algorithm works in this study.[15]

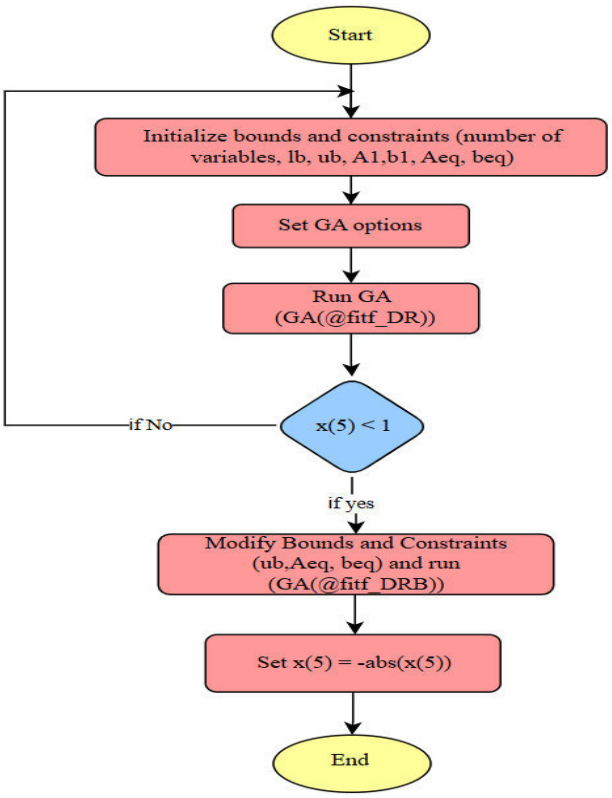


Fig. 2. Flowchart of GA Algorithms.

Just like the processes of evolution, this optimization technique contains four components:

- Inheritance;
- Mutation;
- Selection;
- crossover.

This is one of the most well-known methods of having merits like:

- Solving multiple problems;
- Being easy to understand;
- Being very precise;
- Being easy to integrate with other applications like MATLAB.

Despite all of its benefits, genetic algorithms have drawbacks. These drawbacks include:

- Being trapped in local optimal spots;
- Having a delayed reaction time [12].

This GA approach aims to optimize the energy supply and consumption of DERs for microgrids. This allows these facilities to satisfy their energy demand while also mitigating stability and power quality problems, reducing operating costs, and minimizing carbon emissions [13], [14].

IV. THE SIMULATION RESULTS AND ANALYSIS

We present and apply the suggested simultaneous optimization approach to the imbalanced three-phase system. In Fig. 3, the test system's line diagram is displayed.

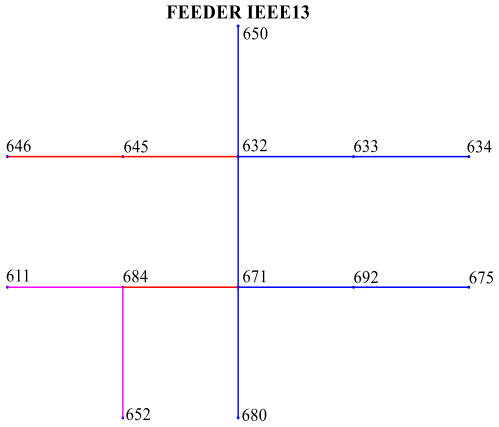


Fig. 3. The IEEE 13-bus distribution system's line diagram.

Fig. 4 illustrates the impact of a demand response strategy on energy consumption over 24 hours by comparing demand load before and after its implementation. Before the demand response, there are two sharp peaks. In contrast, the peaks are significantly reduced after implementing demand response, demonstrating effective peak shaving. Demand response strategies can significantly reduce energy consumption, especially during peak hours. This can lead to improved grid management, cost savings, and enhanced energy efficiency.

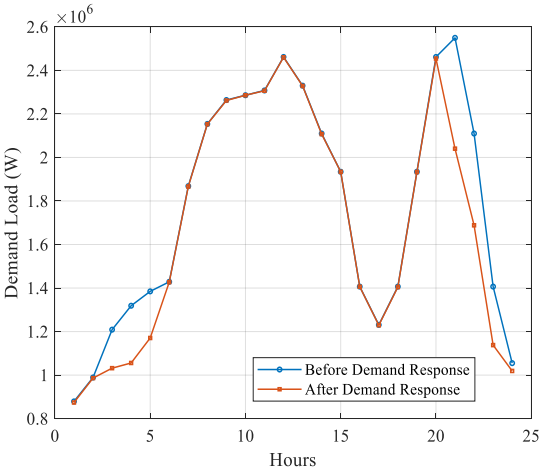


Fig. 4. The hourly load curve's response to DR programming.

The active power losses in a microgrid throughout a 24-hour period are contrasted in Fig. 5. Before optimal management, active losses exhibit significant variability and peaks throughout the 24 hours. Notably, there are sharp increases in losses, with the highest peak occurring between 08:00 PM and 10:00 PM, reaching nearly 10,000 W. Other smaller peaks are observed during earlier hours, particularly

around 09:00 AM and 11:00 AM. In contrast, the losses after optimal management are consistently minimized across all hours, indicating a dramatic reduction in energy losses. This highlights the effectiveness of the optimal management strategy in stabilizing and reducing active losses, even during hours previously characterized by high variability and peaks. Overall, the figure underscores the critical impact of implementing optimal management strategies to enhance system efficiency by significantly lowering active power losses.

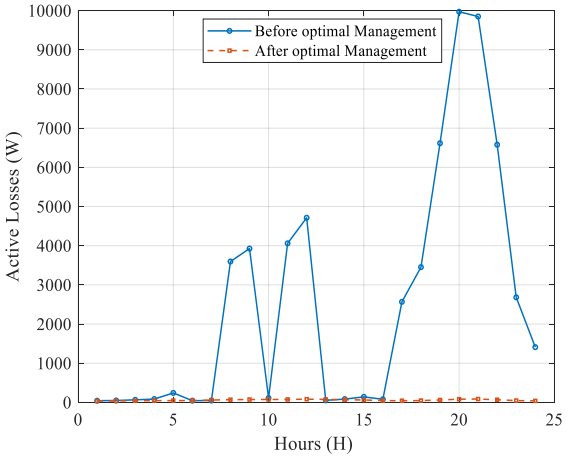


Fig. 5. The microgrid losses before and after the simultaneous GA optimization.

Before implementing the optimization approach in Fig. 6, the system exhibited significant voltage fluctuations, with a prominent peak around 08:00 PM, and after implementing the optimal management strategy, the voltage deviation was significantly reduced, resulting in a smoother profile with much lower overall fluctuations. This suggests that the GA optimization successfully addressed the factors causing voltage instability in the system. While the post-optimization deviation is lower, there is still room for further improvement, indicating the potential for fine-tuning the management strategy to achieve even greater voltage stability.

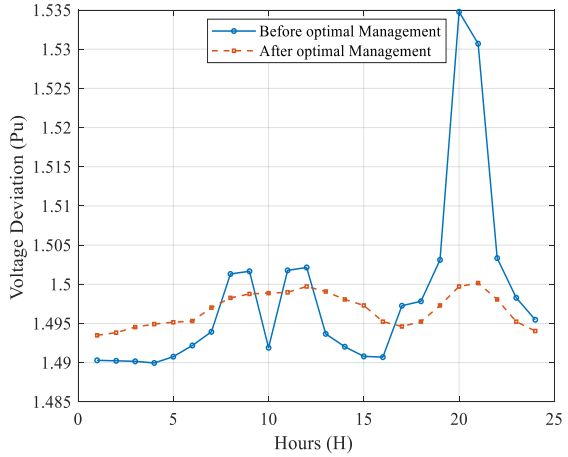


Fig. 6. The total voltage deviation before and after the GA optimization.

Fig. 7 shows PV power generation increasing after daybreak at 05:00 AM, peaking between 09:00 AM and 01:00 PM, and then declining in the evening. An odd peak occurs at 08:00 PM, before decreasing to zero overnight. The GA dynamically adjusts the PV system's operation to maximize power generation during daylight hours and minimize it during nighttime. This optimized operation enhances the microgrid's efficiency.

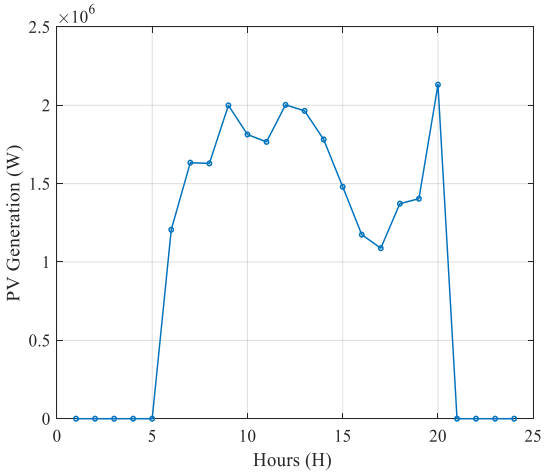


Fig. 7. PV power generation.

In the context of microgrid energy management, Fig. 8 displays wind turbine energy generation optimized by a GA. The generation fluctuates throughout the day, with high-energy output during the early hours around 12 am to 4 am, and a notable drop after 5 am. Between 6 am and 6 pm, the output remains variable but generally lower, reflecting changes in wind availability. A significant surge occurs in the evening around 8 pm, indicating a return to higher wind speeds or optimized turbine performance. These variations showcase the dynamic nature of wind energy and the GA's role in maximizing its contribution to microgrid energy management.

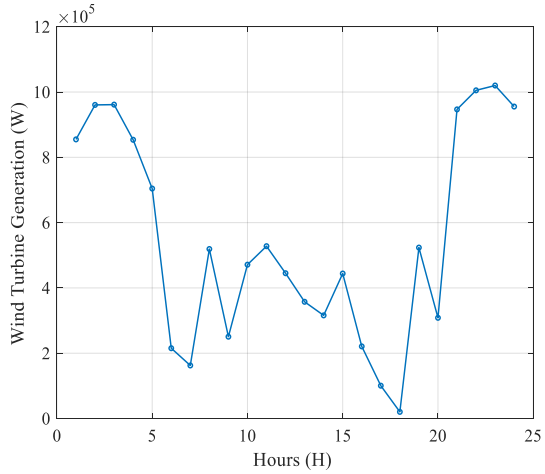


Fig. 8. Wind turbine power generation.

A microgrid energy management system's genetic algorithm optimizes a battery's power production over 24 hours, as displayed in Fig. 9. During some periods, the battery produces energy, with notable increases occurring in the early morning at 05:00 AM and 08:00 PM to 09:00 PM. When the battery is charging or not in use, the battery generation stays low or stays at zero for the remainder of the day, demonstrating effective energy management. This pattern illustrates how the GA helps balance supply and demand in the microgrid by optimizing battery utilization.

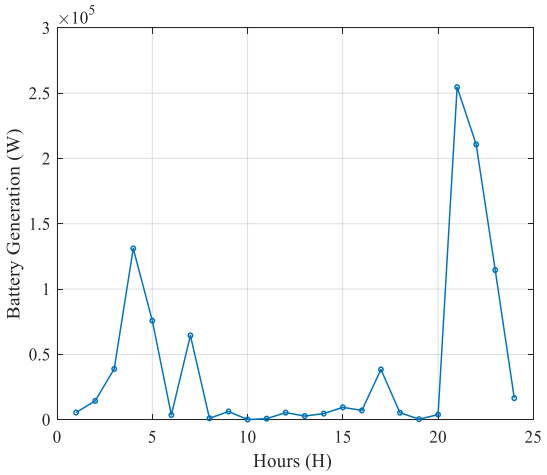


Fig. 9. Battery power generation.

Fig. 10 depicts a diesel generator's power production, which GA optimized in a microgrid management system. The GA dynamically modifies the generator's operation to adapt to current energy demands and the availability of alternative generating sources, producing a highly variable power generation. This flexible operation allows the diesel generator to serve as a reliable backup power source and to support the integration of renewable energy sources into the microgrid. This optimized operation enhances the microgrid's efficiency, stability, and environmental sustainability.

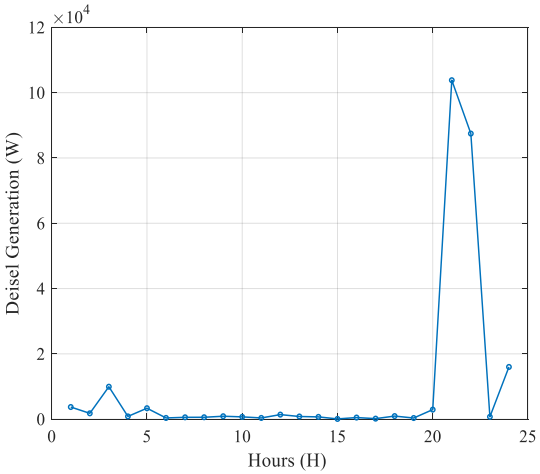


Fig. 10. Diesel generator power generation.

V. CONCLUSIONS

This study demonstrates the effectiveness of using a GA to optimize energy sharing in hybrid microgrids comprising solar energy, wind turbines, battery storage, and diesel generators. By dynamically allocating energy resources, the proposed approach ensures an efficient balance between supply and demand, minimizes reliance on fossil fuels, and enhances overall system performance. The results reveal significant improvements in energy sharing, reduced energy losses, and enhanced power quality. Moreover, the optimization framework promotes efficient battery utilization and extends the operational lifespan of critical components. The findings underscore the potential of GA-based energy management systems in addressing the complexities of hybrid microgrids, particularly in environments with high renewable energy penetration. This approach supports the broader goal of achieving sustainable, cost-effective, and reliable energy systems while mitigating the environmental impact of diesel-based generation. Future work could focus on integrating real-time data processing, considering economic constraints, and expanding the framework to accommodate larger, more complex energy systems.

REFERENCES

[1] H. M.Saleh and A. I.Hassan, "The challenges of sustainable energy transition: A focus on renewable energy," *Applied Chemical Engineering*, vol. 7, p. 2084, Apr. 2024, doi: 10.59429/ace.v7i2.2084.

[2] F. Carreras and H. Kirchsteiger, "An Empirical Approach to Optimize Nonlinear Problems of Domestic Energy Management Systems," *Environmental and Climate Technologies*, vol. 27, no. 1, 2023, doi: 10.2478/rtuect-2023-0023.

[3] Md. R. Hazari, S. Ahmad, E. Ogliari, C. Hossain, E. Jahan, and M. Mannan, "Optimal sizing of islanded microgrid using pelican optimization algorithm for Kutubdia Island of Bangladesh," *Electric Power Systems Research*, vol. 238, Oct. 2024, doi: 10.1016/j.epsr.2024.111088.

[4] Z. Yu, W. Zheng, K. Zeng, R. Zhao, Y. Zhang, and M. Zeng, "Energy optimization management of microgrid using improved soft actor-critic algorithm," *International Journal of Renewable Energy Development*, vol. 13, pp. 329–339, Feb. 2024, doi: 10.61435/ijred.2024.59988.

[5] S. Ahmad, M. Shafiullah, C. B. Ahmed, and M. Alowaifeer, "A Review of Microgrid Energy Management and Control Strategies," 2023. doi: 10.1109/ACCESS.2023.3248511.

[6] M. Cavus and A. Allahham, "Enhanced Microgrid Control through Genetic Predictive Control: Integrating Genetic Algorithms with Model Predictive Control for Improved Non-Linearity and Non-Convexity Handling," *Energies (Basel)*, vol. 17, p. 4458, Sep. 2024, doi: 10.3390/en17174458.

[7] D. Nemova, D. Rao, R. Singh, R. Bhardwaj, and S. Sharma, "Hybrid Energy Microgrids: A Comparative Study of Optimization Techniques," *E3S Web of Conferences*, vol. 511, p. 01018, Apr. 2024, doi: 10.1051/e3sconf/202451101018.

[8] F. Mostofi and H. Shayeghi, "Feasibility and optimal reliable design of renewable hybrid energy system for rural electrification in Iran," *International Journal of Renewable Energy Research*, vol. 2, no. 4, 2012.

[9] K. Gao, T. Wang, C. Han, J. Xie, Y. Ma, and R. Peng, "A review of optimization of microgrid operation," *Energies (Basel)*, vol. 14, no. 10, 2021, doi: 10.3390/en14102842.

[10] A. L. Bukar, "Optimal sizing of an autonomous photovoltaic/wind/battery/diesel generator microgrid using grasshopper optimization algorithm," *Solar Energy*, vol. 188, Jun. 2019, doi: 10.1016/j.solener.2019.06.050.

[11] S. Sharma *et al.*, "Modeling and sensitivity analysis of grid-connected hybrid green microgrid system," *Ain Shams Engineering Journal*, vol. 13, no. 4, 2022, doi: 10.1016/j.asej.2021.101679.

[12] M. Kiehbadroudinezhad, A. Merabet, A. G. Abo-Khalil, T. Salameh, and C. Ghenai, "Intelligent and Optimized Microgrids for Future Supply Power from Renewable Energy Resources: A Review," 2022. doi: 10.3390/en15093359.

[13] C. B. Moorthy and S. Sivamani, "A Concise Review on Applications of Bioinspired Algorithms in Electrical Power System," *Electric Power Components and Systems*, 2024, doi: 10.1080/15325008.2024.2319309.

[14] M. Khaleel, Z. Yusupov, N. Yasser, and H. J. El-Khozondar, "Enhancing Microgrid Performance through Hybrid Energy Storage System Integration: ANFIS and GA Approaches," *International Journal of Electrical Engineering and Sustainability (IJEES)*, vol. 1, no. 2, 2023.

[15] S. Li, W. Chen, S. Jain, D. Jung, and J. Lee, "Optimization of flow behavior models by genetic algorithm: A case study of aluminum alloy," Journal of Materials Research and Technology, vol. 31, pp. 3349–3363, Jul. 2024, doi: 10.1016/J.JMRT.2024.07.048.

Enhancing Efficiency of Three-Level Inverters by Minimizing Switching Losses

Sid-Ahmed Touil
Institute of Electrical and Electronic Engineering
University M'hamed Bougara.
Boumerdes, Algeria
s.touil@univ-boumerdes.dz

Hemza Gueddar
Electrical Engineering Department, University of Jijel
Jijel, Algeria
hamza01gueddar@gmail.com

Nidhal Cherrat
Dept. of Electrical Systems Engineering ISE University of Boumerdes
Laboratoire d'Automatique de Jijel-LAJ, University MSBY
Boumerdes, Algeria
n.cherrat@univ-boumerdes.dz

Tarek Bensidhoum
Smart Materials and Renewable Energy Laboratory
Department of Electromechanics, University of Mohamed El Bachir El Ibrahimi, El-Anasser ,
Bordj-Bou-Arreridj, Algeria.
tarek.bensidhoum@univ-bba.dz

Nasserdine Boudjerda
Renewable Energy Laboratory
University Mohammed Seddik Ben Yahia.
Jijel, Algeria
n_boudjerda@yahoo.fr

Badis Lekouaghet
Research Center In Industrial Technologies
Cheraga, Algiers, Algeria
easbadis@gmail.com

***Abstract*— This paper presents a generalized discontinuous pulse width modulation (DPWM) technique to reduce switching losses in the three-level inverters. By directly reducing the switching number per cycle, this innovative approach addresses a significant source of power losses in electronic converters. The proposed technique is evaluated through simulations and compared to the conventional Hybrid Space Vector Modulation (HSVM) method. The obtained results demonstrate a significant reduction in switching number while maintaining an acceptable total harmonic distortion (THD) in the output voltage, achieving optimal balance between efficiency and performance.**

Keywords—Three-phase inverters; space vector modulation; discontinuous PWM; switching losses.

I. INTRODUCTION

Inverters are controllable voltage sources in terms of frequency, amplitude and shape; they generate sine wave voltage with specific characteristics of frequency amplitude and harmonic content. According to their structures, they are reversible and need at least two current bidirectional switching cells. Depending on the type of application, inverters may require the addition of filtering devices. The main fields of application concern the variable speed drives in electric traction or in industrial applications moreover in addition to the connection of the novel renewable power system generation to the electricity distribution network.

In high power applications, the three-level inverters structure is more suitable compared to the classic two-level structure, as the output voltages and currents contain a much lower harmonic rate required by the international standards. Moreover there is no need for filters. The voltage across each switch is halved, and the chopping frequency of each component is lower. Furthermore, this structure causes less power loss, which requires less cooling equipment and smaller devices at lower costs. In this work, different discontinuous PWM techniques are presented and compared with the conventional PWM technique. Note that the hybrid space vector modulation (HSVM) is particularly studied in this work in both discontinuous and classical forms. The remainder of the paper is organized as follows:

Section II describes the Space Vector Modulation (SVM) and Hybrid SVM relationship, Section III report the various DPWM modulating waveforms generation, the simulation results are presented in Section IV; Finally, Section V concludes the paper.

II. SVM AND HYBRID SVM RELATIONSHIP

The SVM technique is a modulation method with high linearity performances and good spectral properties. It is based on a complex representation of the inverter output voltage and allows a direct calculation of the inverter switching times, therefore the obtained reference voltage vector, rotates in the complex plan [1-3] is given by:

$$\overrightarrow{V_{qd}^{*}} = V_q^{*} - jV_d^{*} \tag{1}$$

Where:

$$V_q^{*} = \frac{1}{3}\left(2V_a^{*} - V_b^{*} - V_c^{*}\right) \tag{2}$$

$$V_d^{*} = \frac{\sqrt{3}}{3}\left(V_c^{*} - V_b^{*}\right) \tag{3}$$

The rotating vector $\overrightarrow{V_{qd}^{*}}$ defines eight state vectors (V_0, V_1 …V_7); V_0 and V_7 are called zero state vectors and the six others define six sectors subdivided in sampling periods T_s in which $\overrightarrow{V_{qd}^{*}}$ is assumed constant. In each sector, two adjacent vectors are active during T_1 and T_2 respectively:

$$T_1 = T_s \frac{\sqrt{3}}{2}\left(V_d^{*}\cos\left(n\frac{\pi}{3}\right) + V_q^{*}\sin\left(n\frac{\pi}{3}\right)\right) \tag{4}$$

$$T_2 = -T_s \frac{\sqrt{3}}{2}\left(V_d^* \cos\left((n-1)\frac{\pi}{3}\right) + V_q^* \sin\left((n-1)\frac{\pi}{3}\right)\right) \quad (5)$$

n : sector number (n = 1… 6).

Zero state vectors V_7 and V_0 are active during k_0T_0 and $(1-k_0)T_0$ respectively, with ($0 \le k_0 \le 1$). The total duration T_0 is given by:

$$T_0 = T_s - T_1 - T_2 \quad (6)$$

For the classical SVM, k_0 = 0.5, therefore V_7 and V_0 have the same durations:

$$k_0 T_0 = (1-k_0)T_0 = T_0/2 \quad (7)$$

In practice, the SVM switching signals can be generated either by space vectors or by comparison of three reference signals with a triangular carrier, exactly in the same way of the sinusoidal modulation (SM); the only difference is in the reference signals. This last method is called hybrid SVM (HSVM) and the appropriate reference signals v_{abc}^{**} can be obtained by adding a zero-sequence signal to the sinusoidal ones as follows:

$$v_{abc}^{**} = v_{abc}^{*} + v_{zs}^{*} \quad (8)$$

Where, v_{abc}^{*} is the vector of reference signals for sinusoidal modulation (SM), i.e. wanted output voltages and v_{zs}^{*} is the zero sequence given by:

$$v_{zs}^{*} = -\left[(1-2k_0) + k_0 v_{\max}^{*} + (1-k_0) v_{\min}^{*}\right] \quad (9)$$

For the classical SVM, vectors V_0 and V_7 have equal durations (k_0 = 0.5), Where $v_{\min}^{*}$, v_{mid}^{*} and $v_{\max}^{*}$ are minimum, middle and maximum values of reference signals v_{abc}^{*}[1, 2].

III. Generalized Discontinuous PWM

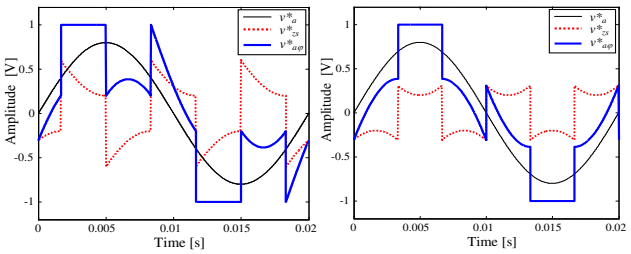
The principle of generalized discontinuous PWM (GDPWM) technique is based on the injection of a zero sequence v_{zs}^{*} to the sinusoidal references v_{abc}^{*} in the same way of HSVM. Where the only difference is in the derivation of the zero sequence. The main goals of this technique are:

- Reducing the number of switching by *1/3* symmetrically in the two half cycles: all switches are inactive during a phase angle of $\pi/3$. This is the purpose of this technique because it allows a reduction of the switching losses in the same ratio while conserving a symmetrical operation of all switches [4].
- Extending the linearity up to $2/\sqrt{3}$ as in the third harmonic injection and HSVM.

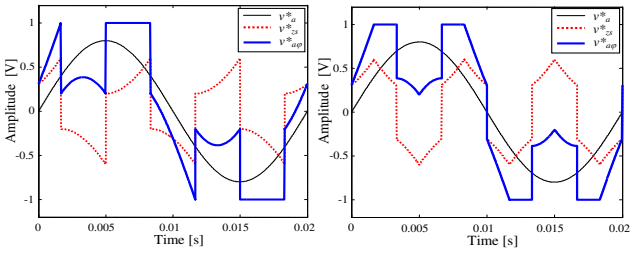
In order to generate the various GDPWM algorithms, we consider the initial vector v_{abc}^{*} of reference signals for the classical PWM technique using sinusoidal modulation. Then we use the new vector of references, $v_{ab\alpha\varphi}^{*}$ phase is shifted of the initial vector v_{abc}^{*}, by φ $(0 \le \varphi < \pi/3)$. For each value of the delay $(0 \le \varphi < \pi/3)$, the maximum and minimum values of $v_{ab\alpha\varphi}^{*}$ can be obtained as: $V_{\max,\varphi} = \max\left(v_{ab\alpha\varphi}^{*}\right)$ and $V_{\min,\varphi} = \min\left(v_{ab\alpha\varphi}^{*}\right)$. Finally, the zero sequence v_{zs}^{*} is built using expression *(9)*, in which the parameter k_0 takes alternately the values *0* or *1* as follows :

$$\begin{cases} if \;\; V_{\max,\varphi} + V_{\min,\varphi} < 0 \quad then: k_0 = 1 \\ if \;\; V_{\max,\varphi} + V_{\min,\varphi} \ge 0 \quad then: k_0 = 0 \end{cases}$$

Various DPWM modulating waveforms can be generated by using $(0 \le \varphi < \pi/3)$. The particular cases of φ equals to *0*, $\pi/6$, $\pi/3$ and $\pi/2$ are called DPWM3, DPWM0, DPWM1 and DPWM2 respectively [5, 6]. For these particular schemes, the initial sinusoidal reference phase voltage, the zero sequence signal and the final reference phase voltage (modulating waveform) are shown in (Fig. 1.) considering the modulation ratio r = 0.8.



a. DPWM0 reference b. DPWM1 reference



a. DPWM2 reference b. DPWM3 reference

Fig. 1. Modulation principale of HSVM

IV. Simulation Results

The structure of the three-level neutral point clamped (NPC) inverter is presented in (Fig.2). Each of the three legs is connected to two series diode and incorporates four main power switch devices with anti-parallel diode in series as shown in (Fig.2).

Each inverter leg requires four control signals [7]:

$$q_{i3} = 1 - q_{i1} \;\text{ and }\; q_{i4} = 1 - q_{i2}, \; i = 1, 2, 3 \quad (10)$$

For the leg 1, q_{11} is obtained by comparing the reference signal to the upper carrier and q_{12} is obtained by comparing the same reference to the lower carrier, while q_{13} and q_{14} are

obtained using (10). All switches operate symmetrically, hence our study is restricted to q_{11} only [7-9].

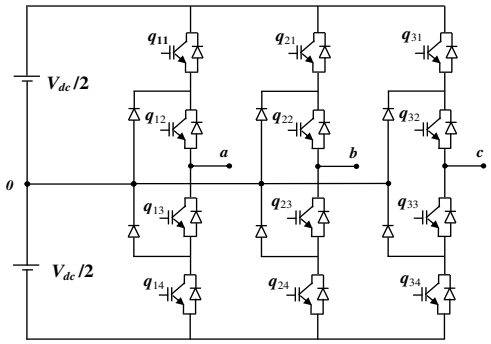
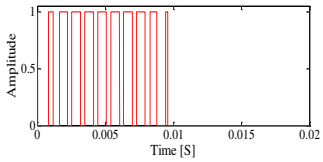
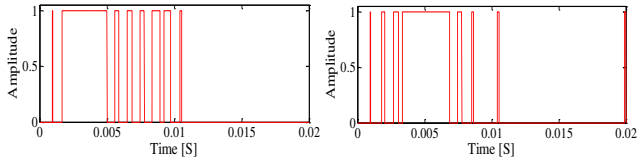


Fig. 2. Modulation principale of HSVM

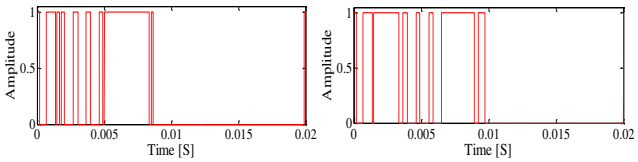
Simulations are conducted on the three level inverter and the results are given in (Fig.3, Fig.4 and Table I).



a. HSVM

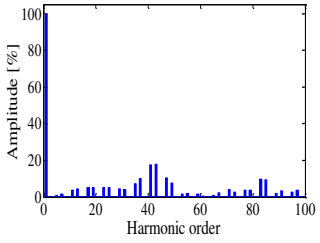


b. DPWM0 c. DPWM1

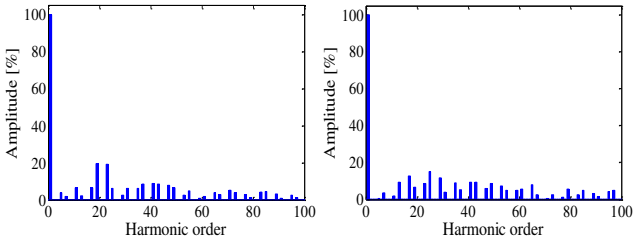


b. DPWM2 c. DPWM3

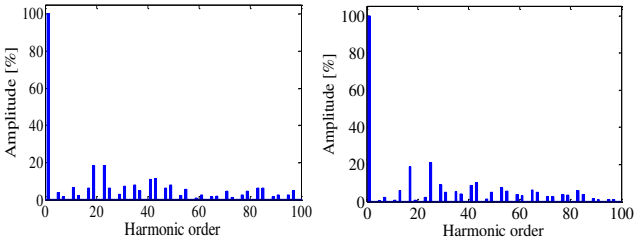
Fig. 3. Control signals for various DPWM schemes (r = 0.8 and M = 21)



a. HSVM



b. DPWM0 c. DPWM1



b. DPWM2 c. DPWM3

Fig. 4. Harmonic spectra of output voltage (-r = 0.8 and M = 21)

Table I. shows obtained results using Matlab/Simulink for both continuous and discontinuous schemes applied to control the three-phase NPC inverter.

TABLE I. COMPARISON OF DPWM SCHEMES FOR THREE-LEVEL INVERTER

PWM Technique	Number of Commutations per fundamental period	THD (%)	THDW (%)
HSVM	21	42.64	1.05
DPWM0	16	42.07	1.70
DPWM1	16	42.47	1.49
DPWM2	18	43.94	1.65
DPWM3	16	42.93	1.53

Notes:

1. The THD is the total harmonic distortion defined as follows:

$$THD(V) = \frac{\sqrt{\sum_{n=2}^{\infty} V_{neff}^2}}{V_{1eff}}$$

2. The THDW is the weighted total harmonic distortion defined as follows:

$$THDW(V) = \frac{\sqrt{\sum_{n=2}^{\infty} \left(\frac{V_{neff}}{n}\right)^2}}{V_{1eff}}$$

It important indices to measure quality of output voltage fed to grid, gives a better measure of harmonic pollution by using the order of each harmonic component as its weight factor.

According to (Table I), we can mention that the results of the discontinuous PWM techniques are close to those of the continuous one. In the other hand, they present the advantage of reducing the number of commutation of the transistors thus reducing the switching losses.

The results show that:

1) The important reduction of the number of the commutations using DPWM techniques is confirmed (Fig. 3 and Table 1).

2) The spectra of DPWM1 (Fig.4) contain less low order harmonics and this is confirmed by the THDW (Table 1).

V. CONCLUSION

The paper evaluates the performance of a three-level inverter controlled by advanced DPWM schemes (DPWM0, DPWM1, DPWM2 and DPWM3) and compared them to the conventional hybrid space vector modulation (HSVM) approach. The primary objective was to minimize switching losses in the inverter by reducing the numbers of commutations. Simulation results demonstrate that the proposed DPWM techniques significantly reduce switching losses while maintaining an acceptable total harmonic distortion (THD) in the output voltage. These findings highlight the potential of DPWM strategies to enhance inverter efficiency. Based on the obtained results, future research should focus on experimentally validating these methods and investigating their practical applications to further assess their effectiveness and reliability.

REFERENCES

[1] Vladimir Blasko, "Analysis of a Hybrid PWM Based on Modified Space Vector and Triangle Comparison Methods", IEEE trans. Ind. Appl. Vol. 33, N0. 3, pp. 756-764, May/June 1997.

[2] N. Boudjerda, M. Mellit, B. Nekhoul, K. El Khamlichi Drissi and K. Kerroum, "Reduction of Conducted Perturbations in DC-AC Converters by a Dual Randomization of Hybrid Space Vector Modulation", International Review of Electrical Engineering (IREE), Vol. 1, N0. 1, pp.154-161, March-April 2006.

[3] Touil, S. A., Boudjerda, N., Boubakir, A., & El Khamlichi Drissi, K. (2019). A sliding mode control and artificial neural network based MPPT for a direct grid‐connected photovoltaic source. Asian Journal of Control, 21(4), 1892-1905.

[4] Touil, S. A., Boudjerda, N., Boubakir, A., & Drissi, K. E. K. (2019). Closed loop discontinuous pulse width modulation control used in inverter grid-connected photovoltaic system for reduced switching losses. Revue roumaine des sciences techniques-serie electrotechnique et energetique, 64(4), 357-363.

[5] A. M. Hava, R. J. Kerkman, T. A. Lipo, "Simple Analytical Tools for Carrier Based PWM Methods", IEEE Power Electronics Specialists Conference. Vol. 2, pp. 1462-1471, June 22-27, 1997.

[6] M. Sushama and A. Jaya Laxmi, "Generalized PWM Algorithm for VSI Fed Induction Motor Drives Based on Unique Zero sequence Signal over the Entire Modulation Region", IEEE proc. India Conference (INDICON), 2011, pp. 1-7.

[7] Fei Wang, "Sine-Triangle versus Space-Vector Modulation for Three-Level PWM Voltage-Source Inverters", IEEE trans. Ind. Appl. Vol. 38, N0. 2, pp. 500-506, March/April 2002.

[8] Touil, S. A., Boudjerda, N., Boubakir, A., & Drissi, K. E. K. Sliding Mode Control and Discontinuous PWM for Minimum Switching Losses in Inverter Grid-Connected PV System, ISTSID (2019).

[9] Touil, S. A., Boudjerda, N., Boubakir, A., & Boudouda, A. (2017, October). Sliding mode control of a grid-connected photovoltaic source via a three-phase inverter using incremental conductance MPPT. In 2017 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B) (pp. 1-6). IEEE.

Artificial Intelligence Advances for Welding Inspection: A Comprehensive Review of X-Ray Digital Radiography Methods

Seghier chahinez & Ayad Amar
Department industrial Engineering
National Higher School of Technology and Engineering
Annaba, Algeria
{c.seghier & a.ayad}@ensti-annaba.dz

Seridi Hassina
Department of computer science
University of Badji Mokhtar
Annaba , Algeria
h.seridi@ensti-annaba.dz

***Abstract*— Welding joints are critical and crucial for structural integrity and safety of industrial components, a reliable inspection is equally vital to ensure quality. While digital radiography (DR) is a common method for defect detection with high image quality, it relies on human expertise, which is sensitive to errors and time consuming. Recent advancements in artificial intelligence (AI) have introduced powerful tools, to enhance defect detection and classification in automated workflows. This review explores the integration of AI techniques in welding inspection via X-ray DR, providing a comprehensive analysis of their capabilities, limitations, and potential to transform the field.**

Keywords— welding process, Digital Radiography, Image processing, Artificial intelligence.

I. INTRODUCTION

Welding joints are critical to the integrity and safety of structures across industries. Traditionally, X-ray radiography for defect detection in welding joints has relied on human expertise. However, recent advancements in artificial intelligence (AI) have introduced opportunities to increase the accuracy and efficiency of defect detection and classification.

Though seemingly straightforward, welding is a complex process that demands precise control. It involves joining two metal pieces by applying heat until they melt and fuse together, forming a permanent bond. This process requires careful adjustment of parameters to regulate the energy source, ensuring the metals liquefy without compromising structural integrity.[1]

The advent of Industry 4.0 has led to increased automation and digitalization in manufacturing and quality control, transforming processes like digital radiography (DR) in welding inspection. Digital radiography captures X-ray images electronically, providing higher image quality, faster processing, and greater efficiency than traditional film-based methods. Although DR is a well-established technique, current practices still require human inspectors to interpret hundreds of weld scans daily—an error-prone and time-consuming task that limits the pace of fully automated workflows. Thus, a central question arises: To what extend can we rely on machines for radiographic image interpretation in welding inspections?[2]

AI methods, particularly neural networks, machine vision, expert systems, case-based reasoning, and fuzzy logic, offer promising solutions for interpreting NDT data. These techniques can enhance the speed and accuracy of defect detection in large-scale automated inspections. In this review, we examine recent AI advancements in the detection and classification of defects in welding joints, as captured by X-ray radiography.

This review is organized as follows:

- An overview of the welding process and defect detection using digital radiography
- A discussion of technologies involved in each phase of DR inspection
- An analysis of AI algorithms used in various inspection phases
- A comparative study of these algorithms

II. WELDING PROCESS AND DIGITAL RADIOGRAPHY

A. Welding process

Welding can be defined as the process that "apply sufficient heat to melt two pieces of metal that are fused together after they cool and harden. In general, a separate piece of metal acts as a filler material that is deposited and cooled between the two pieces in question " [3] there are many welding processes used in manufacturing depending on the source, we can summarize them in figure 01.[3]

While automated welding has surpassed manual welding in many respects, even within automated systems, a surprisingly diverse range of potential defects can be identified and classified.

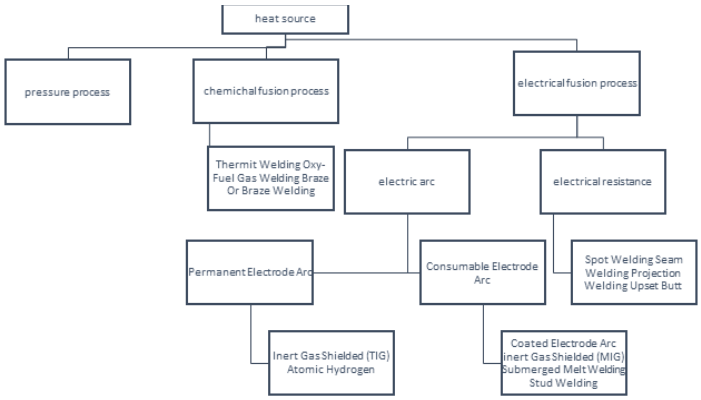


Fig.1 : Types of welding processes.

B. Welding defects :

Welding processes can lead to various defects. While some defects, such as those related to weld metal quality, are influenced by material properties, others arise from human error or insufficient welder skill. In automated pipe welding, common defects are often related to process parameters like welding current, voltage, and speed, which can directly impact weld quality. The most common defects occurring in welds are given in the following :

1. *Cracks:* Crack-type defects are cracks or fractures in the weld or the heat-affected zone. They can be microscopic or visible to the naked eye. Cracks can weaken the structure and propagate under service loads, reducing the strength and service life of the material. Cracks are critical and must be eliminated or rigorously controlled. [4]
2. *Porosity:* Pore-type defects refer to small cavities or gas bubbles trapped in the weld. They can be spherical or elongated. Porosity reduces the strength and tightness of the weld. It can affect the integrity of the joint. Although not as critical as cracks, porosity should be minimized. [4]
3. *Lack of fusion:* the non-penetration type defect occurs when the weld does not fully penetrate the base material. There may be a gap between the joined parts. The lack of penetration reduces the strength of the joint and can lead to premature failure.
4. *Non-metallic inclusions:* These could be caused by weld-metal contamination from contaminants on the joint's surface or the environment. These could be caused by weld-metal contamination from contaminants on the joint's surface or the environment. Poor joint preparation, improper welding technique, and inadequate shielding gas can also contribute to the formation of these defects
5. *undercut:* the exposed upper edges of the bevelled weld preparation have a tendency to melt and channel to the deposited metal in the weld groove, especially during the final or cover pass. resulting in a groove with more or less sharp edges along the weld reinforcement.

Left with a non-exhaustive list of potential defects that can be identified and classified. Non-destructive testing using digital radiography is widely used for internal welding defect detection without altering the safety of the controlled welding joint. Some example of defects are illustrated Fig 02 As a a well-established discipline, radiography inspection is subject to international standards and the computer aided radiography needs to align with the standards of inspection.

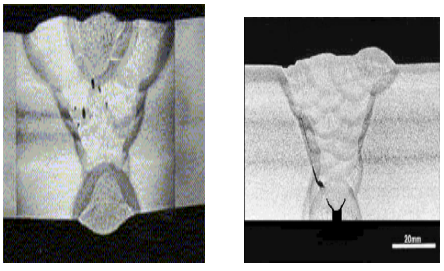


Fig 02 : welding defects a. porosity b. lack of fusion

III. DIGITAL RADIOGRAPHY

A. Definition : "Radiography is non-destructive testing method to find out the internal discontinuities present in a component or an assembly. It is based on differential absorption of penetrating radiation by the part being inspected*."*

B. Principle: Radiography captures the image of X-rays absorbed as they pass through an object . The output is a representation of the internal structure of the object. As the rays flow through the material, any discontinuity creates tiny holes that absorb additional radiation.

C. Parameteres influencing the radiographic image :

The parameters to consider for the exposure are:

- the voltage, measured in kilovolts (kV), it regulates the energy of the radiation and thus its penetration into the material,
- the intensity, in mill amperes (mA), and the time, whose product determines the total amount of radiation received by the object,
- the distance separating the radiation source from the receptor,
- the sensitivity of the receptor
- the nature of the object (atomic number of the constituent elements), and its (or their) thicknesses(s).

D. Radiography inspection standard :

Radiography inspection standards aim to establish a consistent and reliable approach to detecting and evaluating defects in welds using radiographic techniques. These standards, such as ISO 17636-2, ISO 10893-7, ASTM E 2737-10, and ASTM E 2597-7, focus on ensuring the quality and integrity of welds by defining specific requirements for:

1) Equipment: They address the performance and calibration of radiographic equipment, including both traditional film-based systems and modern digital detector arrays (DDAs). For instance, ISO 17636-2 specifies minimum requirements for monitor screens used for digital image evaluation, including brightness, contrast, and resolution.

2) Image Quality: Standards define parameters like spatial resolution, contrast sensitivity, signal-to-noise ratio (SNR), and image unsharpness to guarantee the detectability of different types of defects. ISO 10893-7 sets requirements for normalized SNR, while ISO 17636-2 introduces compensation principles that allow exceeding maximum voltage limits from film radiography by leveraging the higher SNR and contrast sensitivity of DDAs

3) Inspection Procedures: They outline procedures for conducting radiographic inspections, including the placement of image quality indicators (IQIs), exposure settings, and image interpretation criteria.

4) Acceptance Criteria: Standards often establish acceptable limits for the size and type of defects based on the intended application and the potential impact on the component's performance. This ensures that welds meet the required safety and reliability standards.

Those specification needs to be taken into account while integrating AI algorithms in the inspection process.

IV. State of art of AI for digital radiography defect detection and classification

Digital radiography inspection relies on many tasks related to image manipulation from image enhancement to defect detection and classification those tasks can involve computer aided AI methods to help inspectors optimize multiple operation for a more reliable and less time consuming inspection. We summarize the existing methods related to each task in the table 01.

Table 01: literature review on AI methods for welding defect detection

Task	Reference	Method	Principle	Performance
Image manipulation	Li et al., 2024[5]	Synthetic Data augmentation (Single Image Single Defect, SISD) and (Single Image Multi Defects, SIMD), DNN	realize a synthesis based on a small number of real samples by data augmentation	an effective solution for data sample generation in the industrial field. Improving accuracy of defect classification of YOLOV5.6.1X dataset from 66.6% and 72.8%
	He et al,2023[6]	DG-GAN: A High Quality Defect Image Generation Method for Defect Detection	The surface defect detection, a defect image generation method DG-GAN is Based on the idea of the progressive generative adversarial, D2 adversarial loss function, cyclic consistency loss function	The YOLOX detection accuracy improved with the increase in generated defect images (the highest increases were 6.1% and 20.4%, respectively).
Quality enhancement	Wafaa Al-Hameed, 2019[7]	Fuzzy logic, Fuzzy Hyperbolic Threshold	The fuzzy logic system solves many problems by relying on mathematical concepts which are usually easy and uncomplicated.	satisfactory result in contrast enhancement for the input images to give improved quality results.
Image labeling	Dai et al,2020 [8]	Semi-supervised learning :	utilizes a combination of active learning and self-training through a sample query suggestion algorithm for classification.	improved classification accuracy while reducing the need for manual annotations
Defect detection	Madhvacharyula et. Al in 2022 [9]	situ weld defect detection	Real-time or in situ weld defect detection aids in the detection of defects during the welding process	
	Yan Hanbing et. Al,2009.[10]	Binary decision tree	With the help of svm and binary decision tree to detect defaults	Accurate defect detection
	. J Shao et al,2011[11]	Support Vector Machine (SVM)	Feature extraction, binary classification, dimensionality reduction	Accuracy rate: 99.4%
Defect classification	Ajmi et Al in 2020 [12]	Deep learning : data augmentation, ANN	Using data augmentation to leverage the performance of ANN in small data sets	Overcoming the challenge of DL ANN application on small dataset with an overall performance of 95 % in precision
	Kumaresan et al., 2021 [13]	Transfer learning , CNN,	Using a pre-trained datasets with machine learning algorithms like VSM, random forest for feature extraction	Increase the accuracy of CNN for small datasets by using pre-trained datasets with up to 99.1% in binary classification and 97.8% in multiclass classification
	Vishal et al., 2019 [14]	Fuzzy systems: Fuzzy K-NN, Adaptive network based on fuzzy system	Twelve numeric features were extracted to represent each instance of defect	accuracy of 91.57%

V. A COMPARATIVE ANALYSIS OF AI METHOD FOR WELDING DEFECT DETECTION AND CLASSIFICATION:

The previous section summarized the different AI methods employed while manipulating X-ray images for welding defect classification we can divide those algorithms in two classes traditional methods of machine learning and deep learning methods and compare the advantages and limitations of each approach and the further advancement and future directions to overcome the challenges. We can summarize the analysis in the table 02

Table 02 : comparative analysis of AI methods

Approach	Method	Advantages	Limitations
Machine learning	SVM	Great classifier especially in high dimensional spaces; memory efficient, clear margin separation between classes	Not suitable for large datasets, sensible to noise, manual feature selection, depending on expert knowledge to perform classification
	K-NN FUZZY LOGIC	Fuzzy logic handles uncertainty and imprecision in inspection data, model human reasoning processes.	Difficult to implement, Challenging in setting up the fuzzy logic, complex model
	Random forest/ decision tree	Suitable for complex models, improved accuracy comparing to decision tree, scalability handles both large and small data sets, parallel processing	Difficult to interpret, computationally expensive, sensible to noise, manual feature extraction
Deep learning	GANS	High quality images generation, relatively easy to train, potential to learn from little data, and no labelled data, helpful when dealing with small samples of data	Unstable training, slow due to the controversial models of generators and discriminators, risks of unrealistic images, Require large data samples for more accurate training results
	CNN	Powerful in pattern recognition, image classification, automatic feature extraction, handles well large datasets	Interpretability challenge, high computational costs, long time for training, challenging model integration
	Semi–supervised learning	Computational costs reduction, improved prediction by combining both labelled and unlabelled data, flexible Handles large datasets will minimum labelled data	Hard to evaluate in case of little labelled data sample, complexity and requirements for pre-processing techniques, potential noise creation
	Transfer learning and data augmentation	Helps in limited datasets, prevent overfitting, improve model performance by using pre-trained data sets, save computational costs	Risk of suboptimal performance or even misdirected learning, domain specific feature limit potential for learning, high computational costs, time consuming task

VI. CONCLUSION

The integration of artificial intelligence (AI) into welding automation offers transformative potential, however a successful integration addresses many challenges related to effective costs models with high accuracy and more accessible interpretability and for digital radiography inspection integration of AI methods that are aligned with international standards. There is also a high tendency for deep learning approach adoption and multidisciplinary collaborations to make advantage of cutting edge techniques in industries like NDT4.0 that are redefining the future of NDT automation.

REFERENCES

[1] *Fault Diagnosis and Prognosis Techniques for Complex Engineering Systems*. Elsevier, 2021. doi: 10.1016/C2019-0-04569-4.

[2] J. Vrana, N. Meyendorf, N. Ida, and R. Singh, 'Introduction to NDE 4.0', in *Handbook of Nondestructive Evaluation 4.0*, N. Meyendorf, N. Ida, R. Singh, and J. Vrana, Eds., Cham: Springer International Publishing, 2022, pp. 3–30. doi: 10.1007/978-3-030-73206-6_43.

[3] 'ASM Metals Handbook, Vol 17.pdf'.

[4] '31005449.pdf'.

[5] L. Li *et al.*, 'Synthetic data augmentation for high-resolution X-ray welding defect detection and classification based on a small number of real samples', *Eng. Appl. Artif. Intell.*, vol. 133, p. 108379, Jul. 2024, doi: 10.1016/j.engappai.2024.108379.

[6] X. He, Z. Luo, Q. Li, H. Chen, and F. Li, 'DG-GAN: A High Quality Defect Image Generation Method for Defect Detection', *Sensors*, vol. 23, no. 13, p. 5922, Jun. 2023, doi: 10.3390/s23135922.

[7]'FuzzylogicForDefectDetectionofRadiographyImages.pdf'.

[8] 'Dai, W.; Mujeeb, A.; Erdt, M.; Sourin, A. Soldering defect detection in automatic optical inspection. Adv. Eng. Inform. 2020, 43, 101004.'

[9] A. S. Madhvacharyula, A. V. S. Pavan, S. Gorthi, S. Chitral, N. Venkaiah, and D. V. Kiran, 'In situ detection of welding defects: a review', *Weld. World*, vol. 66, no. 4, pp. 611–628, Apr. 2022, doi: 10.1007/s40194-021-01229-6.

[10] 'Yan Hanbing, Zhao Lina, Ju Hui, "Research on SVM Based Classification for Welding Defects in Radiographic Testing" 978-1-4244-4131-0/09/2009 IEEE'.

[11] 'J. Shao, H. Shi, D. Du, L. Wang and H. Cao, "Automatic weld defect detection in real-time X-ray images based on support vector machine," 2011 4th International Congress on Image and Signal Processing, Shanghai, China, 2011, pp. 1842-1846, doi: 10.1109/CISP.2011.6100637. keywords: {Welding;Support vector machines;X-ray imaging;Feature extraction;Real time systems;Image segmentation;Training;Real-time X-ray image;Defect detection;Background subtraction;Support Vector Machine},'.

[12] C. Ajmi, J. Zapata, J. J. Martínez-Álvarez, G. Doménech, and R. Ruiz, 'Using Deep Learning for Defect Classification on a Small Weld X-ray Image Dataset', *J. Nondestruct. Eval.*, vol. 39, no. 3, p. 68, Sep. 2020, doi: 10.1007/s10921-020-00719-9.

[13] S. Kumaresan, K. S. J. Aultrin, S. S. Kumar, and M. D. Anand, 'Transfer Learning With CNN for Classification of Weld Defect', *IEEE Access*, vol. 9, pp. 95097–95108, 2021, doi: 10.1109/ACCESS.2021.3093487.

[14] V. Vishal, R. Ramya, P. Vinay Srinivas, and R. Vimal Samsingh, 'A review of implementation of Artificial Intelligence systems for weld defect classification', *Mater. Today Proc.*, vol. 16, pp. 579–583, 2019, doi: 10.1016/j.matpr.2019.05.131.

Enhancing Energy Efficiency in PV-Powered Ventilation Systems: A Cuckoo Search Algorithm Approach

Ali Mazari
Laboratory of The Applied And Automation And Industrial Diagnostic
University Of Djelfa
Djelfa, Algeria
ali.mazari@univ-djelfa.dz

Okba Fergani
Laboratory of Identification, Commande, Control and Communication (LI3CUB)
University Mohamed Khider Biskra
Biskra, Algeria
okba.fergani@univ-biskra.dz

Hamou Ait Abbas
Laboratory of Electrical and Automatic Systems Engineering (LGSEA)
University of Bouira
Bouira, Algeria
h.aitabbas@univ-bouira.dz

Kouider Laroussi
Laboratory of The Applied and Automation and Industrial Diagnostic
University of Djelfa
Djelfa, Algeria
k.laroussi@univ-djelfa.dz

Billel Naceri
Laboratory of Identification, Commande, Control and Communication (LI3CUB)
University of Biskra
Biskra, Algeria
billel.naceri@univ-biskra.dz

Younes Belhadjer
Laboratory of The Applied and Automation and Industrial Diagnostic
University of Djelfa
Djelfa, Algeria
younes.belhadjer@univ-djelfa.dz

***Abstract*— This study presents the application of a metaheuristic algorithm approach to optimizing the performance of ventilation system powered by photovoltaic (PV) energy, specifically under partial shading conditions to ensure maximum energy extraction from the PV system, with the optimization process controlled by the Cuckoo Search algorithm (CSA) algorithm. The CSA algorithm successfully mitigates the challenges posed by partial shading, achieving 75% of the maximum theoretical power output from the PV system. The extracted power is then utilized to operate a ventilation system with a DC motor, which demonstrates excellent operational characteristics, including a torque of 7.5 Nm and a speed of 350 rad/s. These results are crucial for industries, particularly agriculture, where energy-efficient ventilation is necessary for maintaining optimal environmental conditions. This research highlights the potential of integrating the CSA algorithm with PV systems to provide a sustainable and energy-efficient solution for agricultural and industrial applications, reducing dependence on grid power and promoting renewable energy use.**

Keywords—MPPT, PV system, Cuckoo search algorithm, Ventilation system

I. INTRODUCTION

The rising need for eco-friendly and energy-saving options across industries has sparked greater focus on combining renewable energy systems with building designs [1]. A promising approach involves utilizing solar panels to operate ventilation systems, essential for ensuring healthy indoor air and optimal temperature regulation. Traditional ventilation systems, while essential, often consume a significant amount of energy, contributing to higher operational costs and environmental impact [2]. The integration of PV systems with ventilation units offers a promising approach to reduce these costs and enhance sustainability by utilizing solar energy to power ventilation systems [2].

However, the efficiency of such systems is subject to fluctuations in solar energy generation due to environmental variables such as time of day, weather conditions, and seasonal changes [3]. To address this, power conversion devices like DC-DC boost converters are employed to regulate and step up the voltage output from the PV system, ensuring that sufficient power is supplied to the ventilation units. While these converters offer improved performance, optimizing their operation in real-time remains a challenge [4].

Recent advancements in optimization algorithms have shown potential in addressing such challenges. An example of this is the Cuckoo Search algorithm (CSA), a method modeled after the unique breeding patterns of cuckoo birds and has been successfully applied to complex optimization problems in various fields. This algorithm can effectively adapt to dynamic conditions, optimizing system parameters to maximize energy output and ensure the efficient operation of the PV-powered ventilation system [5].

This paper explores the integration of a PV system with a ventilation unit via a DC-DC boost converter, with the operation optimized using the Cuckoo Search algorithm. The aim is to maximize the power produced by the PV system, ensuring optimal energy utilization for ventilation while reducing reliance on grid power. Through simulations and performance analysis, the paper demonstrates the potential of this hybrid system in enhancing energy efficiency and sustainability in building ventilation applications.

The paper is organized as follows: Section 2 outlines the system design, covering the PV setup and boost converter. Section 3 explores the control mechanisms, while Section 4 presents the findings. Section 5 concludes, summarizing the work and suggesting areas for future research.

II. PV SYSTEM

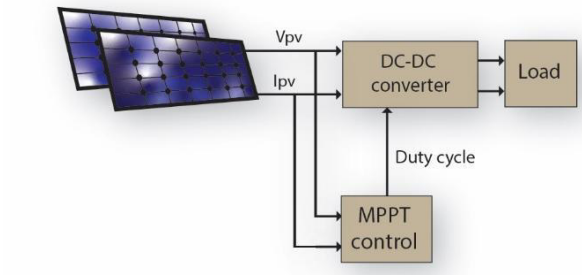


Fig. 1. *PV system overview.*

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

The photovoltaic (PV) system used in this study consists of a series of solar panels that convert sunlight into electrical energy. The electrical energy produced by the PV system is direct current (DC) and is used to power the ventilation system via a DC-DC boost converter. The system's performance is highly dependent on factors such as solar irradiance, panel orientation, and temperature [6].

A. PV Panel Model

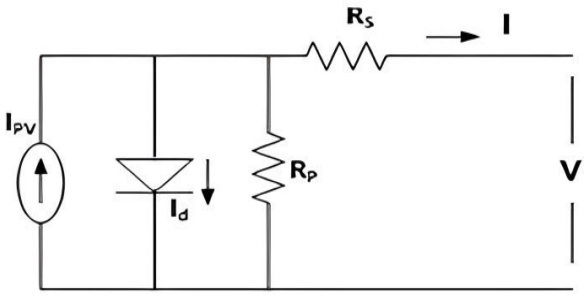


Fig. 2. *PV equivalent circuit [4].*

The output power from a PV panel can be modeled based on the photovoltaic equation [4]:

$$I = I = I_{pv} - I_S\left(e^{\frac{q(V+R_SI)}{N_SKta}} - 1\right) - \frac{V+R_SI}{R_p} \quad (1)$$

I_{pv} : Light-generated current
I_s : Reverse saturation current
R_p : Shunt resistance
R_s : Series resistance
I : PV cell generated current
V : Voltage across the PV cell

The current generated by sunlight in a PV cell, known as the light-generated current, depends on factors like sunlight intensity and the cell's competence in converting light into electricity. The reverse saturation current, which occurs due to diode leakage under reverse bias, plays a key role in shaping the cell's performance under different conditions. Series resistance, caused by electrical connections and the cell's internal resistance, results in voltage losses within the module. The diode ideality factor captures non-ideal diode behavior, including recombination losses, while the thermal voltage, influenced by temperature, affects the diode's characteristics [4].

Shunt resistance provides alternative pathways for current to flow around the diode, impacting the PV cell's efficiency under different light conditions. Table 1 outlines the specifications of the PV module employed in this research, the 1Soltech 1STH-215-P, which includes 20 parallel strings and 2 modules connected in series per string. These parameters are essential for accurately modeling the PV system's performance and understanding how each factor contributes to the overall efficiency and output of the module under different environmental conditions [4].

TABLE I. PV PARAMETERS [4].

Module data	Value
Max Power (W)	213.5
Open-Circuit Voltage Voc (V)	36.3
Short-Circuit Current Isc (A)	9.84
Max Voltage Vmax(V)	29
Max Current Imax(A)	7.35

B. PV Panel Model

Partial shading occurs when certain sections of a solar panel array are obstructed from sunlight, typically due to objects like trees, buildings, or other obstructions. This shading can significantly affect the performance of the entire PV system, even if only a small portion of the array is impacted. When a section of the array is shaded, the power output of the affected solar cells decreases, leading to a drop in the overall system efficiency. The impact is exacerbated by the series connection of PV modules, as shading on one module can cause current mismatch across the array. This mismatch results in voltage drops, reduced current flow, and potential hotspots that could damage the system. To mitigate the effects of partial shading, techniques like module bypass diodes, microinverters, or power optimizers are commonly employed to isolate shaded modules and maintain higher overall energy production. Understanding and handling partial shading is pivotal for ensuring better performance and longevity of PV systems, especially in environments with varying levels of sunlight throughout the day [7].

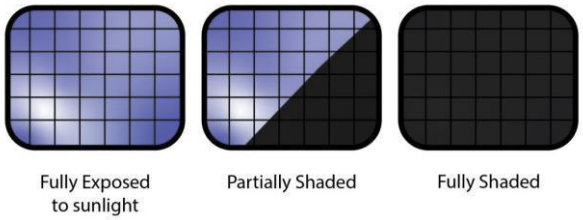


Fig. 3. *Partial shading effect..*

C. DC-DC boost converter

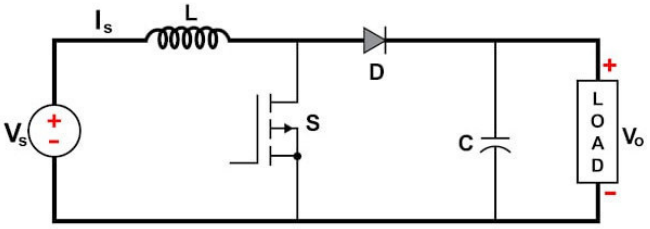


Fig. 4. *Equvalent scheme of a DC-DC boost converter*

A boost converter is an electronic device designed to elevate a lower DC input voltage to a higher DC output level. It is frequently utilized in photovoltaic (PV) systems to raise the voltage from solar panels to meet the needs of connected devices, such as ventilation systems. The converter consists of key components like an inductor, a transistor switch, a diode, and a capacitor. When the switch is activated, the inductor stores energy as a magnetic field. Upon deactivation, the stored energy is discharged, resulting in a higher voltage across the output capacitor [6]. The output voltage is regulated by the switch's duty cycle, where increasing the duty cycle raises the voltage. A key benefit of the DC-DC boost converter is its high efficiency, as it transfers energy directly without significant losses, such as heat generation. However, the converter's performance can be affected by factors such as switching frequency, load conditions, and component characteristics. In PV systems, The boost converter effectively adjusts the energy collected from solar panels to meet the requirements of the connected load, optimizing overall system performance [8].

III. PV SYSTEM CONTROL

A. *MPPT Control by Cuckoo search algorithm*

The CSA is a bio-inspired optimization strategy modeled after the brood parasitism of cuckoo birds, specifically their egg-laying and nest-competition strategy. The algorithm is designed to solve complex optimization problems by exploring and exploiting the solution space. In the context of your system, the CSA is applied to optimize the performance of the PV system and the DC-DC boost converter to maximize power production while ensuring efficient energy utilization [5].

Cuckoo birds lay their eggs in the nests of other species, and the eggs' survival depends on the host bird's ability to detect and reject foreign eggs. Similarly, in the algorithm, potential solutions (representing different configurations of the PV system) are represented by eggs, and the nests are candidate solutions to the optimization problem. The algorithm works by iteratively improving the candidate solutions based on two main operations: levy flights (which are random walks used for exploration) and nest abandonment (which is a mechanism for exploitation). By balancing exploration and exploitation, the Cuckoo Search algorithm efficiently converges to an optimal solution [9].

The Lévy flight equation, discovering the generation of new solutions [10, 11]:

$$x_i^{t+1} = x_i^t + \alpha \otimes l\acute{e}vy(\lambda) \quad (2)$$

Pseudo Code for Cuckoo Search Algorithm

```
Initialize population of nests (solutions)
Define fitness function for PV system performance (e.g., power output)
Define maximum number of iterations or stopping criteria
For each iteration:
    For each nest (solution):
        Generate a new solution using Levy flight
        Assess the quality of the new solution (maximize power output)
        If the new solution outperforms the existing one:
            Update the current solution with the new one
    Evaluate the overall best solution (best power output)
    If stopping criteria is met:
        Terminate
Return the best solution (optimal configuration for PV system)
```

B. *Ventilation system*

A ventilation system plays a critical role in ensuring adequate air circulation within buildings, improving interior air quality, and maintaining a pleasant atmosphere for those inside. In this system, air is exchanged between the indoor and outdoor environment to regulate temperature, humidity, and air quality, and to eliminate indoor contaminants like carbon dioxide and volatile organic compounds (VOCs), and moisture. Ventilation systems can be largely assorted into natural ventilation, mechanical ventilation, and hybrid systems, where mechanical ventilation employ fans or blowers to circulate air, while natural ventilation depends on passive airflow driven by temperature gradients, wind, or pressure differences [12].

The ventilation system is powered by photovoltaic (PV) panels, which harness solar energy to operate the fans or ventilation units. This approach offers several advantages in terms of energy efficiency, sustainability, and cost reduction. Solar-powered ventilation systems are particularly beneficial for reducing energy consumption in buildings, as they rely on a renewable energy source (solar power) rather than conventional grid electricity [13].

Ventilation System with DC Motor-Powered Fans

The ventilation system is powered by DC motors, which drive the fans responsible for circulating air within the building. These motors are energized by the photovoltaic (PV) system, providing a sustainable and energy-efficient method to operate the fans. DC motors are widely used in such systems due to their simplicity, control flexibility, and efficient operation at variable speeds. By adjusting the motor speed, the airflow and ventilation rate can be controlled according to the building's needs, ensuring optimal indoor air quality and thermal comfort [14].

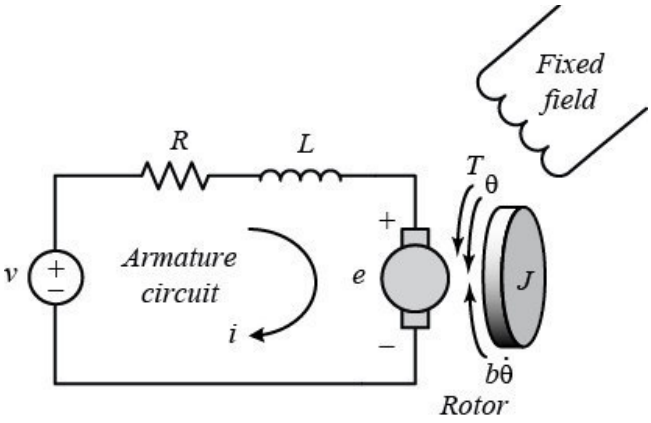


Fig. 5. *Electrical Circuit Diagram of DC motor.*

Optimizing System Performance with the CSA:

The CSA is employed to optimize the parameters of the ventilation system, including the duty cycle of the boost converter and the speed of the DC motor. The algorithm iteratively searches for the optimal settings to maximize the system's energy efficiency, adjusting for varying solar irradiance and environmental conditions. By ensuring that the motor operates at peak efficiency, the algorithm reduces energy waste and enhances the performance of the system. It balances the need for increased airflow with the available power from the PV system, ensuring that the fans provide the required ventilation while minimizing power consumption.

IV. SIMULATION AND RESULT DISCUSSION

To evaluate the effectiveness of the proposed CSA in optimizing power generation for a solar-powered ventilation system, comprehensive simulations were carried out in MATLAB Simulink. The setup included a PV system linked to DC motor-operated fans via a DC-DC boost converter. This setup mimics real-world conditions, where the system's energy comes from solar panels to operate the fans that provide ventilation. The primary aim was to evaluate how effectively the CSA could optimize system performance, particularly by controlling the boost converter and fan speed.

The System diagram (Fig. 5) illustrates the integrated system, including the PV panels, boost converter, and fan-driven ventilation system, all controlled and optimized using the CSA to promote the overall efficiency of the system.

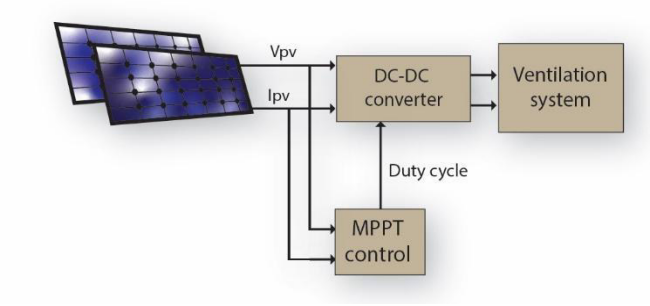


Fig. 6. *Ventilation system connected to PV system.*

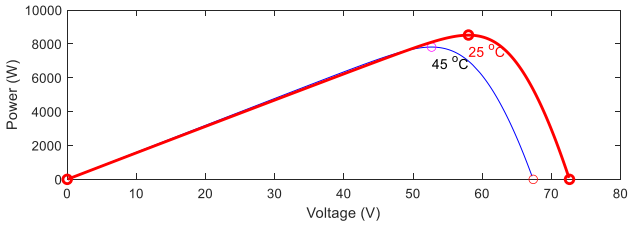


Fig. 7. *Nominal power of PV panel*

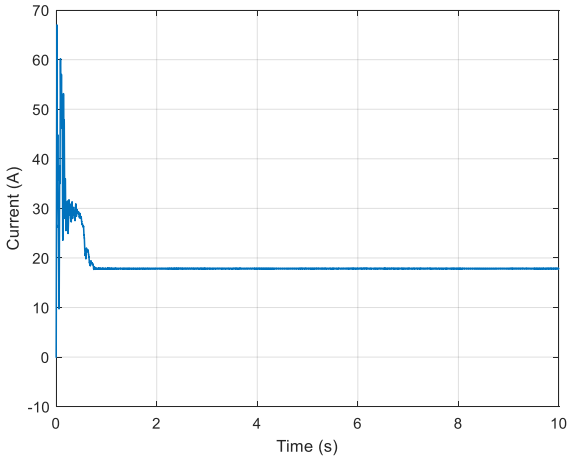


Fig. 8. *Current of the boost converter*

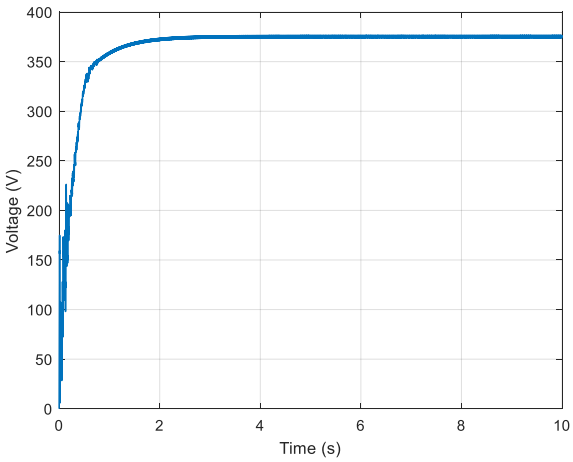


Fig. 9. *Voltage of the boost converter.*

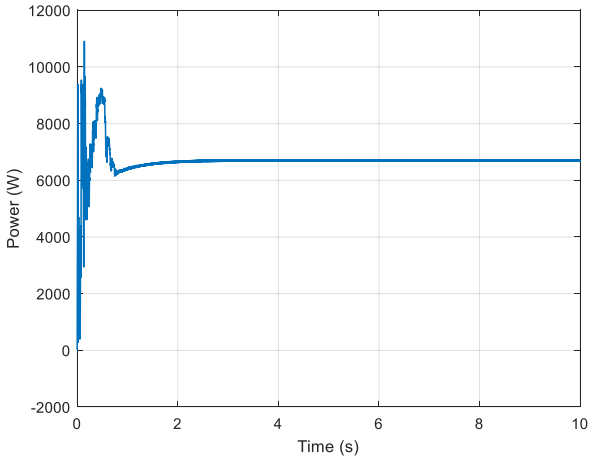


Fig. 10. *The power output from the boost converter.*

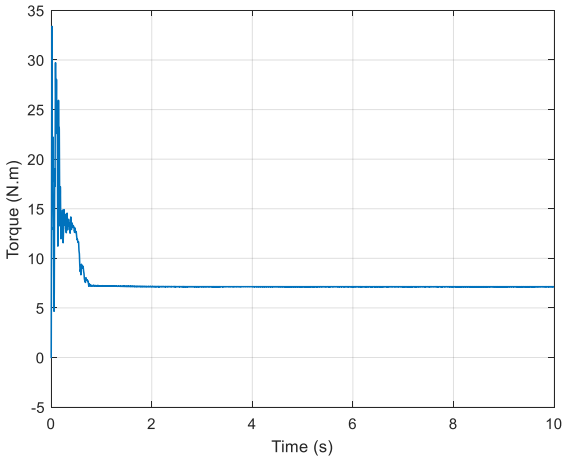


Fig. 11. *Torque of the fan's DC motor*

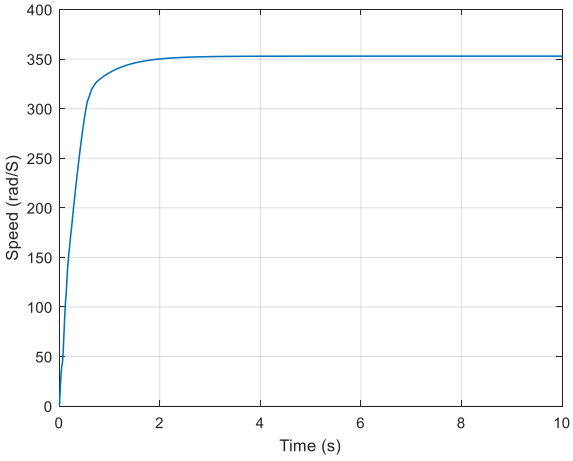


Fig. 12. *Speed of the fan's DC motor.*

The results demonstrate that the CSA optimization algorithm effectively addresses the challenges posed by partial shading conditions in photovoltaic (PV) systems. In this study, the CSA algorithm achieved an impressive 75% of the maximum theoretical power extraction, which is a significant improvement under non-ideal conditions. This enhanced power extraction was then utilized to operate a ventilation

system, demonstrating the algorithm's practical applicability in optimizing energy usage for industrial systems.

The ventilation system powered by the optimized PV output showed excellent performance, achieving a torque of 7.5 Nm and a speed of 350 rad/s, both of which are crucial parameters in industrial applications, especially in agriculture. These values are essential for ensuring the efficient operation of ventilation systems that are vital for maintaining proper airflow and temperature regulation in agricultural environments, such as greenhouses or livestock facilities.

Moreover, by integrating the Cuckoo Search optimization algorithm with a PV-powered system, a sustainable and energy-efficient solution is provided. This system not only maximizes the energy output from the PV panels but also reduces dependency on grid power, contributing to a more sustainable energy model. This approach is particularly advantageous in rural or off-grid areas, where reliable access to electricity may be limited.

V. Conclusion

In summary, this research effectively showcases the use of the Cuckoo Search algorithm to improve the efficiency of a solar-powered ventilation system, particularly in scenarios involving partial shading. The CSA algorithm significantly increases the energy extracted from the PV panels, achieving 75% of the maximum theoretical output. The ventilation system, powered by this optimized energy, achieves a torque of 7.5 Nm and a speed of 350 rad/s, making it highly suitable for industrial and agricultural applications where precise control over environmental conditions is crucial.

The integration of renewable energy sources with advanced optimization techniques, such as Cuckoo Search, offers a sustainable and cost-effective solution to energy management. By optimizing the performance of PV systems and ensuring reliable, efficient operation of ventilation systems, this research provides a viable pathway toward reducing operational costs, lowering environmental impact, and enhancing energy security for industries reliant on ventilation systems. Future work could focus on further refining the algorithm and exploring its scalability for larger industrial systems or multi-fan ventilation setups.

References

[1] W. Rzeźnik, I. Rzeźnik, and P. Mielcarek, "Impact of Using Photovoltaic Panels in Piggery on Greenhouse Gases Emission," 2018, doi: 10.15544/rd.2017.089.

[2] E. Giama, "Review on ventilation systems for building applications in terms of energy efficiency and environmental impact assessment," Energies, vol. 15, no. 1, p. 98, 2021.

[3] Y. Ohnishi, Y. Takahashi, K. Fujiwara, K. Hidaka, and H. Morita, "Modeling of <i>I – V</I> Characteristics for Crystalline Silicon Photovoltaic Modules Based on a Simplified Equivalent Circuit and a Temperature Correction," Electrical Engineering in Japan, vol. 215, no. 4, 2022, doi: 10.1002/eej.23410.

[4] O. Fergani, R. Mechgoug, A. Afulay Bouzid, N. Tkouti, and A. Mazari, "A New Modified Bacterial Foraging MPPT Technique with Dynamic Mutation Rates for Photovoltaic Systems under Partial Shading Conditions," International Journal of Engineering, vol. 37, no. 8, pp. 1569-1579, 2024.

[5] T.-P. Dao, "Cuckoo Search Algorithm: Statistical-Based Optimization Approach and Engineering Applications," pp. 79-99, 2020, doi: 10.1007/978-981-15-5163-5_4.

[6] A. Amir, A. Amir, H. S. Che, A. Elkhateb, and N. Abd Rahim, "Comparative analysis of high voltage gain DC-DC converter topologies for photovoltaic systems," Renewable energy, vol. 136, pp. 1147-1163, 2019.

[7] A. Mohapatra, B. Nayak, P. Das, and K. B. Mohanty, "A review on MPPT techniques of PV system under partial shading condition," Renewable and Sustainable Energy Reviews, vol. 80, pp. 854-867, 2017.

[8] P. Ainapure, P. Kulkarni, S. U. Patil, and A. Tiwari, "Design, Simulation and Implementation of MPPT Controlled Buck-Boost Converter Extracting Power From Solar Panel for Microgrid Applications," International Journal of Electrical Engineering & Technology, vol. 10, no. 6, 2019, doi: 10.34218/ijeet.10.6.2019.001.

[9] F. K. Abo-Elyousr, A. M. Abdelshafy, and A. Y. Abdelaziz, "MPPT-based particle swarm and cuckoo search algorithms for PV systems," Modern Maximum Power Point Tracking techniques for photovoltaic energy systems, pp. 379-400, 2020.

[10] F. Okba, R. Mechgoug, and A. B. Afulay, "Revolutionizing PV Pumping Systems with PMSM Machine and Advanced MPPT Algorithm Integration: A Comparative Study of PSO, GWO, and CSA Techniques," NeuroQuantology, vol. 21, no. 6, p. 1852, 2023.

[11] C. Zheyuan, A. T. Hammid, A. N. Kareem, M. Jiang, M. N. Mohammed, and N. M. Kumar, "A Rigid Cuckoo Search Algorithm for Solving Short-Term Hydrothermal Scheduling Problem," Sustainability, vol. 13, no. 8, 2021, doi: 10.3390/su13084277.

[12] S. Revathi, N. Sivakumaran, and T. Radhakrishnan, "Design of solar-powered forced ventilation system and energy-efficient thermal comfort operation of greenhouse," Materials Today: Proceedings, vol. 46, pp. 9893-9900, 2021.

[13] L. Klimeš, P. Charvát, and J. Hejčík, "Comparison of the energy conversion efficiency of a solar chimney and a solar PV-powered fan for ventilation applications," Energies, vol. 11, no. 4, p. 912, 2018.

[14] S. Shastri, U. Sharma, and B. Singh, "Design and analysis of brushless dc motors for ceiling fan application," in 2020 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), 2020: IEEE, pp. 1-6.

Ensemble Tree Model based MPPT for Standalone Photovoltaic System

1st Mohamed Abderrazak Ghellouci
Electrical Engineering Department
Laboratory of Electrical Engineering and Renewable Energy, University of Souk Ahras Souk Ahras, Algeria
m.ghellouci@univ-soukahras.dz

2nd Samira Boumous
Electrical Engineering Department
Laboratory of Electrical Engineering and Renewable Energy, University of Souk Ahras Souk Ahras, Algeria
samira.boumous@univ-soukahras.dz

3rd Zouhir Boumous
Electrical Engineering Department
Laboratory of Electrical Engineering and Renewable Energy, University of Souk Ahras Souk Ahras, Algeria
zohir.boumous@univ-soukahras.dz

4th Sabrine Morakeb
Electrical Engineering Department
Laboratory of Electrical Engineering and Renewable Energy, University of Souk Ahras Souk Ahras, Algeria
s.morakeb@univ-soukahras.dz

5th Mohamed Mounir Abbes
Electrical Engineering Department
Laboratory of Electrical Engineering and Renewable Energy, University of Souk Ahras Souk Ahras, Algeria
m.abbes@univ-soukahras.dz

***Abstract*—This paper explores the application of Ensemble Tree Model (ETM) for maximum power point tracking (MPPT) in photovoltaic (PV) systems. The Gradient Boosting Machines method is employed to train (ETM), which offers rapid and precise (MPPT) capabilities. The (GBM) is used to determine the reference voltage corresponding to the maximum power point under varying atmospheric conditions. By effectively controlling the DC-DC boost converter, accurate (MPPT) can be achieved. MATLAB/Simulink simulations validate the theoretical analysis.**

***Index Terms*— Ensemble Tree Model (ETM), maximum power point tracking (MPPT), The Gradient Boosting Machines (GBM), photovoltaic systems (PV).**

I. INTRODUCTION

Countries worldwide are increasingly focusing on renewable energy development to address the impending depletion of fossil fuels. Renewable energy sources not only offer energy densities comparable to or exceeding those of fossil fuels but also provide a clean alternative without harmful emissions like nitrogenous compounds, sulfate compounds, and dust. Hydrogen, a promising candidate for future energy, aligns with these environmental and energy security goals due to its clean, sustainable, and emission-free nature. As a result, research into hydrogen production and applications is expanding rapidly.

Photovoltaic (PV) technology harnesses the sun's energy by converting sunlight into electricity. This process, known as the photoelectric effect, was first described by Edmund Becquerel in 1839. (PV) systems offer a clean and sustainable energy solution, requiring no fuel, producing no emissions, and having no moving parts. Their versatility allows for various sizes and shapes, and they rely on the abundant and inexhaustible energy of the sun. When sunlight strikes certain materials, it liberates electrical charge carriers, enabling the capture of light energy as electricity. This captured energy can be stored as direct current (DC) and then converted into alternating current (AC) for use in standard electrical power systems[1].

Photovoltaic (PV) arrays transform sunlight into direct current (DC) electricity. The generated power output varies with solar illumination and environmental temperature. Additionally, the load demand influences the power produced. Under consistent conditions, a (PV) array exhibits a current-voltage characteristic with a single point of maximum power output. To ensure optimal power delivery to the load, a Maximum Power Point Tracking (MPPT) algorithm is essential. This algorithm actively controls power converters to continuously identify and operate at the PV array's maximum power point [2].

II. MODELING PHOTOVOLTAIC SYSTEMS

II.1 A Simplified Model of a Photovoltaic Cell

An idealized photovoltaic (PV) cell can be represented by the equivalent circuit shown in Figure 1.

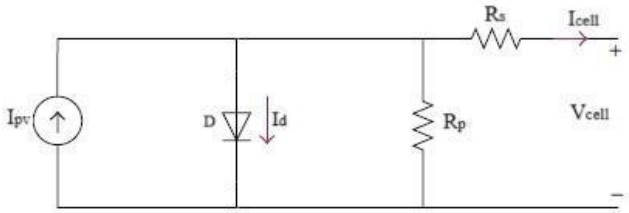


Fig1. Equivalent circuit of PV model.

The fundamental equation governing the (I-V) characteristic of this ideal cell is derived from semiconductor theory[3] thus :

$$I = I_{pv} - I_{0,cell} * \left[e\left(\frac{qV}{akT}\right) - 1\right] \quad (1)$$

$$Id = I_{0,cell} * \left[e\left(\frac{qV}{akT}\right) - 1\right] \quad (2)$$

with $(I_{pv,cell})$ represents the light-generated current, which is directly related to solar irradiance. (Id) is given by the Shockley diode equation, while $(I_{0,cell})$is the reverse saturation or leakage current of the diode. The symbol (q) denotes the electron charge $(1.60217646 \times 10^{-19}\,C)$, (k) is the Boltzmann constant $(1.3806503 \times 10^{-23}\,J/K)$, (T) is the temperature of the p-n junction measured in Kelvin, and $(a)$is known as the diode ideality factor[4].

II.2 Modeling a Photovoltaic Array

Equations (1) and (2) for the PV cell do not accurately depict the V-I characteristics of a real (PV) array. Practical arrays consist of multiple interconnected (PV) cells, and to observe the characteristics at the (PV) array's terminals, additional parameters must be incorporated into the basic equation [3,4]:

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

$$I = I_{pv} - I_0 * \left[e\left(\frac{V + R_s I}{V_t a}\right) - 1\right] - \frac{V + R_s I}{R_p} \qquad (3)$$

(I_{pv}) and (I_0) represent the (PV) current and saturation current of the array, respectively. The thermal voltage of the array, with N_s cells connected in series, is given by $V_t = \frac{N_s kt}{q}$.

When cells are connected in parallel, the current increases, and when connected in series, they provide higher output voltages. If the array consists of (N_p) parallel cell connections, the (PV) and saturation currents can be expressed as $I_{pv} = N_p I_{pv,cell}$ and $I_0 = N_s I_{0,cell}$.

In Equation (3), (R_s) represents the array's equivalent series resistance, while (R_p) is the equivalent parallel resistance. Equation (3) outlines the single-diode model illustrated in Figure 1[4].

(PV) array datasheets typically provide key information, including the nominal open-circuit voltage $(V_{oc,n})$, nominal short-circuit current $(I_{sc,n})$, voltage at the maximum power point (V_{mpp}), current at the maximum power point (I_{mpp}), open-circuit voltage, temperature coefficient (K_v), short-circuit current temperature coefficient (K_I), and the maximum experimentally achieved peak output power (P_{max}). These values are usually referenced to nominal or standard test conditions (STC_s), which specify temperature and solar irradiance levels. Some manufacturers also provide (I-V) curves for various irradiance and temperature levels, which can aid in fine-tuning and validating the desired mathematical (I-V) model. Generally, this is the standard information available from (PV) array datasheets[4].

Electric generators are broadly categorized as current or voltage sources, but practical (PV) devices exhibit a hybrid behavior—acting either as a current or voltage source, depending on the operating point. A practical (PV) device has a series resistance (R_s) that plays a more significant role when the device functions as a voltage source and a parallel resistance (R_p) that is more influential in the current-source region.

The series resistance (R_s) comprises several internal structural resistances, including the contact resistance between the metal base and the p-type semiconductor, the resistances of both the (P) and (n) regions, the contact resistance of the (n) layer with the metal grid, and the resistance of the grid itself[5].

The parallel resistance (R_p) mainly results from the leakage current across the $(P - n)$ junction and varies based on the cell's manufacturing process. The (R_p) value is usually high, and some models simplify calculations by ignoring it. Similarly, (R_s) is often very low, and in certain cases, it may also be disregarded to simplify the model.

The (I-V) characteristics of a (PV) array, as illustrated in Figure 2, are influenced by both the device's internal properties (R_s and R_p) and external factors such as the level of solar irradiation and temperature.

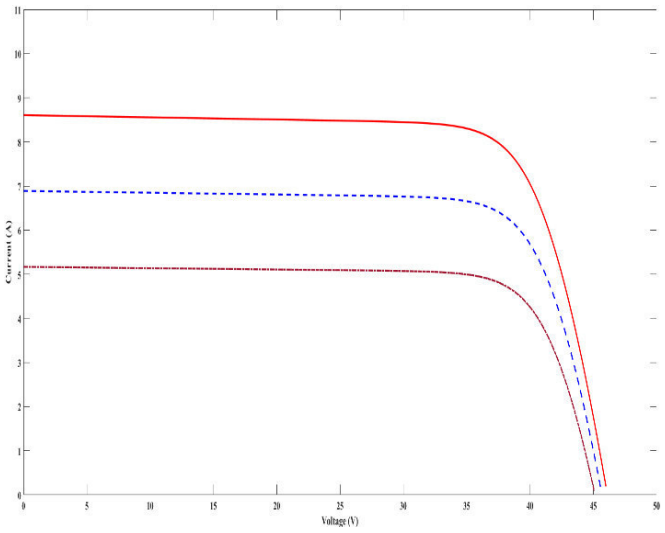


Fig2. I-V characteristic of PV.

The amount of sunlight incident on the device significantly impacts the generation of charge carriers, which in turn affects the current produced by the (PV) device. However, determining the light-generated current (I_{pv}) of the individual cells without accounting for the effects of series and parallel resistances is challenging.

Datasheets typically provide only the nominal short-circuit current $(I_{sc,n})$, which represents the maximum current output at the device's terminals.

In modeling (PV) devices, it is common to assume $(I_{sc,n} \approx I_{pv})$, as practical devices usually have low series resistance and high parallel resistance. The light-generated current of a PV cell varies linearly with solar irradiation and is also affected by temperature, as described by Equation (4),[2,4,6-8]

$$I_{pv} = \left(I_{pv,n} + k_I \Delta T\right)\frac{G}{G_n} \qquad (4)$$

Here, $(I_{pv,n})$ (in amperes) represents the light-generated current under nominal conditions (typically $25^{\circ}C$ and $1000\ W/m^2$), $\Delta T = T - T_n$ accounts for the difference between the actual temperature (T) and the nominal temperature (T_n) in Kelvin. (G) (in watts per square meter) is the irradiance on the device's surface, while G_n is the nominal irradiance. $(V_{t,n})$ denotes the thermal voltage of (N_s) series-connected cells at the nominal temperature (T_n).

The saturation current (I_0) of the (PV) cells that make up the device depends on the saturation current density of the semiconductor (J_0 typically given in A/cm^2) and the effective cell area[9]. The current density (J_0) is determined by the intrinsic properties of the (PV) cell, which are influenced by physical parameters such as the electron diffusion coefficient in the semiconductor, the minority carrier lifetime, and the intrinsic carrier density. In this study, the diode saturation current (I_0) is approximated as a fixed value of 0.81nA.

The diode constant (a) can be chosen arbitrarily, with many authors proposing methods to estimate its optimal value. Generally,

$1 \leq a \leq 1.5$, with the selection depending on other parameters of the I-V model. Empirical studies, provide values for (a) .

Since it represents the ideality of the diode and is determined empirically, any initial value can be selected for model adjustment. If needed, (a) can be refined later to improve model accuracy. This constant influences the curvature of the (V-I) curve, and adjusting (a) can help fine-tune the model to better match real-world performance[4].

III. Maximum Power-Point Tracking

III.1 MPPT methods

The maximum power of a solar array varies with changes in environmental temperature and irradiance (Figure 3). Since the available energy from solar panels continuously fluctuates due to atmospheric conditions, a real-time maximum power point tracker is an essential component of photovoltaic (PV) systems. Maximum power point tracking (MPPT) techniques discussed in the literature [10] can be classified into three categories [11].

1. **Direct methods**
2. **Artificial intelligence methods**
3. **Indirect methods**

Direct methods, also referred to as true-seeking methods, determine the maximum power point (MPP) by continuously adjusting the operating point of the (PV) array. Among these, the Perturb and Observe (P&O) [12], Hill Climbing (HC) [13], and Incremental Conductance (INC) [14] methods are widely used. The (P&O) method involves altering the (PV) array's operating voltage to find the (MPP), while the (HC) method similarly adjusts the duty cycle of the (DC-DC) converter. These methods are simple but exhibit steady-state oscillations, which limit their use in low-power applications. The (INC) method reduces steady-state oscillations by leveraging the fact that the slope of the power-voltage curve is zero at the (MPP).

Artificial intelligence and indirect methods aim to enhance the dynamic performance of (MPPT). Focusing on the non-linear characteristics of (PV) arrays, artificial intelligence methods offer fast but computationally intensive solutions to the (MPPT) challenge.

Indirect methods estimate the (MPP) from the output characteristics of the array. Fractional open-circuit voltage (OCV) [15] and short-circuit current (SCC) [16] approaches provide simple and effective means to identify the (MPP).

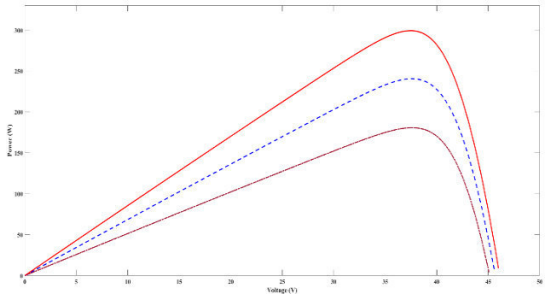


Fig3. P-V characteristic of PV.

$T = 303K, G = 600w/m^2$

$T = 300K, G = 800w/m^2$

$T = 298K, G = 1000w/m^2$

III.2 Ensemble Tree Models based MPPT

This paper presents an ensemble tree model approach for maximum power point tracking (MPPT) in photovoltaic (PV) systems. A gradient boosting machine (GBM) is employed to predict the optimal operating point of the PV system. The (GBM) takes temperature (T) and irradiance (G) as input variables and produces the voltage at the maximum power point (Vmp) as the output.

To train the (GBM), a comprehensive dataset of input-output pairs is generated using a (PV) system simulation model. The model is trained to accurately predict the (Vmp) under various operating conditions.

Once trained, the (GBM) model is integrated into the MPPT control system. The (GBM) receives real-time measurements of temperature and irradiance and predicts the optimal (Vmp).

The DC-DC boost converter is then adjusted to operate at the predicted (Vmp), ensuring maximum power extraction from the PV system.

The following equation illustrates how to calculate Duty Cycle :

$$D = 1 - \sqrt{\frac{Vmpp}{Impp} \times \frac{Iout}{Vout}} \qquad (5)$$

A resistive load, composed of four parallel-connected resistors, was employed at the output of the DC-DC boost converter.

Parameter	Value
Isc	8.56A
Voc	46.12V
I0	0.31nA
IPV	8.01A
A	1.037
Rp	201.5533Ω
Rs	0.37297Ω
Ns	72
Np	3
Kv	−0.362V/K
KI	0.046005A/K
LBOOST	0.20µH
CBOOST	27µF

Table1. PV System with MPPT, DC-DC Boost And Control Unit.

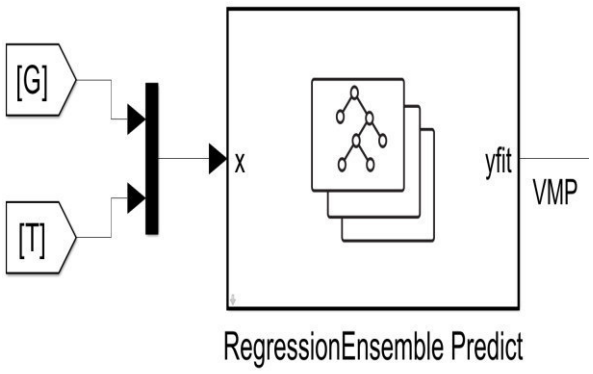


Fig4. Gradient Boosting Machines Structure.

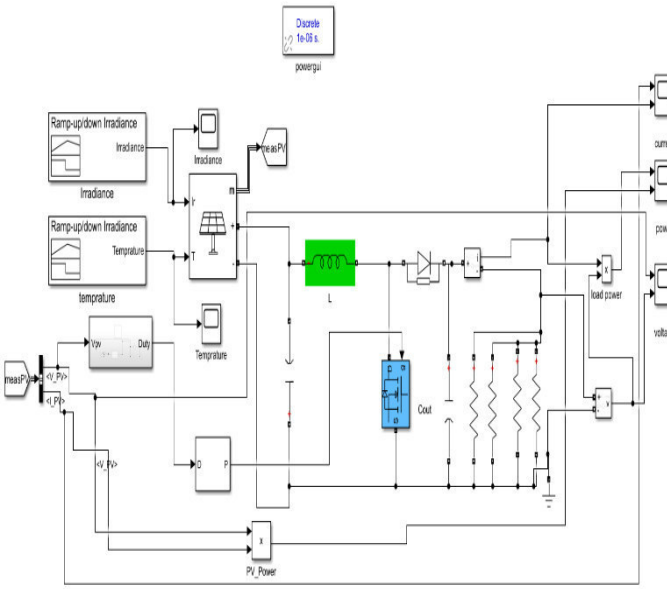


Fig5.PV System with MPPT, DC-DC Boost And Control Unit.

IV. Simulation and results

To validate the theoretical analysis discussed in the previous sections, a stand-alone photovoltaic (PV) system connected to a boost (DC-DC) converter is simulated using MATLAB/SIMULINK. The characteristics of the simulated model are detailed in Table 1.

Initially, the artificial ensemble tree model (ETM) is trained using a dataset comprising temperature, irradiance, and the voltage corresponding to the maximum power point (MPP).

The training process utilizes the scaled conjugate gradient method. Once the neuron weights are determined, the ensemble tree model is implemented in MATLAB/SIMULINK and integrated into the control unit.

For any given temperature (T) and irradiance (G) inputs to the GBM, the output automatically provides the reference voltage for the maximum power point (Vmpp).

The simulation is conducted under three distinct scenarios, each with three different temperature and irradiance levels. The specific temperature and irradiance values used in the simulation are illustrated in Figure 6(a),(b) respectively.

The ensemble tree model's output, representing the reference voltage of the maximum power point (Vmpp), for these three scenarios is shown in Figure 7(a),(b) respectively.

By applying (Vpv) and (Ipv) on DC-DC boost a duty cycle linked to (PWM) computed : switching pulses are generated to ensure that the (PV) output voltage and current track (Vmpp) and (Impp,).

Figure 8(a),(b) and Figure 9(a),(b) presents the output voltage and current of the (PV), demonstrating that they effectively follow (Vmpp) and (Impp).

A comparison between the maximum power reference and the power extracted from the PV system (Ppv,mpp) Figure 10(a),(b) confirms that the maximum power reference is successfully tracked.

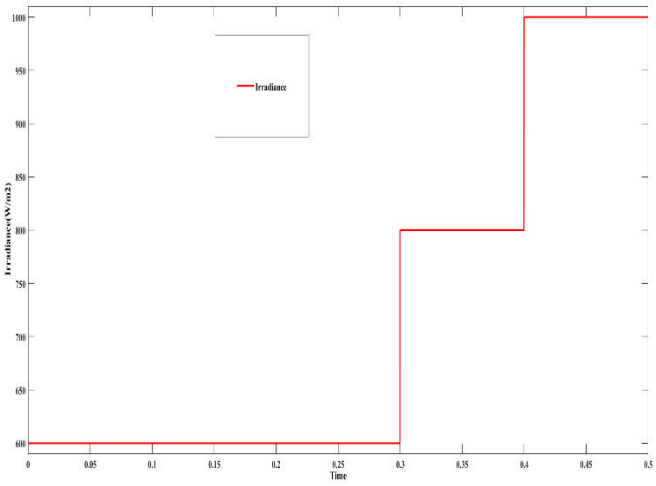


Fig6(a). Three different irradiances applied to simulation results.

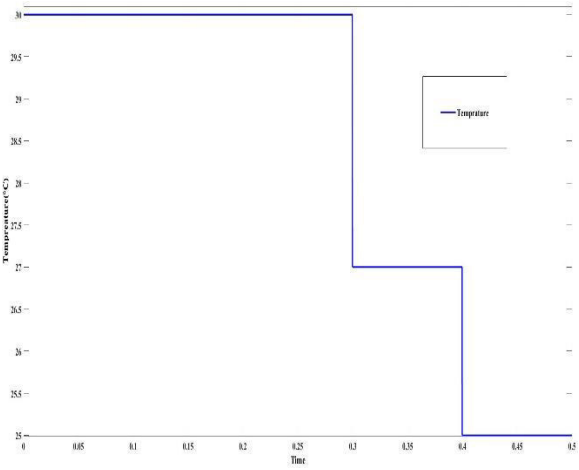


Fig6(b). Three different temperatures applied to simulation results.

In this figure:

The GBM model effectively predicts the optimal current for maximum power extraction.

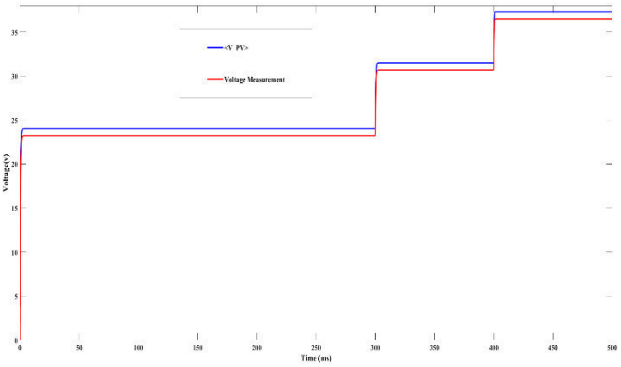


Fig7(a). Reference voltage of maximum power point generated by GBM, (b) Reference voltage of PV array.

In this figure :

The GBM model effectively predicts the optimal voltage for maximum power extraction.

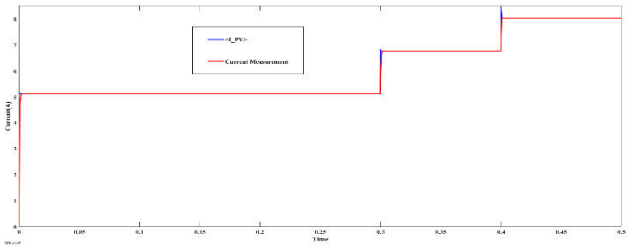


Fig9(a). Reference current of maximum power point generated by GBM, (b) Reference current of PV array.

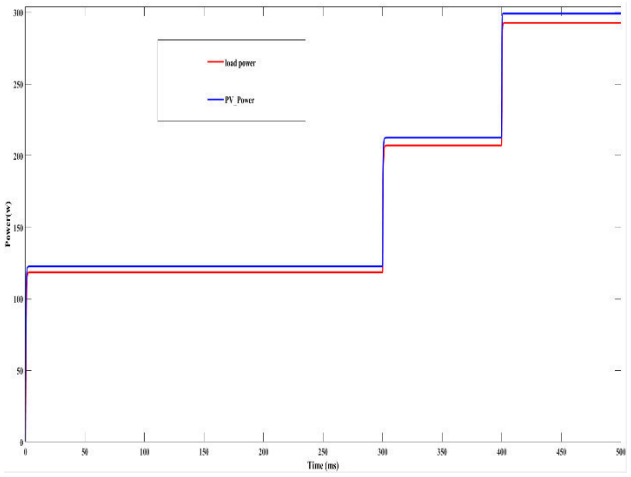


Fig7(a). Reference power of maximum power point generated by GBM, (b) Reference power of PV array

In this Figure :
The stepwise increase in both reference and actual power signifies changes in environmental conditions.
The GBM accurately predicts the MPP, leading to efficient power extraction.
During steady-state conditions, the system maintains the PV array at the MPP, ensuring optimal performance.

These figures highlight the control system's excellent dynamic performance in tracking voltage, current, and maximum power point.

V. Conclusion

This research explores the application of ensemble tree models for optimizing maximum power point tracking (MPPT) in photovoltaic (PV) systems. By leveraging the predictive capabilities of ensemble tree models, the system can accurately and efficiently identify the optimal operating point of the PV array, even under dynamic environmental conditions.

Compared to traditional MPPT techniques, ensemble tree model-based approaches offer superior dynamic performance, faster response times, and higher energy harvesting efficiency. The integration of a DC-DC boost converter further enhances the system's ability to extract maximum power from the PV array, ultimately maximizing energy yield.

The results obtained from the simulation demonstrates the effectiveness of the proposed approach. The ensemble tree model accurately predicts the optimal operating point, leading to significant improvements in power output and overall system efficiency.

References

[1] S. J. Lee, H. Y. Park, G. H. Kim et al., "The experimental analysis of the gridconnected PV system applied by POS MPPT," in *Proceedings of the International Conference on Electrical Machines and Systems (ICEMS '07)*, pp. 1786–1791, Seoul, Korea, October 2007.

[2] H. H. Lee, L. M. Phuong, P. Q. Dzung, N. T. Dan Vu, and L. D. Khoa, "The new maximum power point tracking algorithm using ANN-based solar PV systems," in *Proceedings of the IEEE Region 10 Conference (TENCON '10)*, pp. 2179–2184, Fukuoka, Japan, November 2010.

[3] H. S. Rauschenbach, *Solar Cell Array Design Handbook*, Van Nostrand Reinhold, NewYork, NY, USA,1980.

[4] M. G. Villalva, J. R. Gazoli, and E. R. Filho, "Comprehensive approach to modeling and simulation of photovoltaic arrays," *IEEE Transactions on Power Electronics*, vol. 24, no. 5, pp. 1198–1208, 2009.

[5] F. Lasnier and T. G. Ang, *Photovoltaic Engineering Handbook*, Adam Hilger, New York, NY, USA, 1990.

[6] W. De Soto, S. A. Klein, andW. A. Beckman, "Improvement and validation of amodel for photovoltaic array performance," *Solar Energy*, vol. 80, no. 1, pp. 78–88, 2006.

[7] Q. Kou, S. A. Klein, and W. A. Beckman, "A method for estimating the long-term performance of direct-coupled PV pumping systems," *Solar Energy*, vol. 64, no. 1–3, pp. 33–40, 1998.

[8] A. Driesse, S. Harrison, and P. Jain, "Evaluating the effectiveness of maximum power point tracking methods in photovoltaic power systems using array performance models," in *Proceedings of the IEEE 38th Annual Power Electronics Specialists Conference (PESC '07)*, pp. 145–151, Orlando, Fla, USA, June 2007.

[9] K. Nishioka, N. Sakitani, Y. Uraoka, and T. Fuyuki, "Analysis of multicrystalline silicon solar cells by modified 3-diode equivalent circuit model taking leakage current through periphery into consideration," *Solar Energy Materials and Solar Cells*, vol. 91, no. 13, pp. 1222–1227, 2007.

[10] T. Esram and P. L. Chapman, "Comparison of photovoltaic array maximum power point tracking techniques," *IEEE Transactions on Energy Conversion*, vol. 22, no. 2, pp. 439–449, 2007.

[11] V. Salas, E. Ol´ıas, A. Barrado, and A. L´azaro, "Review of the maximum power point tracking algorithms for stand-alone photovoltaic systems," *Solar Energy Materials and Solar Cells*, vol. 90, no. 11, pp. 1555–1578, 2006.

[12] O. Wasynczuk, "Dynamic behavior of a class of photovoltaic power systems," *IEEE Transactions on Power Apparatus and Systems*, vol. 102, no. 9, pp. 3031–3037, 1983.

[13] E. Koutroulis, K. Kalaitzakis, and N. C. Voulgaris,"Development of a microcontroller-based, photovoltaic maximum power point tracking control system," *IEEE Transactions on Power Electronics*, vol. 16, no. 1, pp. 46–54, 2001.

[14] K. H. Hussein, I. Muta, T. Hoshino, and M. Osakada, "Maximum photovoltaic power tracking: an algorithm for rapidly changing atmospheric conditions," *IEE Proceedings: Generation, Transmission and Distribution*, vol. 142, no. 1, pp. 59–64, 1995.

[15] J. J. Schoeman and J. D. vanWyk, "A simplified maximal power controller for terrestrial photovoltaic panel arrays," in *Proceedings of the IEEE 13th Annual Power Electronics Specialists Conference (PESC '82)*,pp. 361–367, 1982.

[16] M. A. S. Masoum, H. Dehbonei, and E. F. Fuchs, "Theoretical and experimental analyses of photovoltaic systems with voltage- and current-based maximum power-point tracking," *IEEE Transactions on Energy Conversion*, vol. 17, no. 4, pp. 514–522, 2002.

Hardware in The Loop Simulation of Sinusoidal Pulse Width Modulation Inverter˘s Controller Based on OPAL-RT & FPGA.

BOUYAHIAOUI Hadjer
Signals and Systems Laboratory,
Institute of Electrical and Electronic
Engineering, University M˘hamed
BOUGARA of Boumerdesy
h.bouyahiaoui@univ-boumerdes.dz

AZZOUGUI Yasmina*
Signals and Systems Laboratory,
Institute of Electrical and Electronic
Engineering, University M˘hamed
BOUGARA of Boumerdes
y.azzougui@univ-boumerdes .dz

Mohammed Salim BENMERABET
Direction Central Recherche et
D¶veloppement ¯ Sonatrach.
s.benmerabet@hotmail.fr

Abstract´ Hardware-in-the-loop (HIL) simulation is an essential tool for validating power electronics systems. This paper employed HIL simulation to test an FPGA-based controller for a three-level power inverter. In this paper two digital control circuits (Bipolar and Unipolar Sinusoidal Pulse Width Modulation) were designed in VHDL and integrated into the FPGA-based controller. A detailed MATLAB/Simulink model, utilizing Simscape blocks, emulated the three-level inverter for real-time testing. The HIL simulation results demonstrated a significant reduction in Total Harmonic Distortion (THD), improving the power quality of the inverter. The close agreement between HIL and pre-simulated results, with less than 21% error, validates the effectiveness of HIL simulation for power electronics systems.

Keywords´ Pulse Width Modulation (PWM), Total Harmonic Distortion (THD), Hardware in the Loop (HIL), Unipolar & Bipolar Sinusoidal Pulse Width Modulation, VHDL, FPGA, Inverter, OPAL-RT, RT-LAB.

I. Introduction

Renewable energy is becoming increasingly important as we move towards a more sustainable future. Power electronics play a vital role in the generation, transmission, and distribution of renewable energy. Inverters are a type of power electronics device that converts DC power to AC power. They are used in a variety of applications like Uninterruptible Power Supply (UPS), Variable Speed Drive (VSD) and in renewable energy (RE) such as solar photovoltaic systems, wind turbines, and electric vehicles.

The control of inverters is a complex task. The controller must ensure that the inverter output voltage and frequency are accurately controlled, even under varying load conditions [1]. Hardware-in-the-loop (HIL) simulation is a powerful tool that can be used to test and validate inverter controllers. HIL simulation allows the controller to be tested in real time with a simulated inverter model. This can help to identify and correct controller design flaws early in the development process, which can save time and money [2].

The objective of this paper is to investigate the use of HIL simulation for testing and validating inverter controllers. The specific focus of this paper is on the development and testing of an FPGA-based controller for a three-level power inverter. The controller was designed using VHDL and tested using the OPAL-RT HIL simulation tool. The results of the HIL simulation were compared to pre-simulated MATLAB÷ results and expected results from . The results showed that the controller was able to control the inverter output voltage and frequency under varying load conditions template.

II. Theoretical Background and Literature Review

A. The inverter

The inverter also known as DC -AC converter, is a one type of the power electronics that converts DC power to AC power at desired output amplitude and frequency, inverters can be voltage-source inverters (VSIs), current-source inverters (CSIs) and impedance-source inverters (ZSIs) [1]. Fig. II-1 represents a typical block diagram representation of VSI, CSI and ZSI, they are mostly used in AC power supplies (power inverters), motor drives (VFD) and systems where the objective is to produce a sinusoidal AC output whose magnitude and frequency both can be controlled. Practically, inverters are used in both single-phase and Multi-phase AC systems.

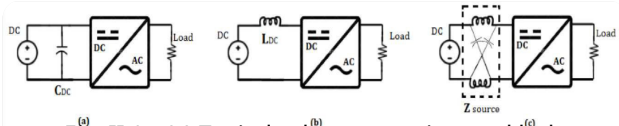


Fig. II 1 (a) Typical voltage-source inverter block diagram and (b) current¯source inverter block diagram (c) Z-source inverter indicating input energy storage elements

Inverters are available on the market in various configurations and power ratings. Fig. II-2 illustrate an abstract representation of inverters in a system where: element (1) represents the DC power source that feeds the inverter, it could be a battery, a rectifier or a DC power supply.

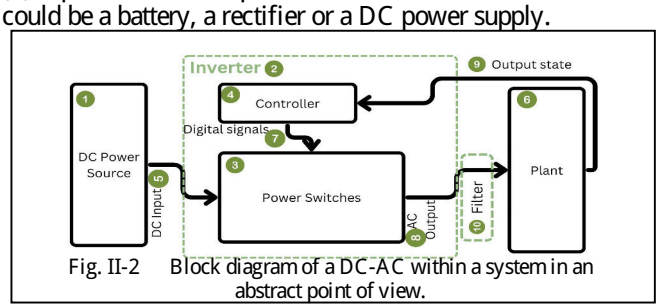


Fig. II-2 Block diagram of a DC-AC within a system in an abstract point of view.

element (2) is the inverter, in within element (3) represents a set of power semiconductor switches that are connected together in specific structure to direct the flow of DC current, there exist numerous different structures and topologies [1], these power semiconductor switches are controlled via element (4) the controller of the inverter, this last generates digital signals based on different control and modulation techniques aimed to turn ON or OFF the power semiconductor switches. The controller can be available as a separate off-the-shelf module as :converter˘s control unit˘ for specific application, as it could be a circuit integrated internally within the system, they can be single Integrated Circuit (IC) chips

XXX-X-XXXX-XXXX-X/XX/$XX.00 ø 20XX IEEE

that generate the control commands (either ASIC or FPGA-based controller) like the TL494 PWM IC, as they can be a SoC (System on Chip) with a microcontroller that generates the command signals, integrated with safety circuits and Human Machine Interface (HMI) to manipulate the output signals, etc. The rest of the element are summarized in Table I bellow:

TABLE I. INVERTER MAIN COMPONENTS

Element	Name	Description	Example
5	DC input	The Dc input is the DC power that feeds the inverter and will be converted to AC power	220Volts, 350Volts
6	The output load (plant)	The characteristics of the load can greatly affect the performance of the inverter. For instance, resistive loads are the easiest for an inverter to handle. However, reactive loads can be more challenging due to their power requirement.	Resistor, incandescent lights, AC motor, transformers
7	Digital signals	These digital signals control the switches, they are crucial for the operation of the inverter, they turn high or low depending on the type of the modulation controlling method used.	Pulse Width Modulation,
8	AC output	This output is the desired AC signal which has been converted from a DC signal.	Square wave, PWM signal
9	Output state	Feedback signal, generally used to control the actions needed to keep a specific variable of the inverter under control, for instance, in a motor drive system where the speed must be controlled electronically through the inverter. This signal requires sensors to be measured	Phase of the AC signal state,
10	Output filter	The output filter removes the high-frequency components of the PWM wave, to produce a nearly sinusoidal output.	Low pass LC filter

B. Basic Voltage-Source Inverter˜s structures

VSIs basic structures are the simplest inverter topologies that can output an AC signal from single DC source and can be classified as: 1) Single phase half bridge, 2) H-bridge structure and 3) Multi-phase inverter structures. In this paper we are interested in the H-bridge structure, also known as Full-bridge inverter or H-bridge cell as illustrated in Fig. I-3 is a basic inverter structure where the switching circuit consists of two half-bridge legs connected in parallel with a DC link capacitor C_d as shown in Fig. II-3, The switches can be controlled to provide Three-Level voltages ($-V_{dc}$, 0 and $+V_{dc}$), Table II-2 shows the possible switching states of the inverter and the corresponding output voltages. Fig. II-4 presents the output waveform across points A and B: $V_{AB}=V_{A0}-V_{B0}$, along with corresponding gating signals.

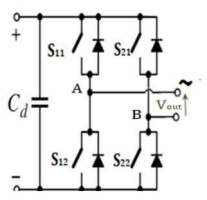


Fig. II 3 H-bridge inverter structure.

Table II-2 H-bridge state table

Switching State				Output
S_{11}	S_{12}	S_{21}	S_{22}	
0	1	0	1	0
0	1	1	0	$-V_{dc}$
1	0	0	1	V_{dc}
1	0	1	0	0

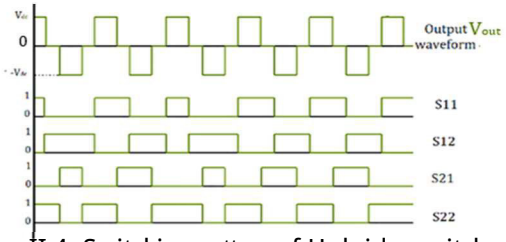


Fig. II 4 Switching pattern of H- bridge switches and the outputted three level pattern

III. SINUSOIDAL PULSE WIDTH MODULATION G

PWM is a technique for generating a variable-width pulse train from a fixed-frequency clock signal. The width of the pulse train is modulated according to a reference signal. The duty cycle is the ratio of the pulse in ON state (T_{on}) to the period of the clock signal T. A duty cycle of 50% means that the pulse is High for half of the period [3], in this case a rectangular waveform is generated. The duty cycle can be adjusted to control the average power delivered to the load [4].

To generate PWM gating signals, a reference signal, a carrier signal, and a comparator are needed. The maximum amplitude or peak value of the reference and carrier signals are denoted by $6_{Hēðéú}$ and $6_{ì ēðéú}$, respectively. Their frequencies are denoted by $ò_{éāí}$ $ò_{H}$ respectively.

The ratio of $6_{Hēðéú}$ to $6_{ì ēðéú}$ is known as the modulation index or amplitude modulation index, m_a:

$$m_a = \frac{V_{r,peak}}{V_{c,peak}} \quad (1)$$

The ratio of $ò_{}$ to $ò_{H}$ is known as the frequency modulation index, m_f :

$$m_f = \frac{f_c}{f_r} \quad (2)$$

Sinusoidal pulse width modulation or SPWM is a modulation technique that uses a sinusoidal as a reference signal to generate a pulse width modulated (PWM) signal. The PWM signal is then used to control the switching of power semiconductor devices in the inverter. In the case of an H-bridge inverter, four switching signals are required to control the four semiconductor devices as shown in Table II-2.

The two types of SPWM signals that can be generated by an H-bridge inverter are bipolar SPWM and unipolar SPWM [1] [5] [6] as shown in Fig. III-1.

A. Bipolar SPWM

Bipolar SPWM is generated when $3_{□□}$ is the logic-inverse of 傻 (i.e. $3_{□□}$ 嗯 this will generate a two level SPWM signal (-Vdc and +Vdc) [1] as shown in Fig. III-1 and illustrated in Table II-2.

B. Unipolar SPWM

Unipolar SPWM is generated when 傻 is leading 傻 by 180é as described in Table II-2, this will generate a three level SPWM signal (-Vdc, 0 and +Vdc) as shown in Fig. III-1, this is named Unipolar SPWM because in positive half cycle only levels (+Vdc, 0) occur while in negative half cycle only signal (-Vdc, 0) occur [7].

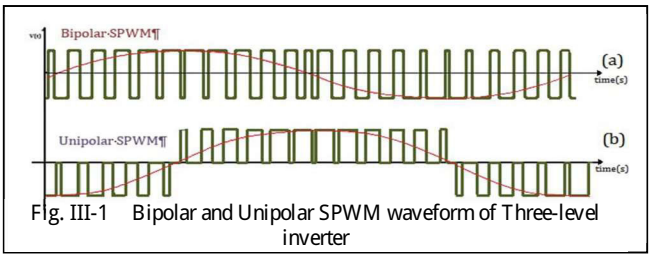


Fig. III-1 Bipolar and Unipolar SPWM waveform of Three-level inverter

IV. Overview of Real Time Simulation

Real-time simulation (RTS) runs computer models at the same speed as real life [8]. This allows for realistic interactions between simulated and real-world systems. While commonly used in gaming, RTS is also essential for industrial training and controller tuning. Tools like LabVIEW, SolidThinking Embed, and Simulink streamline the creation of real-time simulations.

Several commercial tools are available in the market to facilitate the development and deployment of real-time simulations [7], including: Nvidia, dSPACE, Typhoon HiL, OPAL-RT, PLECS, Siemens... etc. In this paper we used the OPAL-RT simulator: OP4510.

Real time simulation can be classified into: Full digital simulation like Software- in-the-Loop (SIL), Model-In-the-Loop (MIL), ... etc and semi-physical real time simulation such as Hardware-In-the-Loop (HIL), Rapid Control Prototyping (RCP), ... etc. In this paper both MIL and HIL were used, Model-In-the-Loop is the simulation of an embedded system in an early development phase of modeling in the field of model-based software development. Embedded systems communicate with their environment and often expect sensor signals as input and then stimulate the physical system. In order to function properly, the environment of the embedded system must be simulated. If the embedded system (model) is simulated in a loop together with the environmental model, this is called Model in the Loop Simulation. [3], it is a cost-effective way to test algorithms for embedded systems. Development and simulation environments for model-based development are, for example, MATLAB/Simulink, PLECS, Dymola, ASCET or the free software Scilab/X cos.

MIL is typically used in the early stages of embedded system development. It can be used to investigate specified model behaviors. Data collected during MIL simulation helps to validate that the model behaves as expected and can even be used for reference during the later stages and next design and testing phase. The system may need to be tested using Software in the Loop (SIL), FPGA in the Loop (FIL) or Power Hardware in the Loop (PHIL) simulations. These simulations are more complex than MIL, but they provide a more realistic test of the system's performance, for instance Hardware in the Loop is a procedure in which a real embedded system (e.g., real electronic control unit) is connected to a real time simulator via its inputs and outputs and test the controller on simulated environment, as illustrated in Fig. III-2, the HIL simulator serves as a replica of the system's real environment. If the target controller is not used and only a simulation of the software is performed, this is referred to as Software in the Loop (SiL) [9].

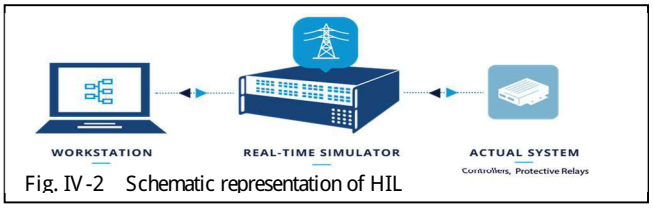
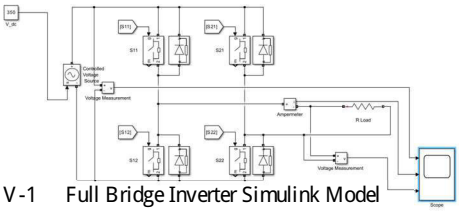


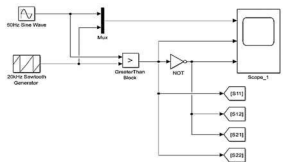
Fig. IV-2 Schematic representation of HIL

V. MATLAB® Simulation and Results

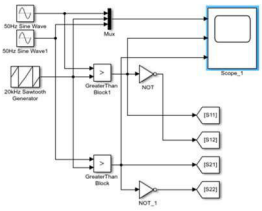
As mentioned earlier, before moving to HIL an MIL simulation was first implemented were the DUT is FPGA-based inverter's controller. The testing environment of this DUT contains a single-phase H-bridge inverter, a resistive load and wire connections. Therefore, a model for the environment (load and H-bridge inverter) was implemented first as shown in Fig. V-1 then a model for the DUT, as mentioned previously; two different SPWM controller circuits were designed and tested: Bipolar Sinusoidal PWM and Unipolar Sinusoidal PWM as shown in Fig. V-2 and Fig. V-3 respectively.


Fig. V-1 Full Bridge Inverter Simulink Model

The first circuit that have been implemented was the Bipolar SPWM as shown in Fig. V-3 since it is the simplest SPWM modulation circuit, it is composed of a single sine signal as a reference and a sawtooth signal as a carrier and a comparator that outputs the gating signals, after getting the results we improved the controller by designing a more complex SPWM controlling circuit which is Unipolar SPWM as shown in Fig. V-4, which is composed of an additional sine signal that is leading the other reference signal by 180é.


Fig. V-2 Bipolar SPWM control circuit blocks.

In both controllers we set the modulation index 侄 to 0.9 $m_a = \frac{V_{r,peak}}{V_{c,peak}} = 0.9$, the frequency of the carrier is 偏嗯 and the frequency of the refence is 偏嗯 侑. The modulation frequency index is 侄 嗯 嗯


Fig. V-3 Unipolar SPWM control circuit blocks.

The output of the scopes in Fig. V-2 and Fig. V-3 are observed in Fig. V-4 and Fig. V-5 respectively, which represents the expected gating signals. As can be observed in Fig. V-4, the duty cycle of S11and S22 is proportional to the reference signal, contrary to S12 and S21 which is inversely proportional to the reference signal. In other hand the output of the scope in Fig. V-3 is observed in Fig. V-5, as we can see the duty cycle of S11 is proportional to the red reference signal, contrary to S21 which is proportional to the blue reference signal.

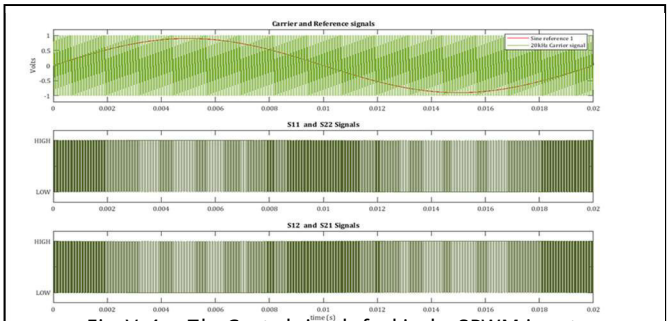


Fig. V-4 The Control signals for bipolar SPWM inverter

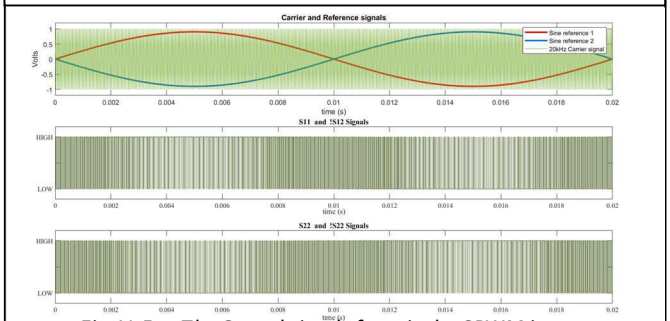


Fig. V-5 The Control signals for unipolar SPWM inverter

These values have been saved as a reference to compare them with the output of the real FPGA-Based controller.

The results from the scopes: after running the simulation the obtained results are shown in Fig. V-6 and Fig. V-7, due to the high switching frequency, some output signals appear as a rectangle, we calculated their mean value which is inserted and shown in red. In Bipolar SPWM circuit˜s output the mean value of the output voltage in red is similar to the reference voltage for the different loads. The THD of the output voltage is high and equals to 121.35%, but the THD of two-level inverters is generally very large and varies depending on the modulation signals [1].

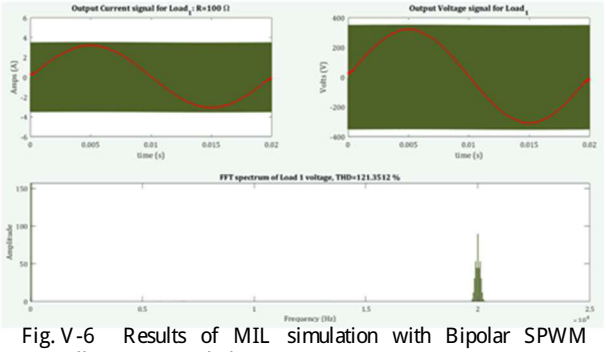


Fig. V-6 Results of MIL simulation with Bipolar SPWM controller using Simulink

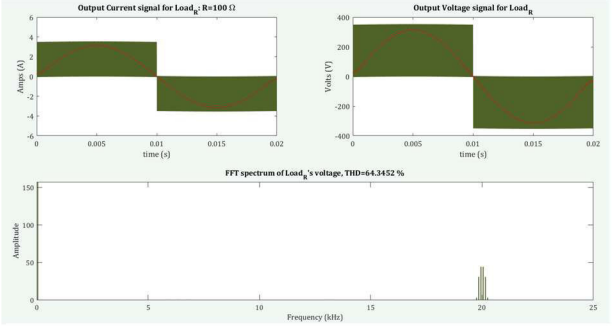


Fig. V-7 Results of MIL simulation with Unipolar SPWM controller using Simulink

In other hand, the output of the unipolar SPWM allows the appearance of a third voltage level as shown in Fig. V-7 it is a three-level inverter output using an H-Bridge inverter. The overall THD is lower than the bipolar strategy, reaching 63.28%. these results were saved as reference for next HIL simulation.

The next step is to generate real gating signals using FPGA board. These days most small and medium-sized three-level inverters˜ switching frequency is more than 10 kHz [10]. If we want to carry out accurate and real-time hardware-in-the-loop simulation test, the simulation based on CPU cannot meet the requirements, while the simulation based on FPGA is needed, the minimum simulation step length is 2us which can easily be generated by FPGAs, in this paper a DE2 development board that contain a Cyclone II FPGA chip from ALTERA (Intel) was used as the FPGA-based controller. After generating these real-time signals, we used the oscilloscope to test and compare the output signals with the MIL results that were presented earlier in this scope-

VI. Hardware In The Loop Setup of The Experiment

After testing the controller˜s output, the hardware setup shown in Fig. VI-1 for HIL simulation has been adopted: we used op4510 as the Real-time simulator from OPAL-RT TECHNOLOGIES, the OP4510 is an entry-level simulator that contains an FPGA carrier, which can accept four standard OPAL-RT mezzanine boards [11], contains two digital and two analog boards in addition to the RS422 signals, these mezzanine boards interface using a DB37 connector at the back of the chassis of the op4510, In general, the simulator (target simulator) is connected to a computer (host computer) as illustrated in Fig. VI-2 via ethernet cable with TCP/IP communication protocol. The host computer is supposed to send and control the model, display the graphical results, and receive or read the stored simulation results files in the simulator and some other functions. While the target simulator is to load, build and run the model, during simulation the simulator can store the data without affecting the simulation operation˜s results and at the end it sends the file to the host computer, in our case we used a standard computer as host, we used FPGA as a programable chip to design the controller of the three-level inverter shown in Fig. VI-2 ,



Fig. VI-1 The Final Hardware Setup of the experiment

RT-LAB is a software that connects the Simulink models and the real-time simulator [11], building an HIL simulation in RT-LAB involves creating a model that includes Simulink blocks, RT-LAB library˜s blocks and respect some specific requirements. In general, Any Simulink model can be implemented in RT-LAB, but some modifications must be made to distribute the model and transfer it into the simulation environment.

RT-LAB˜s model consists of two types of subsystems: one that contains the computational elements of the model,

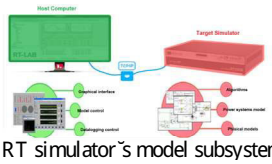


Fig. VI-2 Opal RT simulator˜s model subsystems

(named Master & Slaves) and a consol subsystem that is used as a graphic interface subsystem [11]. In top view there must be only one Master subsystem but multiple Slave subsystems could be added; depending on the number of targets in the model; The data between the computational subsystems and GUI subsystems will be exchanged asynchronously through the TCP/IP link.

The top view of our RT-Lab model is shown in Fig. VI-3, Fig. VI-4 and Fig. VI-5 represent the H-bridge model inside the master subsystem and the consol subsystem respectively.

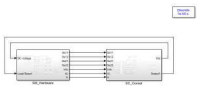


Fig. VI-3 The top view of our HIL OPAL-RT model

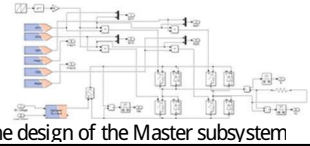


Fig. VI-4 The design of the Master subsystem

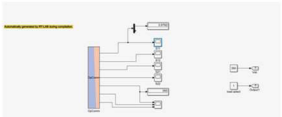


Fig. VI-5 The Design of the Consol sub-system.

VII. FPGA-BASED CONTROLLER DESIGN

The controller's software was designed using VHDL, implemented with Quartus II, and tested against MIL simulations. After successful verification, the FPGA outputs were connected to the op4510 for real-time HIL simulation using RT-LAB. The aim is to create a Bipolar SPWM control circuit for an H-bridge, ensuring compliance with constraints (1) and (2) with

$$m_f = \frac{20kHz}{50Hz} = 400 \quad \text{and}$$

$$m_a = \frac{V_{r,peak}}{V_{c,peak}} = 0.9$$

therefore a Bipolar SPWM circuit similar to the one in Fig. V-3 was built using VHDL and presented in Fig. VII-1, both designs contain same elements (a sine signal, a sawtooth signal and a comparator): the final VHDL design contains four main components (blocks) in top-level.

The main functionalities of the above VHDL˜s Bipolar SPWM controller design is summarized in the next table.

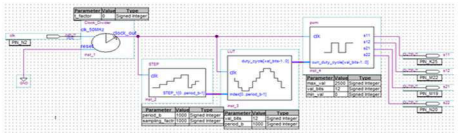


Fig. VII-1 Bipolar SPWM circuit schematic

TABLE II. PSUEDO CODE OF THE BIPOLAR SPWM SYSTEM

1. Generate main clock signal
2. Increase STEP_1 internal counter every clock signal
2. In parallel, update the output of the LUT table.
3. Compare the counter in pwm block with output of LUT block.
4. Generate the Bipolar PWM gating signal based on the output of the comparison

The outputs gating signals s11, s12, s21 and s22 are presented in Fig. VII-2, The gating signals appear to resemble the gating signals generated using Simulink shown in Fig. V-4 , it can be observed in the bellow graph that the frequency of the gating signal is 50Hz, also s11 and s22 are similar while s12 and s21 are the inverse of s11 and that˜s exactly as expected, therefore these gating signals we connected to the op4510 to test theme in HIL simulator.

In other hand, Fig. VII-3 shows the Unipolar SPWM controlling circuit as mentioned previously the Unipolar SPWM controlling circuit require two sinusoidal references that are out-of-phase, additionally each leg of the H-bridge inverter is controlled based on the comparison between the sawtooth and one of the references; While there are similarities between Unipolar and Bipolar SPWM circuits, Unipolar SPWM utilizes an additional reference. Leveraging the previously designed VHDL blocks from the Bipolar SPWM implementation, we created the Unipolar SPWM circuit as depicted in Fig. VII-3. The final design contains four top-level blocks: 'Clock_Divider', 'STEP', 'pwm', and 'ULUT'.



Fig. VII-2 Gating signals generated from the FPGA-Based controller

Table III summarizes the main functionalities of this VHDL-based Unipolar SPWM controller.

TABLE III. PSUEDO CODE OF THE UNIPOLAR SPWM SYSTEM

1. Generate main clock signal.
2. Increase STEP block˜s internal counter every clock signal.
2. In parallel update the output of the ULUT table.
3. Compare the counter in pwm blocks with outputs of ULUT block.
4. Generate the Unipolar PWM gating signal based on the output of the comparison

As shown in Fig. VII-3 the final Unipolar SPWM controller design has two inputs labeled :clk˜ and :reset˜ and five outputs s11, s12, s21 and s22, and bug, the output of the gating signals is presented in Fig. VII-4.

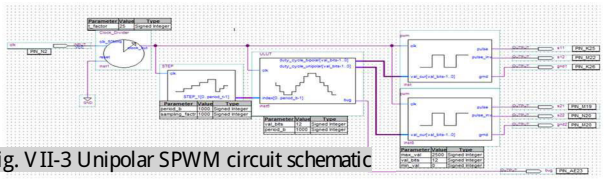


Fig. VII-3 Unipolar SPWM circuit schematic

Fig. VII-4 (a) shows that the frequency matches the design in Fig. VII-3 and Fig. VII-4 (b) shows that the duty cycle of S11 is leading S21 by 180∘therefore these gating signals we connected to the op4510 to test theme in HIL simulator.



Fig. VII-4 Unipolar SPWM gating signals generated from the FPGA-Based controller

VIII. RESULTS AND COMPARISON

As shown in Fig. VI-5, we put multiple scopes in the consol subsystem to help us view the behavior of the signals during the HIL simulation. The gating signals generated from the FPGA-Based controller are shown in Fig. VII-2 and Fig.

VII-4 for Bipolar SPWM and Unipolar SPWM control circuits respectively:

Fig. VIII-1 and VIII-2 present the HIL simulation results for Bipolar and Unipolar SPWM controllers, respectively. We plotted in red color the mean value of the output waveforms to help us observe the change of duty cycle of the output. Fig. VIII-1 shows the Bipolar SPWM results, highlighting a significant THD (145.80%) in the R load output voltage, compared to 112.38% in MIL. Unipolar SPWM (Fig. VIII-2) reduced the THD to 69.3599%, closer to the reference (63.286%), similarly to the reference the harmonics are significant at 20kHz.

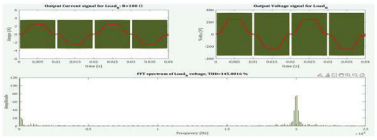


Fig. VIII-1 Results of HIL simulation with Bipolar SPWM controller using Simulink

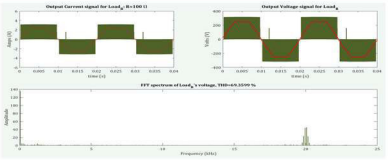


Fig. VIII-2 Results of HIL simulation with Unipolar SPWM controller using Simulink

After obtaining the HIL results we calculated the percentage of errors of the THD to see which design is better, using the following formula and presented the values in Table III-2:

$$\text{Error } \% = \frac{\left| \text{Reference values} - \text{Experimental values} \right|}{\text{Reference values}} * 100$$

Table IV summarizes the error percentages. Bipolar SPWM exhibited a 20.14% error, potentially due to external noise or instrument malfunction. Unipolar SPWM demonstrated a lower error of 9.59%, indicating closer alignment with MIL results. This suggests that HIL is a suitable testing method for inverter controllers before implementing them with a real IGBT H-bridge inverter (Fig. VIII-3) as shown in Fig. VIII-3, additionally we could compare our results with processor in the loop of single phase H-bridge inverter outputs from [12], were they obtained THD of the output voltage across the output load using TMS320F28379d processor as controller, Table IIV summarizes the error percentages between our results and results from [12]. Bipolar SPWM exhibited very high error percentages, Unipolar SPWM demonstrated a lower error of 30.8%, indicating closer alignment with our results.

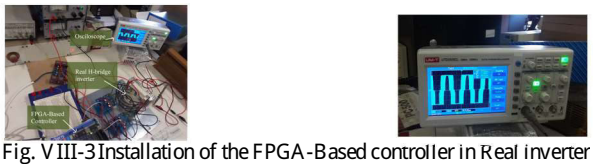


Fig. VIII-3 Installation of the FPGA-Based controller in Real inverter

TABLE IV. ERROR PERCENT BETWEEN HIL AND MIL RESULTS

		Reference values	Experimental HIL results	Error percent %
THD of Voltage across the $Load_R$: R=100ῼ	Bipolar SPWM	121.3512%	145.8016%	20.14
	Unipolar SPWM	63.286%	69.3599%	9.59

As shown in Fig. VIII-3, the H-bridge inverter successfully generates the desired output waveform. The output signal closely matches the reference signal, indicating that the controller effectively drives the inverter.

TABLE V. ERROR PERCENT BETWEEN HIL AND RESULTS FROM [12]

	Reference [12] outputs	Experimental HIL results	Error percent %
Bipolar SPWM	70%	145.8016%	108.288
Unipolar SPWM	53%	69.3599%	30.8

CONCLUSION

This project aimed to develop and test an FPGA-based controller for an H-bridge power inverter. We also investigated the use of Hardware-in-the-Loop (HIL) simulation, specifically with OPAL-RT, for validating these controllers. Our research demonstrated the effectiveness of HIL simulation for evaluating FPGA-based inverter controllers. By extensively testing both Bipolar and Unipolar Sinusoidal Pulse Width Modulation (SPWM) control algorithms, we observed minimal error between Model-in-the-Loop (MIL) and HIL results. This suggests that HIL is a highly promising testing strategy for achieving high-quality inverter performance.

REFERENCES

[1] E. C. Dos Santos and E. R. Cabral Da Silva, ADVANCED POWER ELECTRONICS CONVERTERS:PWM Converters Processing AC Voltages, Canada: IEEE Press, 2015.

[2] L. P. S. Anjali Krishna R, "A Brief Review on Multi Level InverterTopologies,," International Conference on Circuit, Power and Computing Technologies, pp. 1-6, 2016.

[3] P. K. SACHIN MAHESHRI, "Simulation of single phase SPWM (Unipolar) inverter," International Journal of Innovative Research in Advanced Engineering (IJIRAE), vol. 1, no. 9, pp. 2349-2163, October 2014.

[4] M. N. Bin Wu, HIGH-POWER CONVERTERS AND AC DRIVES, Hoboken, New Jersey and Canada: John Wiley & Sons, Inc, 2017.

[5] A. C. A. A. S. M. Obbu Chandra Sekhar, "Simulation of Three-Level Diode Clamped Multilevel Inverter Using SPWM Technique, Space Vector Strategy & SHE Technique," RECENT, vol. 17, no. 2, pp. 107-115, 2016.

[6] K. C. J. M. K. D. Akanksha Sinha, "An inclusive review on di喀erent multi-level inverter topologies, their modulation and control strategies for a grid connected photo-voltaic system," Elsevier Ltd, p. 633¯657, 1 June 2018.

[7] J. S. H. B. Jana, "A review of inverter topologies for single-phase grid-connected photovoltaic systems.," Renew. Sustain. Energy Rev, vol. 1, no. 71, p. 1256¯1270, 2017.

[8] J. Belanger, P. Venne and J. N. Paquin, "The what, where and why of real-time simulation.," Planet Rt, vol. 1, no. 1, pp. 25-29, 2010/10/4.

[9] Y. S. ,. R. M. Mohamed A.Ghalib, "Design and Implementation of a Pure Sine Wave Single Phase Inverter for Photovoltaic Applications," Conf. Informatics, Electron, pp. 1-8, 2014.

[10] Mathworks, "What is FPGA-In-the-Loop," Mathworks, [Online]. Available: https://www.mathworks.com/help/hdlverifier/ug/fpga-in-the-loop-fil-simulation.html. [Accessed 7 4 2024].

[11] Opal-rt Academy, "RT-LAB Documentation," 4 2023. [Online]. Available: https://opal-rt.atlassian.net/wiki/spaces/PRD/pages/144249433/Mezzanines. [Accessed 9 2023].

[12] B. Satabdi and S. Susovon, "Comparative Analysis of Various Sinusoidal Pulse Width Modulation Techniques," Third International Conference on Power, Control and Computing Technologies (ICPC2T), pp. 200-205, 2024.

MPPT Design based on partial Swarm optimization and predictive control for photovoltaic system

Nora .KACIMI
Electrical Engineering system
Department , Faculty of technology
Boumerdes university.
E-mail : kacimi.nora@yahoo.fr

Abdelhakim .IDIR
Institution: Faculty of Technology
University of M'sila F.H.C,.
University of Boumerdes.
E-mail: abdelhakim.idir@univ-msila.dz

Belgacem. MECHERI
Electrical Engineering system
Department , Faculty of technology
Boumerdes university.
E -mail: mecheri.begacem@univ-boumerdes.dz

Aicha .DJALAB
Electrical Engineering system
Department , Faculty of technology
Boumerdes university.

Said .GROUNI
Tamangrasset university
E-mail : said.grouni@yahoo.com

***Abstract*—Photovoltaic (PV) systems under partial shading condition have more one maximum power point (MPP), as consequence multiple peaks. The present paper develops and discusses global maximum power point tracking (GMPPT) algorithm under partial shading condition. The main contribution of the suggested method is to control DC/DC converter using a Particle Swarm Optimization (PSO) and predictive control (MPC) the maximum peak. PSO is used to locate maximum power and MP to tracking it. The proposed system is simulated and evaluated using Matlab/Simulink The performance of proposed approach detected the Global Peak with fewer power oscillations.**

Keywords — Global maximum power, Predictive control. Photovoltaic system, Particle swarm optimization.

I. Introduction

In recent decades, electric power generations by renewables energy have been increased. Solar energy is one of the most exploited renewable energy resources due to several advantages [1]. Currently PV systems encounter many challenges to maximize their efficiency and output power under varying weather conditions and particularly low irradiance levels; it is difficult for PV systems to provide the maximum output, which results in low efficiency [2]. The PV system have a unique maximum power point (MPP), at which operates with the maximum output and the highest efficiency. The control technique to achieve maximum power point tracking (MPPT) is a very important technology. Various technique for maximum power point tracking (MPPT) have been proposed in literature as perturb and observation ,incremental conductance (INC) [3], Fuzzy logic control, [4] Neural network[5]….etc.

These methods have simple structures but they may be stuck at a local MPP (LMPP) instead of tracking the global MPP (GMPP) under partially shaded conditions (PSC) because P–V curve displays multiple peaks with several local peaks and a global peak (GP), which cannot be differentiated by conventional algorithms. Global maximum power point tracking (GMPPT) is the technique to extract the possible maximum amount of power from the installed photovoltaic (PV) system under the partial shading conditions [6]

Various methods have been proposed for tracking the MPP in partial shading conditions are proposed [7-12]. Authors in Reference [13] have proposed to use a hybrid Artificial Neural Network and Particle Swarm Optimization algorithm to detect the global peak power. The ANN algorithm initializes optimal voltage which is used then in the PSO algorithm to reach the GP location. In this study, we proposed a combination the particle swarm optimization algorithm and predictive control to extract the GMPP under partial shaded condition.

Particle Swarm Optimization (PSO) is used to search for continuous variable for optimization problems [14]. In order to get a fast response and ensures a good transition of energy between the PV system and the load. We proposed also to use the predictive control .The model predictive control used to solve a finite-horizon optimal control problem at each sampling instant and get control actions for both the present time and a future period [15]. In practical, the modification in real time is a desirable property, which can compensate inevitable modeling error. The implementation of a PV array MPPT using MPC combines two important keys, speed, and reliability, evading intolerable oscillations despite the increased speed. In fact, model predictive control reconstructs instant operating model of the system at each sampling time, then predicts future states and optimizes dynamic response while taking into account the future states [16].

The remainder of this paper is structured as follows: Section 2 briefly presents the model of the solar PV and Boost converter. In Section 3 describes the algorithms that track the GMPP and details of used the PSO algorithm and predictive control. Next the simulations results are presented in Section 4. Finally, an appropriate conclusion and future work are pointed.

II. PV array under partial shading conditions

II.1 - Mathematical representation of solar PV array

A photovoltaic cell is an electrical device which converts the energy of light directly in to electricity by the photovoltaic effect. Figure 1 shows the one-diode equivalent circuit used to obtain the model of PV cell [17].

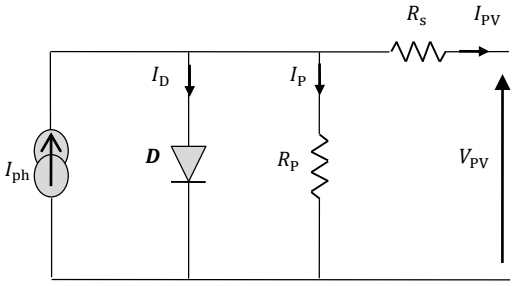


Figure 1. Equivalent electrical circuit of one diode for PV solar

The relation between the current I_{pv} and the voltage V_{pv} is given by equation (1):

$$I_{PV} = I_{Ph} - I_0 * \left(e^{\frac{V_{PV}+R_S*I_{PV}}{Vt*A}} - 1\right) - \frac{V_{PV} + R_s * I_{PV}}{R_P} \quad (1)$$

Where:
Rs is the equivalent series resistance ;,
I_{ph}, is the current generated by the incident light; Rp is the equivalent parallel resistance and

$$V_t = \frac{N_S*T*K}{q} \quad (2)$$

Where:

T is the temperature,
q is the charge of an electron,
k is the Boltzmann constant, and
Ns is the number of cells connected in series.

A PV array consists of several connected PV panels formed by solar cells connected in series and parallel. Equations 3 represent a PV array [18].

$$I_{PV} = N_{pp} I_{Ph} - N_{pp}\, I_0 * \left(e^{\frac{N_{ss}V_{PV}+R_s*I_{PV}(N_{ss}/N_{pp})}{Vt*A*N_{ss}}} - 1\right) - \frac{N_{ss}V_{PV} + R_s * I_{PV}(N_{ss}/N_{pp})}{R_p(N_{ss}/N_{pp})} \quad (3)$$

Where
Npp is the number of PV panels connected in parallel and Nss is the number of PV panels connected in series.

II.2 .DC/DC Boost converter

In order to achieve MPPT in photovoltaic systems, an intermediate DC/DC converter is connected. In our study the boost circuit is chosen. (figure 2).

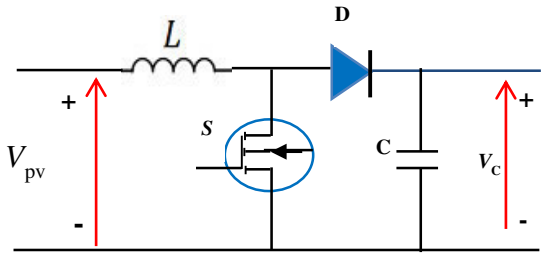


Figure. 2. Equivalent circuits of boost converter

The operation equation of boost converter can be described as by follows M

$$\begin{cases} L\frac{di_L}{dt}(t) = -(1-s)V_c(t) + V_{pv}(t) \\ C\frac{dV_c}{dt}(t) = (1-s)I_L(t) + \frac{1}{R}V_c(t) \end{cases} \quad (4)$$

Where, L, C, R represent: the inductance, capacitance and resistance load of the boost converter respectively. And S the switching state, which can be taken to 1 or 0.

Using the first order approximation for the derivative in (5), (6), the predicted current and voltage in discrete time are given by [19]:

$$I_L(k+1) = I_L(k) - (1-S)\frac{T_s}{L}V_c(k) + \frac{T_s}{L}V_{pv}(k) \quad (5)$$

$$V_c(k+1) = V_c(k) + (1-S)\frac{T_s}{C}I_L(k) + \frac{T_s}{RC} \quad (6)$$

Where Ts is sample time. S is switching state which can present by '1' or'0'.

The variables I_L and V_c can be predicted for one step horizon. The principal characteristic of Model Predictive Control is predicting the future behaviour of the control variables [20]. In this paper, the model predictive control used is based on a discrete-time of current system. . The predicted variables will be used to obtain the optimal switching state by minimizing a cost function. In this work in one-step horizon MPC, the variables I_{PV}, V_{PV} and V_C are measured in time k and adopted to estimate the future comportment in time (k+1) as show in figure 3. The implemented switching state is determined by optimizing a cost function g_s [21].

$$g_{s=0,1} = I_L(k+1) - I^*_{ref} \quad (7)$$

To determine the reference current, I^*_{ref} the proposed approach used PSO .

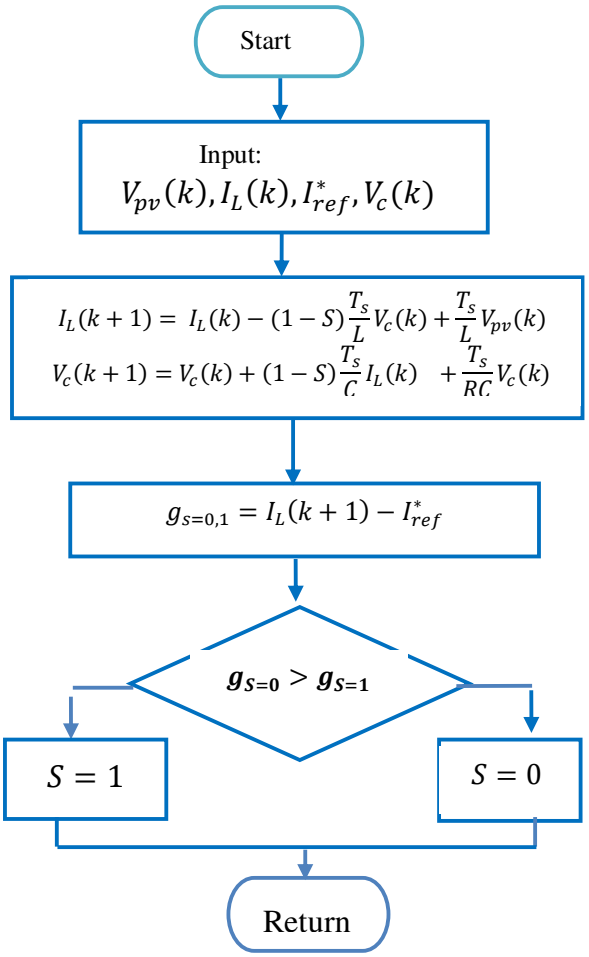


Figure 3. Flowchart of the MPPT algorithm

III. Global Maximum Power tracking MPPT based PSO- MPC

The particle swarm optimization (PSO) is a powerful searching algorithm which was modeled from the pattern observed in bird flocking and fish schooling [1,22].Figure.3 shows a block diagram of proposed system .

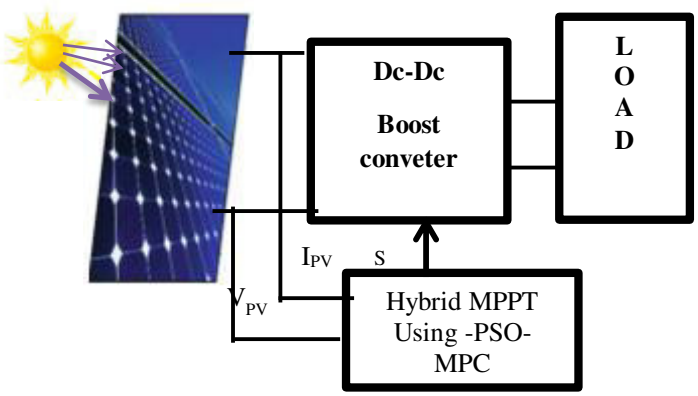


Figure 3. Block diagram of proposed system.

In this algorithm, a group of unique agents, called a swarm of particles is used. Each particle has a position in the search space of the optimization problem . and The algorithm discovers the fitness value and selects the optimal solution from whole swarm. The particle position is given by:

$$X_i(k+1) = X_i(k) + V_i(k+1) \quad i = 1,2,.N) \quad (8)$$

Where
vi(k) is the velocity of particle i,
k is the number of perturbation iteration and
N is the total number of particles.

The velocity is determined by

$$V_i(k+1) = \omega V_i(k) + C1r1(P_{best,i}\, X_i(k)) + C2r2(G_{best} - X_i(k)). (10)$$

where ω is the inertia weight, c Pbest,i is the best, and Gbest is the best solution of the whole swarm

The different steps of PSO are as follows:[23]

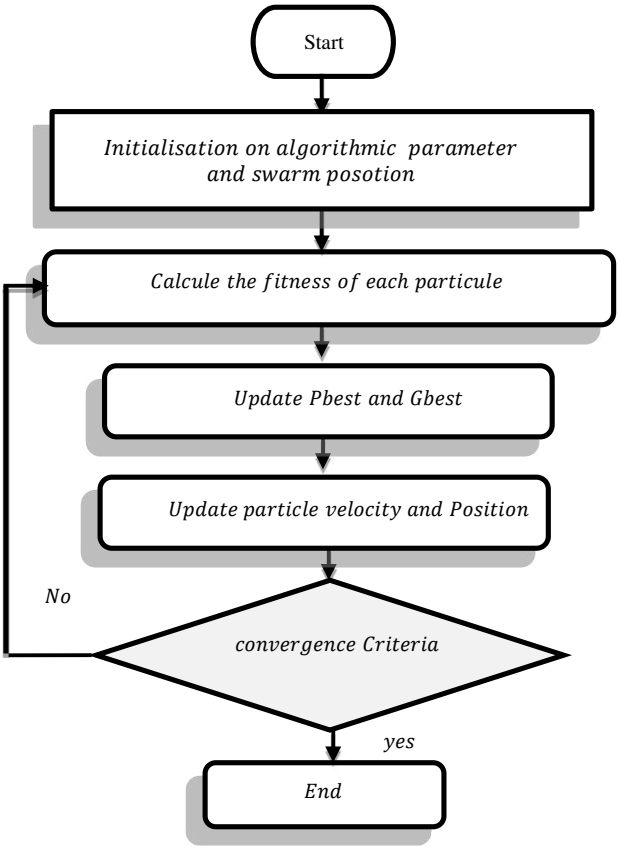


Fig 7. Flowchart diagram of the implemented MPC.

IV. Simulation results

To evaluate the performance of the suggested method under PSCs, the PV system includes three panels, a DC-DC boost converter and load as show in figure 4. A PSO a population of three particles is closed . To obtained the current reference I_{ref}* .The Boost converter is controlled by MPC .

The main parameters of the module used are shown in Table I. The PV panel includes 2 modules connected in parallel with other 2 en series. The characteristics of the PV array for different uniform irradiance levels are illustrated in Figures 4.

Table I. Solar panel specifications

Electrical parameters of the PV system	Value
Maximum power (Pmpp)	110 W
Open circuit voltage (Voc)	43.5V
Short circuit current (Isc)	3.45 A
PV output voltage at MPP	35 V
PV output current at MPP	3.15 A
Number of cells connected in series (Ns)	72
Electrical parameters of Boost converter	
Inductance (L)	45 mH
Capacitance (C)	1110 µF
Load resistive (R)	80 Ω

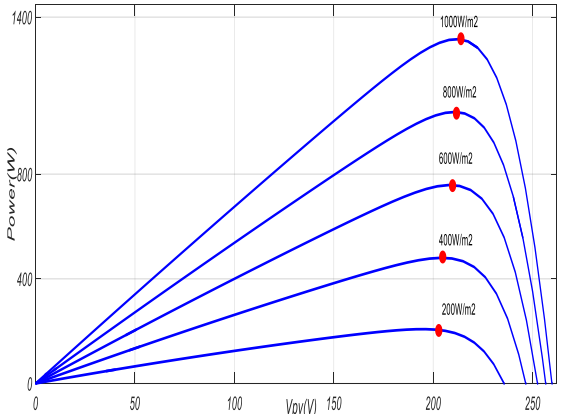


Figure 5. The P-V characteristics curve of PV generators under uniform shading

To examined the performance of proposed technique in tracking GMPP. Three different patterns are introduced in Table 2. Each pattern produces three peaks as show in figure 6 .

TABLE 2 SHADING PATTERNS TAKEN FOR THIS STUDY

Case	Ambient Conditions			Global Peak Pmax (W)
	PV1	PV2	PV3	
Case I	1000 W/m2	800 W/m2	600 W/m2	846W
	25°C	25°C	25°C	
Case II	800 W/m2	600 W/m2	200 W/m2	538.4W
	25°C	25°C	25°C	
Case III	700 W/m2	400 W/m2	100 W/m2	349W
	25°C	25°C	25°C	

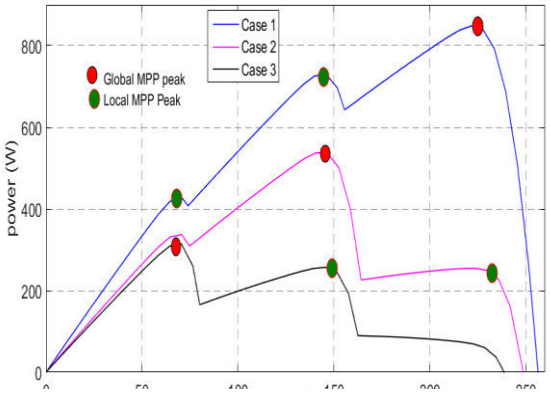


Figure 6. The P-V characteristics curve of PV generators under PSC.

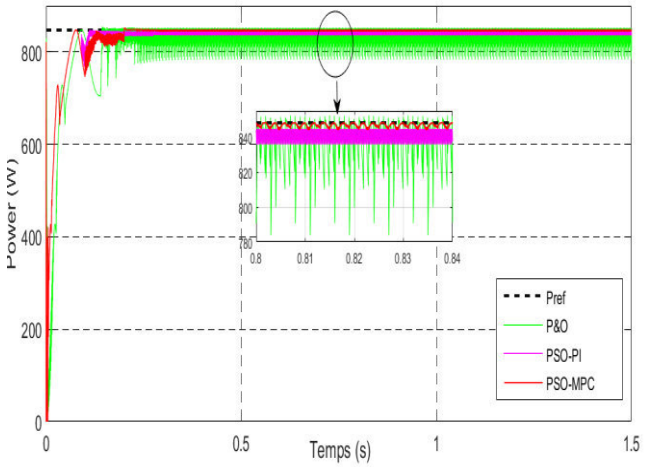


Figure 7. PV output Power under PSC1

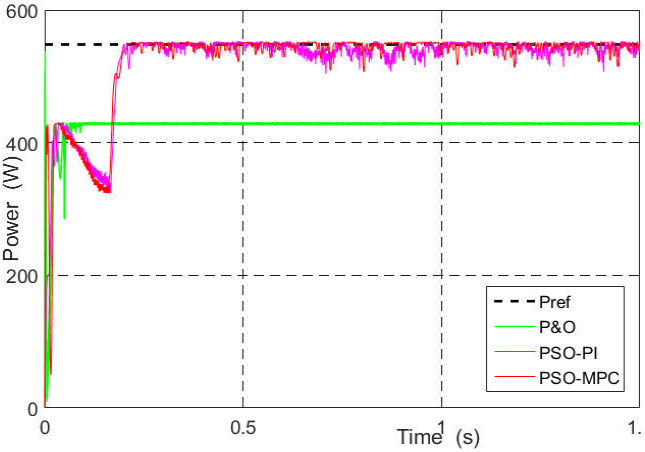


Figure 8. PV output Power under PSC2

The results of the tracking power outputs are depicted in figure 7and figure 8.

Figure 7 and Figure 8 show the proposed output power for the four different MPPT techniques (Perturb and Observe (P&O), incremental conductance (IncCond) , hybrid IncCond-PSO and IncCond-PSO-MPC . The power tracking GMMP by PSO-MPC better than the conventional P&O algorithm .

V. Conclusion

The principal objective of this paper is to present the MPPT method based PSO algorithm and predictive control for extraction of GMPP in PV system. The performance of the proposed technic is investigated under partially shaded conditions. The simulation results showed that the PSO algorithm combined with MPC is able to tracking the Global Maximum Power Point (GMPP).

REFERENCES

[1] J. Shi, W. Zhang, Y. Zhang, F. Xue, T. Yang "MPPT for PV systems based on a dormant PSO algorithm'' Electric Power Systems Research 123 (2015) 100–107

[2] A. F. Mirza,, Q.Ling , M. Y.Javed, M. Mansoor "Novel MPPT techniques for photovoltaic systems under uniform irradiance and Partial shading" Solar Energy 184 (2019) 628–648

[3] V. Salas, E. Olıas, A. Barrado, A. Lazaro "Review of the maximum power point tracking algorithms for stand-alone photovoltaic systems'' Solar Energy Materials & Solar Cells 90 (2006) 1555–1578.

[4] U. Yilmaz, Ali Kircay, S. Borekci ''PV system fuzzy logic MPPT method and PI control as a charge controller'' Renewable and Sustainable Energy Reviews 81 (2018) 994–1001.

[5] Y.H. Liu , Ch.-L. Liu a , J-W. Huang , J-H.Chen "Neural-network-based maximum power point tracking methods for photovoltaic systems operating under fast changing environments'' Solar Energy 89 (2013) 42–53. Elsevier.

[6] S. Veerapen , H. Wen,, X. Li, Y. Du, Y.Yang, Y.Wang, W. Xiao " A novel global maximum power point tracking algorithm for photovoltaic system with variable perturbation frequency and zero oscillation" Solar Energy 181 (2019) 345–356.

[7] Rihab Mahjoub Essefi , Mansour Souissi, Hsan Hadj Abdallah ''Maximum Power Point Tracking Control Using Neural Networks for Stand-Alone Photovoltaic Systems'' International Journal of Modern Nonlinear Theory and Application, 2014.

[8] U. Yilmaz, Ali Kircay, S. Borekci ''PV system fuzzy logic MPPT method and PI control as a charge controller'' Renewable and Sustainable Energy Reviews 81 (2018) 994–1001

[9] E. Lodhi, R. N .Shafqat, and Kerrouche K. D. E "Application of Particle Swarm Optimization for Extracting Global Maximum Power Point in PV System under Partial Shadow Conditions " International Journal of Electronics and Electrical Engineering Vol. 5, No. 3, June 2017

[10] A. Guichi, A. Talha, E.M. Berkouk, S. Mekhilef, S. Gassab" A new method for intermediate power point tracking for PV generator under partially shaded conditions in hybrid system" Solar Energy 170 (2018) 974–987. Elsevier.

[11] N. Kacimi, S.Grouni ,A. Idir and M.S Boucherit " New improved hybrid MPPT based on neural network- model predictive control- kalman-filter for photovoltaic system". Indonesian journal of electrical engineering and Computer Science journal Vol 20 No 3.2020

[12] Nora Kacimi, Said Grouni, Abdelhakim Idir , *M*ohamed Seghir Boucherit "A New combined method For tracking the global maximum power point of photovoltaic systems " *Rev. Roum. Sci. Techn.– Électrotechn. et Énerg. Vol. 67, No. 3, October 2022.*

[13] S. Z. Said, L.Thiaw "Performance of Artificial Neural Network and Particle Swarm Optimization Technique based Maximum Power Point Tracking for Photovoltaic System Under DifferentEnvironmental Conditions " IOP Conf. Series: Journal of Physics: Conf. Series 1049 (2018) 012047. doi : 10.1088/1742-6596/1049/1/012047

[14] S.Z. Said, L. Thiaw, C.W.Wabuge "Maximum Power Point Tracking of Photovoltaic Generators Partially Shaded Using a Hybrid Artificial Neural Network and Particle Swarm Optimization Algorithm" International Journal of Energy and Power Engineering 2017, 6(6): 91-99.

[15] B. Boukezata, J.-P. Gaubert, A. Chaoui, M. Hachemi ''Predictive current control in multifunctional grid connected inverter interfaced by PV system'' Solar Energy 139 (2016) 130–141

[16] A. Dehghanzadeh, G.Farahani, H. Vahedi, K. Al-Haddad''Model predictive control design for DC-DC converters applied to a photovoltaic system'' Electrical Power and Energy Systems 103(2018) 537-544

[17] M. Kermadi ,El M. Berkouk ''Artificial intelligence-based maximum power point tracking controllers for Photovoltaic systems: Comparative study" Renewable and Sustainable Energy Reviews 69 (2017) 369–386

[18] A. Guichi, A. Talha,, E.M. Berkouk, S. Mekhilef, S. Gassab '' A new method for intermediate power point tracking for PV generator under partially shaded conditions in hybrid system''Solar Energy 170 (2018) 974–987

[19] L. Piegari R. Rizzo '' Adaptive perturb and observe algorithm for photovoltaic maximumpower point tracking'' Published in IET Renewable Power Generation, 2010, Vol. 4, Iss. 4, pp. 317– 328.

[20] B.Talbi, F.Krim, T.Rekioua, A. Laib, and H. Feroura "Design and hardware validation of modified P&O algorithm by fuzzy logic approach based on model predictive control for MPPT of PV systems" Journal of Renewable and Sustainable Energy 9, 043503 (2017); doi: 10.1063/1.4999961

[21] M. Metry, Mohammad B. Shadmand, R.S. Balog, and H. Abu Rub '' High Efficiency MPPT by Model Predictive Control Considering Load Disturbances for Photovoltaic Applications Under Dynamic Weather Condition'' IECON2015-Yokohama 978-1-4799-1762-4/15 2015 IEEE.

[22] S. Babu, J. P. Ram,T. Dragicevic,M. Miyatake,F. Blaabjerg , and N. Rajasekar" Particle Swarm Optimization Based Solar PV Array Reconfiguration of the Maximum Power Extraction Under Partial Shading Conditions'' IEEE Transactions on suitable energy .Vol 9.No 1 , January 2018.

[23] N Kacimi , Abdelhakim IDIR , Said Grouni, Mohamed Seghir Boucherit "An Improved MPPT Control Strategy for PV connected to grid using IncCond-PSO-MPC Approach" *Csee Journal Of Power And Energy Systems, (2022)*

Modelling and Simulation of the Interleaved Boost Converter

*

Okba Mohammed Cherif
(LREEI),FHC (UMBB)
University M'hamed BOUGARA
Boumerdes,Algeria
o.mohammedcherif@univ-boumerdes.dz

Ziyad Younsi
LEREESI,Higher National School of Renewable Energies
Batna,Algeria
z.younsi@hns-re2sd.dz

Belferdi Souad
LEREESI,Higher National School of Renewable Energies
Batna,Algeria
s.belferdi@hns-re2sd.dz

Nadji Bouchra
(LREEI),FHC (UMBB)
University M'hamed BOUGARA
Boumerdes,Algeria
b.nadji@univ-boumerdes.dz

Sid Ahmed Tadjer
(LREEI),FHC (UMBB)
University M'hamed BOUGARA
Boumerdes,Algeria
s.tadjer@univ-boumerdes.dz

Bencherif Hichem
LEREESI,Higher National School of Renewable Energies
Batna,Algeria
bencherif.hichem@hns-re2sd.dz

***Abstract*—This paper presents a detailed analysis of the interleaved boost converter (IBC), focusing on its operational principles, mathematical modelling, and simulation results. The IBC is characterized by its ability to efficiently convert low input voltages to higher output voltages while minimizing current ripple through interleaving techniques. We explore various operational modes and derive averaged and small-signal models that describe the converter's behaviour under different load conditions. Simulation results demonstrate the effectiveness of the IBC in achieving high voltage gain with reduced output ripple, making it suitable for applications in renewable energy systems and electric vehicles. The findings underscore the importance of advanced modelling approaches in optimizing IBC designs for improved performance in modern power electronics.**

***Index Terms*—IBC, Modelling, Small Signal, Simulation**

I. Introduction

The Interleaved Boost Converter (IBC) has emerged as a pivotal technology in the field of power electronics, particularly for applications requiring high voltage gain and efficiency. With the increasing reliance on renewable energy sources and electric vehicles [1], the demand for efficient power conversion systems has become more critical than ever. The IBC offers a robust solution by utilizing multiple phases, which enhance performance metrics such as reduced output voltage ripple, improved transient response, and overall efficiency [2], [3]. The operational principle of the IBC involves interleaving two or more boost converter phases, resulting in a significant reduction in both input and output current ripple. This interleaving technique not only improves the converter's performance but also facilitates better thermal management by distributing the thermal load across multiple components. Furthermore, the IBC's design flexibility allows it to be adapted for a wide range of applications, from low-power consumer electronics to high-power industrial systems. Recent advancements in modelling and simulation techniques have provided a deeper understanding of the IBC's dynamics. Engineers can now optimize designs using precise mathematical models that account for various operational conditions. For example, studies have shown that interleaved configurations can achieve efficiencies exceeding 98%, significantly reducing current ripple compared to traditional boost converters [4], [5]. Additionally, advanced control strategies have been shown to further enhance the performance of IBCs under varying load conditions [6]. This paper aims to explore these modelling techniques, with a focus on averaged and small-signal models that describe the converter's behaviour under different load scenarios. By providing insights into the IBC's operational characteristics and control strategies, this study contributes to the ongoing development of efficient power conversion technologies that are essential for modern energy systems.

II. Operation Principle of the (IBC)

Figure.1 shows the equivalent circuit of the Interleaved Boost Converter (IBC), where the switches Q1 and Q2 are driven by control signals that are shifted by 180° relative to each other. The duty cycle is set to less than 0.5.

III. Modes of Operation

A. Mode 1 [0, Td]:

In this period Q_1, D_2 are ON and Q_2, D_1 are OFF (Figure.2)

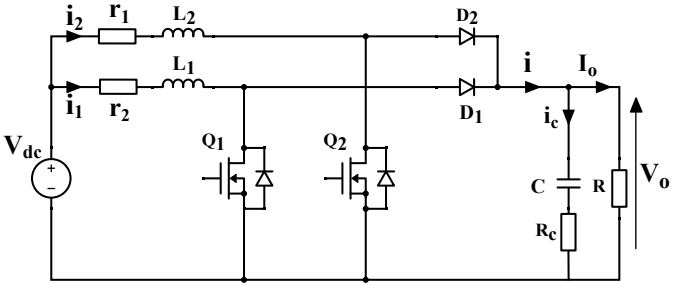


Fig. 1. Interleaved Boost converter IBC.

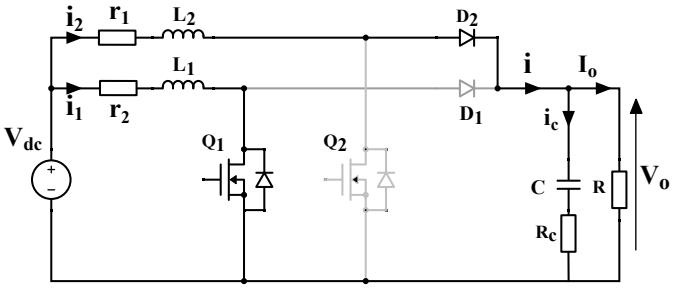


Fig. 2. Circuit model of the Mode 1.

The state-space dynamics of the first mode is given by:

$$\begin{pmatrix} L\frac{di_1}{dt} \\ L\frac{di_2}{dt} \\ C\frac{dv_c}{dt} \end{pmatrix} = \begin{pmatrix} -(r+R_{ds}) & 0 & 0 \\ 0 & -(r+R_d+\frac{RR_c}{R+R_c}) & \frac{-R}{R+R_c} \\ 0 & \frac{R}{R+R_c} & \frac{-1}{R+R_c} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ v_c \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_{dc} \\ \\ v_d \end{pmatrix} \tag{1}$$

$$v_o = \begin{pmatrix} 0 & \frac{RR_c}{R+R_c} & \frac{R}{R+R_c} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ v_c \end{pmatrix} \tag{2}$$

B. Mode 2 [Td, T/2]:

Switches Q_1, Q_2 are OFF and Diode D_1, D_2 are ON. The circuit model of this period is shown in Figure.3

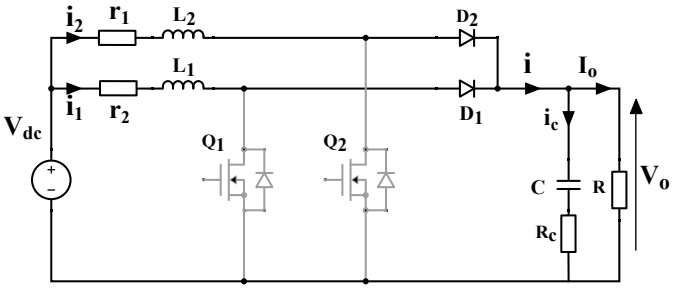


Fig. 3. Circuit model of the Mode 2.

The state-space model of this stage is given by :

$$\begin{pmatrix} L\frac{di_1}{dt} \\ L\frac{di_2}{dt} \\ C\frac{dv_c}{dt} \end{pmatrix} = \begin{pmatrix} -(r+R_d+\frac{RR_c}{R+R_c}) & \frac{RR_c}{R+R_c} & \frac{-R}{R+R_c} \\ \frac{-RR_c}{R+R_c} & -(r+R_d+\frac{RR_c}{R+R_c}) & \frac{-R}{R+R_c} \\ \frac{R}{R+R_c} & \frac{R}{R+R_c} & \frac{-1}{R+R_c} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ v_c \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_{dc} \\ \\ v_d \end{pmatrix} \tag{3}$$

The output voltage is expressed as:

$$v_o = \begin{pmatrix} 0 & \frac{RR_c}{R+R_c} & \frac{R}{R+R_c} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ v_c \end{pmatrix} \tag{4}$$

C. Mode 3 [T/2, T(0.5+d)]:

During this period Q_1, D_2 are OFF, and Q_2, D_1 are ON (see Figure.4).

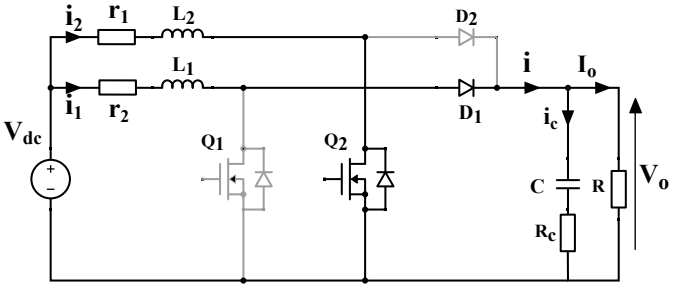


Fig. 4. Circuit model of the Mode 3.

The equations that describe this period are given by:

$$\begin{pmatrix} L\frac{di_1}{dt} \\ L\frac{di_2}{dt} \\ C\frac{dv_c}{dt} \end{pmatrix} = \begin{pmatrix} -(r+R_d+\frac{RR_c}{R+R_c}) & 0 & \frac{-R}{R+R_c} \\ 0 & -(r+R_{ds}) & 0 \\ \frac{R}{R+R_c} & 0 & \frac{-1}{R+R_c} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ v_c \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_{dc} \\ \\ v_d \end{pmatrix} \tag{5}$$

$$v_o = \begin{pmatrix} \frac{RR_c}{R+R_c} & 0 & \frac{R}{R+R_c} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ v_c \end{pmatrix} \tag{6}$$

D. Mode 4 [T(d+0.5), T]:

In this stage theQ_1, Q_2Are OFF and D_1, D_2 are ON, the circuit model of mode 4 is shown in Figure.5

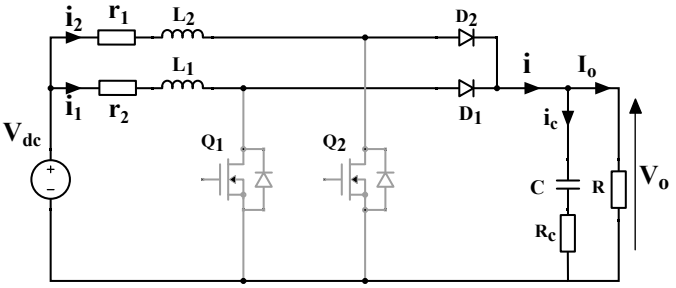


Fig. 5. Circuit model of the Mode 4.

The dynamic behavior of the system during this stage is modeled as:

$$\begin{pmatrix} L\frac{di_1}{dt} \\ L\frac{di_2}{dt} \\ C\frac{dv_c}{dt} \end{pmatrix} = \begin{pmatrix} -(r+R_d+\frac{RR_c}{R+R_c}) & \frac{RR_c}{R+R_c} & \frac{-R}{R+R_c} \\ \frac{-RR_c}{R+R_c} & -(r+R_d+\frac{RR_c}{R+R_c}) & \frac{-R}{R+R_c} \\ \frac{R}{R+R_c} & \frac{R}{R+R_c} & \frac{-1}{R+R_c} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ v_c \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_{dc} \\ \\ v_d \end{pmatrix} \tag{7}$$

The relationship defining the output voltage is expressed as:

$$v_o = \begin{pmatrix} \frac{RR_c}{R+R_c} & \frac{RR_c}{R+R_c} & \frac{R}{R+R_c} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ v_c \end{pmatrix} \tag{8}$$

IV. AVERAGED MODEL

This section focuses on developing the averaged large-signal model of the system. The purpose of this model is to simplify the analysis by averaging the dynamic behavior over one switching period, enabling a clear understanding of the system's overall performance The equations (1)-(8) are in the form:

$$\begin{cases} \dot{x} = A_n x + B_n u \\ y = C_n x \end{cases} \tag{9}$$

where A_n, B_n, and C_n are the matrices of the corresponding operation mode $n = 1, 2, 3, 4$.

To obtain the averaged model, we need to average the matrices A_n, B_n, and C_n, which can be calculated as follows:

$$\begin{cases} A = A_1 \times d + A_2 \times (0.5 - d) + A_3 \times d + A_4 \times (0.5 - d) \\ B = B_1 \times d + B_2 \times (0.5 - d) + B_3 \times d + B_4 \times (0.5 - d) \\ C = C_1 \times d + C_2 \times (0.5 - d) + C_3 \times d + C_4 \times (0.5 - d) \end{cases} \tag{10}$$

Therefore, we obtain the average Model of the interleaved Boost Converter as follows:

$$\begin{pmatrix} L\frac{di_1}{dt} \\ L\frac{di_2}{dt} \\ C\frac{dv_c}{dt} \end{pmatrix} = \begin{pmatrix} -(r + dR_{ds} + d'R_d + \frac{dd'RR_c}{R+R_c}) & \frac{d^2RR_c}{R+R_c} & -d' \\ \frac{d^2RR_c}{R+R_c} & -(r + dR_{ds} + d'R_d + \frac{dd'RR_c}{R+R_c}) & -d' \\ d' & -d' & \frac{-1}{R} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ v_o \end{pmatrix} + \begin{pmatrix} 1 & -d' \\ 1 & -d' \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_{dc} \\ \\ v_d \end{pmatrix} \tag{11}$$

By taking: $i = i_1 + i_2$, we can reduce the equation 11 to:

$$\begin{pmatrix} L\frac{di}{dt} \\ C\frac{dv_c}{dt} \end{pmatrix} = \begin{pmatrix} -(r + dR_{ds} + d'R_d + \frac{d(1-2d))RR_c}{R+R_c}) & -2d' \\ d' & \frac{-1}{R} \end{pmatrix} \begin{pmatrix} i \\ v_o \end{pmatrix} + \begin{pmatrix} 2 & -2d' \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_{dc} \\ \\ v_d \end{pmatrix} \tag{12}$$

From the equation (12) we can easily construct the equivalent steady state average large signal circuit model of the interleaved boost converter as shown in the Figure.6, by taking $\frac{dx}{dt}$ at steady state. From this circuit model the voltage gain and the efficiency can be easily calculated.

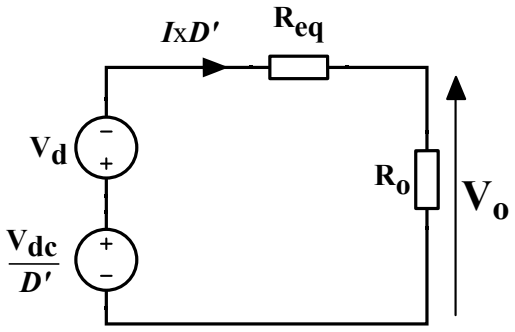


Fig. 6. Averaged model equivalent circuit model.

V. SMALL SIGNAL MODEL

In this section, we develop the small-signal model of the Interleaved Boost Converter (IBC) to derive its various transfer functions. Each variable is expressed as the sum of two components: the steady-state DC value and the small-signal perturbation, as illustrated in equation (13).

$$\begin{cases} i &= \hat{i} + I \\ v_{dc} &= \hat{v}_{dc} + V_{dc} \\ v_o &= \hat{v}_o + V_o \\ d &= \hat{d} + D \\ v_d &= \hat{v}_d + V_d \end{cases} \tag{13}$$

Substituting equation (13) into equation (12) yields the small-signal model, as expressed in equation (14)

$$\begin{pmatrix} L\frac{di}{dt} \\ C\frac{dv_c}{dt} \end{pmatrix} = \begin{pmatrix} -(r + DR_{ds} + DR_d + \frac{D(1-2D))RR_c}{R+R_c}) & -2D' \\ D' & \frac{-1}{R} \end{pmatrix} \begin{pmatrix} i \\ v_o \end{pmatrix} + \begin{pmatrix} 2 & -2d' \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_{dc} \\ \\ v_d \end{pmatrix} \tag{14}$$

The small-signal circuit model is derived from equation (14) and is illustrated in Figure.7

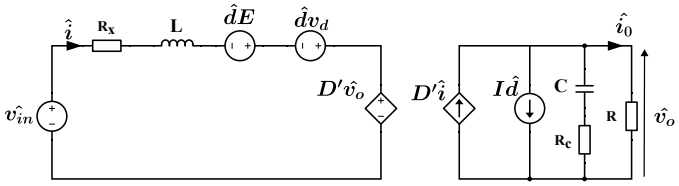


Fig. 7. Small signal circuit model

Where:

$$\begin{cases} R_x = \frac{1}{2}\left(r + DR_{ds} + D'R_d + \frac{D(1-2D)RR_c}{R+R_c}\right), \\ R_{eq} = \frac{R_x}{D'^2}, \\ E = \frac{1}{2}\left(-R_{ds} + R_d + \frac{(1+2D)RR_c + 2D'R^2}{R+R_c} + 2V_d\right) \end{cases}$$

VI. SIMULATION RESULTS

The table below presents the values of the various components used in the Interleaved Boost Converter (IBC)

TABLE I
COMPONENT VALUES OF THE INTERLEAVED BOOST CONVERTER (IBC)

Parameter	Value	Parameter	Value
Vdc	180V	R	20Ω
Vo	300V	Rc	0.03Ω
Duty cycle d	0.4	r	0.3Ω
Inductance L	0.125mH	C	120 μF

Figure.8 shows the output voltage of the Interleaved Boost Converter (IBC) simulated in LTspice software. Figure.9 illustrates the currents in inductors L_1 and L_2, while Figure.8 shows the output current.

In Figure.10, which shows the combined inductor currents $I_1 + I_2$, it is evident that the inductor current ripple is significantly reduced in the Interleaved Boost Converter (IBC) compared to conventional boost converters. This reduction in ripple is one of the primary advantages of the IBC topology, and it plays an important role in enhancing overall system

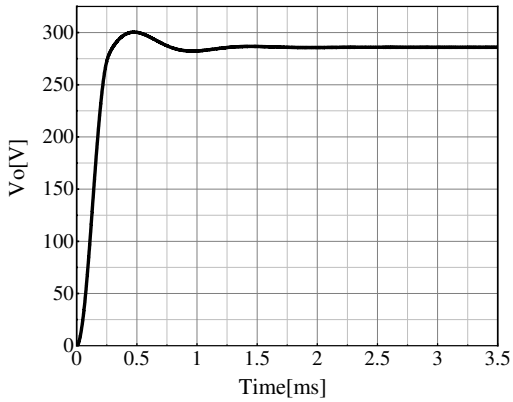


Fig. 8. Caption

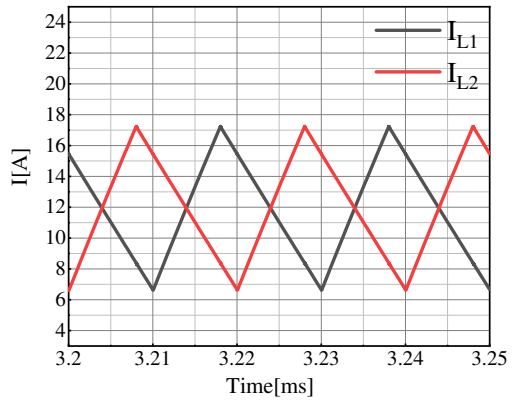


Fig. 9. Inductors currents

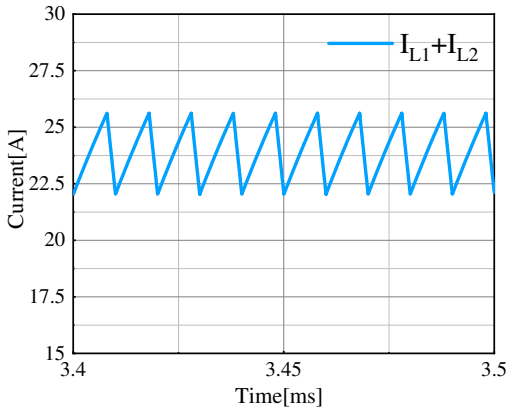


Fig. 10. Total inductor current

performance. The key benefit of reducing the current ripple is the reduction of electromagnetic interference (EMI) and the associated stresses on the components, especially the inductors and switches. This ultimately reduces the need for oversized components, thus enhancing the thermal performance and reliability of the converter. The efficiency improvement, along with reduced stress on components, allows the IBC to operate at higher power levels with better performance and longer lifetimes compared to traditional boost converters.

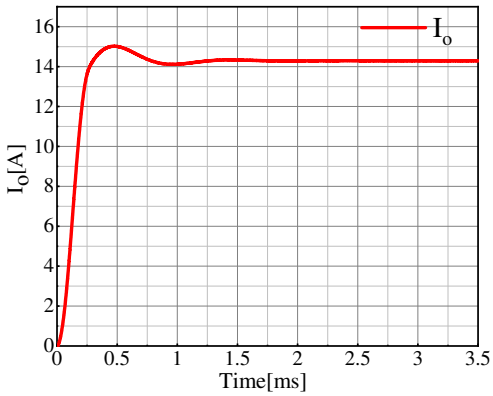


Fig. 11. Output current

A. *Transfer function and bode diagram*

To control the Interleaved Boost Converter (IBC), the control-to-output transfer function is derived using the small-signal model. This transfer function establishes the relationship between the duty cycle and the output voltage. Equation.15 presents the transfer function of the IBC, where the input variable is the duty cycle (d), and the output is the output voltage (V_o). Only the duty cycle-to-output voltage transfer function, G_{Vd}, is calculated. The Bode plot of the transfer function G_{Vd} is shown in Figure.12.

$$G_{vd} = \frac{-0.7362s^2 + 8.454 \times 10^4 s + 2.382 \times 10^{10}}{s^2 1779s + 4.842 \times 10^7} \quad (15)$$

The same approach can be applied to obtain the desired transfer function for other relationships, such as the duty cycle to output current or the input voltage to output voltage.

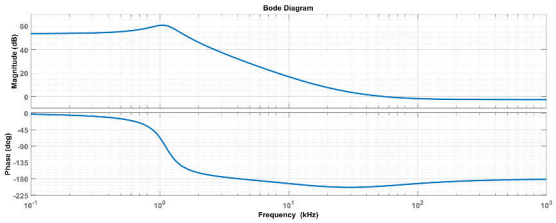


Fig. 12. Bode diagram of the G_{vd} transfer function

VII. CONCLUSION

In this paper, a detailed analysis of the different modes of operation of the Interleaved Boost Converter (IBC) was presented. The averaged large-signal model was developed, enabling the extraction of voltage gain and efficiency while considering various parasitic variables. The small-signal model was also derived, providing valuable insights into the converter's dynamic behavior. The developed model is simple and allows the IBC to be analyzed similarly to a traditional boost converter, simplifying its design and optimization for high-performance power conversion applications. Furthermore, compared to the traditional Boost Converter, the IBC

offers several advantages, including improved efficiency due to reduced ripple currents, enhanced thermal performance through current sharing, and reduced stress on individual components, leading to greater reliability and longer lifetimes. These benefits make the IBC particularly suitable for high-power and high-performance applications where efficiency, thermal management, and reliability are critical.

REFERENCES

[1] G. G. Ramanathan and N. Urasaki, "Non-isolated interleaved hybrid boost converter for renewable energy applications," *Energies*, vol. 15, no. 2, p. 610, 2022.

[2] S. Kamtip and K. Bhumkittipich, "Design and analysis of interleaved boost converter for renewable energy applications," in *Proceedings of the [Conference Name]*. Rajamangala University of Technology Thanyaburi, 2011.

[3] G. Ramya, V. Ganapathy, and P. Suresh, "Comprehensive analysis of interleaved boost converter with simplified h-bridge multilevel inverter based static synchronous compensator system," *Electric Power Systems Research*, vol. 176, p. 105936, 2019.

[4] K. L. Shenoy, C. G. Nayak, R. P. Mandi *et al.*, "Design and implementation of interleaved boost converter," *International Journal of Engineering and Technology (IJET)*, vol. 9, no. 3S, pp. 496–502, 2017.

[5] M. Hosseinpour, E. Seifi, A. Seifi, and M. Shahparasti, "Design and analysis of an interleaved step-up dc–dc converter with enhanced characteristics," *Scientific Reports*, vol. 14, no. 1, p. 14413, 2024.

[6] R. A. Aprilianto, E. Firmansyah, and F. D. Wijaya, "Review on control strategy for improving the interleaved converter performances," in *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. IEEE, 2021, pp. 312–317.

Mathematical Model in Simulink/Matlab of a Battery Electric Vehicle

Said BELHADJ
Department of Electrical Systems Engineering
M'Hamed BOUGARA University
BOUMERDES, ALGERIA
s.belhadj@univ-boumerdes.dz

***Abstract*—Electric vehicles are attracting growing interest among researchers, becoming promising alternatives for achieving more sustainable and cleaner energy emissions in the transportation sector. They have demonstrated a significant capacity to reduce the consumption of fossil fuels and other high CO2-emitting combustibles. This paper presents a simulation model of an electric vehicle on the Matlab-Simulink platform, based on mathematical equations, to evaluate its operation during energy regeneration phases by analyzing the energy flow. The components of a battery electric vehicle (BEV) system were examined, and the relevant electrical system components along with their corresponding equations were identified for validation. We developed this mathematical model of the longitudinal dynamics of an electric vehicle, incorporating all necessary components to generate the forces required to propel the vehicle while accounting for the resistive forces acting on it. The study focuses on analyzing each component to understand the longitudinal dynamics of the vehicle under various conditions. The simulation results are presented and discussed.**

Keywords—Mathematical Modeling, Simulation, Battery Electric Vehicle, Matlab-Simulink.

I. INTRODUCTION

Today, the global automotive industry is undergoing a restructuring process, with new technologies being developed. Many companies are shifting toward strategies for producing electric vehicles. To standardize the emissions of harmful substances from cars, major automakers are implementing new strategies to reinforce these standards. Numerous automotive companies and leaders in the IT industry view electric vehicles as one of the most promising areas for commercial development and are considering strengthening their technological capabilities to ensure future market competitiveness [1].

Energy conservation is considered one of the greatest global challenges, with many hurdles to overcome in the energy sector. While the future of energy remains uncertain, transportation is expected to play a critical role in energy conservation in the coming years [2]. The automotive industry is rapidly transitioning from internal combustion engine (ICE) vehicles to electric vehicles (EVs) as the primary mode of transportation.

This shift is driven by urban air pollution and the fact that internal combustion engines are the second-largest contributor to global warming, accounting for approximately 21% of greenhouse gas emissions [3]. Electric vehicles (EVs) contribute to making our lives easier and safer. EVs can produce, store, and transport electricity. Moreover, they are more economical and environmentally friendly compared to conventional cars [4]. The depletion of fossil fuels and the rise in their prices have significantly increased interest in electric vehicles (EVs) [5]. This is due to the advantages of EVs as clean and quiet technology with greater efficiency compared to internal combustion engine (ICE) vehicles [6]. Numerous simulation and modeling software tools have been developed to study the operation of electric and hybrid propulsion systems [7]. In their study, the authors present a set of modeling, simulation, and analysis tools for the drivetrain of four types of vehicles: an electric vehicle (EV), a parallel hybrid electric vehicle (HEV), a series HEV, and an internal combustion engine vehicle (ICE), using Matlab-Simulink [8-13].

The model developed by Wang and Zhang [8] demonstrates simulation results showing that the electric vehicle exhibits good output characteristics. A study by Raza et al. developed ADVISOR software to build a model for evaluating the performance of the Prius, validated through simulation tests and compared with pure fuel vehicles [9]. The results revealed that under both CYC_UDDS and CYC_HWFET operating conditions, the Prius outperforms pure fuel vehicles in terms of power and fuel economy [10].

Rani et al. highlighted the essential role of virtual platforms in addressing technological advances in the electric vehicle industry [11]. A review by Mosleuzzaman provides insights into how innovations in electrical engineering are shaping the future of EV powertrain design and sustainable transportation [12]. Xu et al. introduced a quasi-static simulation-based approach for estimating EV energy consumption under various road vehicle operating conditions [13].

Lipu et al. conducted an analysis and evaluation of the role of AI-based approaches in enhancing the battery management system (BMS) of EVs. Cutting-edge AI approaches are critically reviewed in terms of their objectives, contributions, advantages, and limitations. Furthermore, several significant issues and challenges, as well as crucial guidelines and recommendations, are provided for potential future developments. This statistical analysis can guide future researchers in the development of emerging BMS technologies for the sustainable management and operation of EVs [14]. In the following section, MATLAB-Simulink was used to design the components of the battery electric vehicle (BEV) and integrate the entire system. Additionally, the complete model was simulated, and its equations were validated. Moreover, the simulation results will be presented and discussed, while the final section will conclude the study.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

II. MODELING METHOD

We used a drivetrain as illustrated in Figure 1. The components used in the electric vehicle (EV) model are: the electric motor, powered by an inverter, which is in turn supplied via a battery connected to a DC-DC converter, and a vehicle interface. The latter ensures communication between the commands of the motor controller and those of the battery controller. The regulation and management of the power supplied to the motor and the energy coming from the battery, are handled by the motor and battery controllers. The battery constitutes the energy storage system. Voltage, current, and frequency are regulated by the controllers according to the motor's demands in the four quadrants of operation.

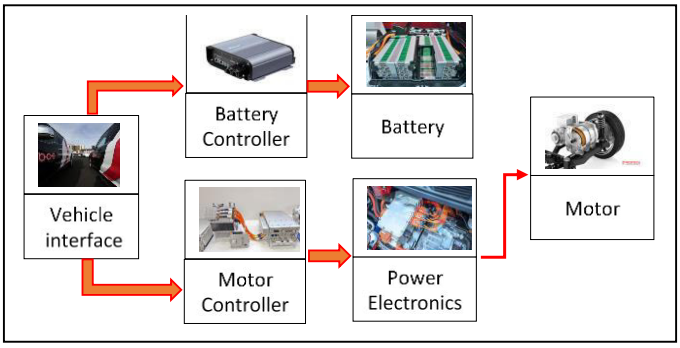


Fig. 1. Drivetrain of an Electric Vehicle (EV).

Figure 4 shows the operating modes of the motor according to the motor speed and the applied torque, following a coordinate axis system. The drivetrain operates in motor mode when it is in the first and third quadrants, and in generator mode when it is in the second and fourth quadrants.

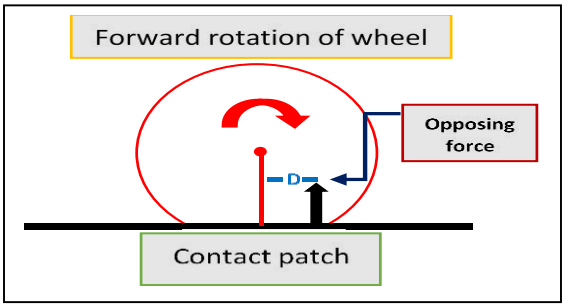


Fig. 2. Cross-sectional view of the forces acting on the wheel

A. *Electric Motor*

In electric vehicles, four different types of electric motors have been used so far: the DC motor, the asynchronous motor, the permanent magnet synchronous motor, and the variable reluctance motor. After evaluation, the permanent magnet synchronous motor is considered the most suitable for the model.

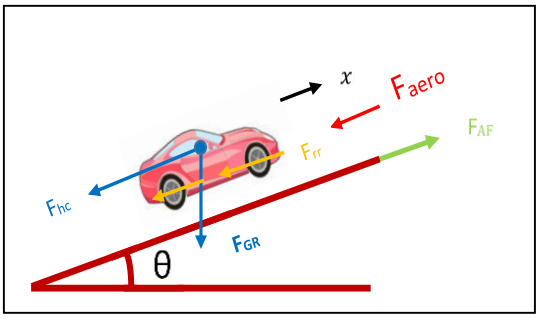


Fig. 3. Longitudinal forces acting on a vehicle

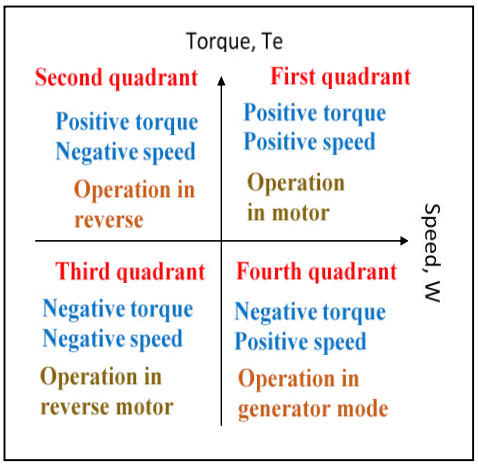


Fig. 4. Four-quadrant operation of the drivetrain.

B. Vehicle resistance forces

During driving, resistance forces act on vehicles. These resistance forces include: aerodynamic resistance, tire rolling resistance, slope resistance, and inertia resistance.

1) *Rolling resistance:*

This is a force that opposes the movement of an object. It is the force acting on the vehicle due to the properties of tires, known as hysteresis. Because of hysteresis, the contact area of the tire at the leading edge is greater than at the trailing edge (the contact is smaller at the trailing edge due to the slow retraction of the tire to its unloaded state). This creates a moment opposite to the direction of motion. This opposing moment is defined as rolling resistance. The rolling resistance coefficient of tires (У) varies depending on the road surface.

$$F_{rr} = У . N \quad (01)$$

У: Rolling resistance coefficient
N: Vehicle weight.

2) *Drag/Aerodynamic resistance:*

Atmospheric air acts as an obstacle to the vehicle's movement, generating a resistance force called aerodynamic drag. This force increases significantly at high speeds (> 40 km/h). The factors influencing this force are the aerodynamic coefficient and the frontal area. Profiles are designed for four-wheeled vehicles to reduce energy losses, as air disrupts movement by creating turbulence.

$$F_{aero} = \frac{1}{2}.\rho.A.V^2.C_d \quad (02)$$

ρ: Air density [Kg/m^2]
A: Frontal area of the vehicle [m^2]
V : Vehicle speed [m/s]
C_d : Drag coefficient.

The aerodynamic resistance coefficient is determined experimentally based on the shape of the vehicle.

TABLE I. VEHICLE TYPE AND CD [15]

Vehicle Type	CD
Cabriole	0.5 – 0.7
Car	0.3 – 0.4
Bus	0.6 – 0.7
Truck	0.8 – 1.5
Optimal Design	0.2 – 0.3

3) Resistance due to the slope:

The resistance encountered during an uphill climb due to gravity.

$$F_{GR} = m.g.sin\theta \quad (03)$$

m: Mass of vehicle [Kg]
g: Acceleration of gravity [m.s^{-2}]
θ : Inclination angle [Degree]

TABLE II. ROAD SURFACE AND CD [16]

Road Surface	CR
Concrete or Asphalt	0.013
Small Gravel Ground	0.02
Macadamized Road	0.025
Soil Road	0.1-0.35

4) Inertia resistance

During motion, an inertia force is generated in the opposite direction, according to Newton's second law, particularly during acceleration or deceleration phases. This resistance is called inertia resistance, and it is given by:

$$F_{AF} = m.a \quad (04)$$

m: Mass of vehicle [Kg]
a: Acceleration of the vehicle [m.s^{-2}]

$$a = \frac{V_f - V_i}{t} \quad (05)$$

V_f: Final Velocity [m/s]
V_i: Initial velocity [m/s]
t: time [s]

5) Calculation of top speed

The maximum speed of the vehicle is determined by the engine speed and the gear reduction used. For a specific tire size, the wheel speed (rpm) can be calculated using:

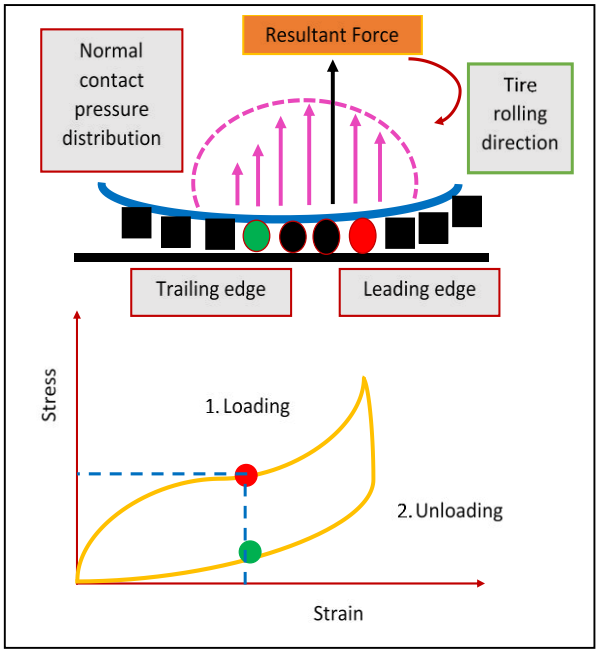


Fig. 5. (Above) lateral view of the tire-road contact pressure distribution in the vertical direction and (below) stress-strain curve of the rubber compound. The tread block in the front half (red) is under load. The tread block in the rear half (green) is unloaded.

$$V = \frac{\pi DN}{60} \quad (06)$$

$$N_{wheel} = \frac{60V}{\pi D} \quad (07)$$

V: Vehicle velocity [m/s]

D: Tire diameter [m]

The gear reduction ratio is necessary to reduce the engine speed, as the engine speed is always higher than the wheel speed. In general, it is the engine's top speed that is chosen. For an electric motor, it can range from 4,000 to 13,000 rpm (ideal case), while the wheel speed is around 900 to 1,100 rpm for 100 km/h. Therefore, the reduction ratio = Engine speed (RPM) / Wheel speed (RPM). We obtained a maximum speed of 96.8 km/h during the simulation in the NEDC cycle.

6) Calculation of peak power:

Power is a function of speed and force. Power is calculated in two steps: peak torque based on acceleration and top speed.

• For maximum torque based on acceleration, power = total traction force * final speed (during acceleration).
• For top speed, power = (rolling resistance force + drag resistance at top speed [m/s]).

The highest value is chosen as the peak power for the motor.

C. Longitudinal dynamic model of a vehicle

The interaction between the car, tires, road, and the driving system can be described within the framework of the concept of the longitudinal dynamic model of a vehicle. Three variables define the state of the dynamic model of the motion of an electric vehicle: the longitudinal wheel slip, the velocity at the vehicle's center of gravity, and the circumferential speed of the wheels. Identifying each of the forces acting in

the direction of a vehicle is essential for controlling its behavior in motion.

$$\lambda_x(t) = \frac{V_v(t) - V_k(t)}{V_v(t)} \quad (08)$$

Where $\lambda_x(t)$ represents the longitudinal slip of the driven wheel, and $V_v(t)$, $V_k(t)$ are the speed of the electric vehicle. To model an electric vehicle (EV), all the mathematical equations representing each component of the EV powertrain have been determined. The motor, battery, motor controller, and proportional-integral (P-I) controller have been modeled on the Matlab-Simulink platform as individual block diagrams to form an EV propulsion system using the equations that will be defined. For a permanent magnet synchronous motor (PMSM), the torque T_e developed in the PMSM is given by:

$$T_e = \frac{3}{2}\frac{P}{2}(\psi_r I_q + (L_d - L_q) I_d I_q) \quad (09)$$

Where P represents the number of poles of the machine. The electrical torque obtained in Eq. (09) is divided into two components: the "mutual reactance torque" and the "reluctance torque," the latter resulting from the difference in reluctance between the q-axis and the d-axis. However, for the PMSM, when $L_q = L_d = L_S$, the torque generated by the PMSM is given by:

$$T_e = \frac{3}{2}\frac{P}{2}(\psi_r I_q) \quad (10)$$

However, by using the dq reference frame, the PMSM model can be represented in the rotating dq frame as follows:

$$V_q = R_s I_q + \omega_r (L_d I_d + \lambda_r) + P L_q I_q \quad (11)$$

$$V_d = R_s I_d - \omega_r L_q I_q + P(L_d I_d + \lambda_r) \quad (12)$$

Assuming there are no losses due to friction or inertia, the developed electrical torque, T_e, is equal to the output mechanical torque, T_m. Therefore, the developed electrical power is equal to the developed mechanical power.

D. Battery charge controller model

The battery charge controller is important for extending its lifespan during operation. When modeling and designing a Battery Electric Vehicle (BEV), the battery management system (BMS) model used is taken into account, where the charge controller enhances the BMS's performance. The latter is evaluated based on the displayed information, such as voltage, current, temperature, state of charge (SoC), as well as the measurement of cell balancing within the battery. Each model contains a specific battery pack, depending on the type of battery used. The state of charge (SoC) is calculated using the following equation:

1) Battery pack

It is composed of M modules connected in parallel, with each module consisting of N cells connected in series.

$$SOD = \frac{1}{Q_T}\int_0^t \alpha\,[i(\tau)].\beta[T(\tau)]d_\tau \quad (13)$$

Where *SOD* is the state of discharge, t represents time, *T* is the temperature, and *i* is the current.

2) State of Charge (SoC)

In percentage, the SoC expresses the remaining capacity of the battery, which can be influenced by temperature, discharge rate, and battery lifespan. As shown in the equation below, SoC is the ratio of the available residual charge to the nominal capacity:

$$SOC = \frac{Q(t)}{Q_{nom}} \quad (14)$$

Where Q(t) is the available residual charge and Q_{nom} is the nominal capacity.

3) Regenerative braking

The regenerative braking system captures the decrease in the vehicle's kinetic energy and converts it into electrical energy to be fed back into the battery.

$$W_c = \frac{1}{\eta_c}\left(\frac{mv^2}{2} + mgh\right) \quad (15)$$

Where:
W_c is the energy stored in the vehicle's power source,
m is the total mass of the vehicle,
V is the speed of the vehicle,
h is the maximum height difference of the BEV,
η_c is the energy efficiency of the power source.

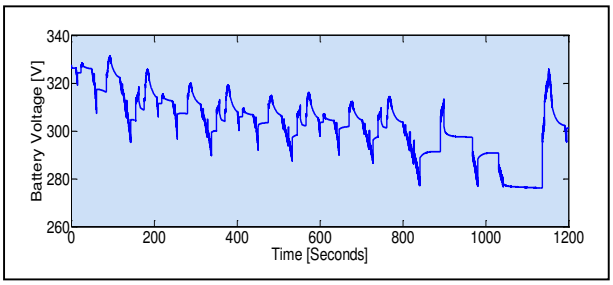


Fig.6. Values of the voltage across the battery terminals.

III. RESULTS AND DISCUSSION

In the first quadrant, the motor consumes energy to provide effort, and the battery is discharged. In the third quadrant, the motor decelerates.

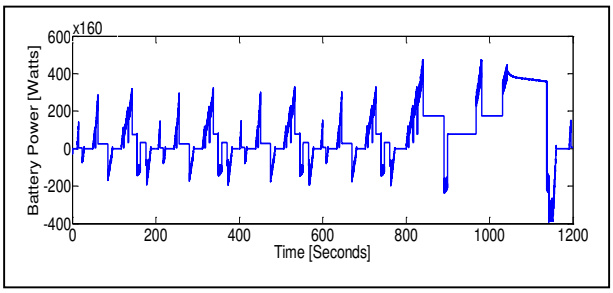


Fig.7. Values of the power supplied by the battery to the motor during the simulation.

When the driver presses the brake pedal, the motor switches to generator mode, and the wheels transfer kinetic energy to the generator. According to the simulation figures, the battery charges and discharges across all four quadrants. While

rotating, the generator converts part of the kinetic energy into electrical energy. This energy is stored in a high-voltage battery. The regenerative braking torque of the electric motor, resulting from energy production, slows down the vehicle.

At low speeds, for example, just before the vehicle stops, the electric motor can no longer provide regenerative braking torque, and the conventional brake must be activated. To maintain the required deceleration, the braking torque provided by the conventional brake is continuously adjusted to match the regenerative braking torque.

This process is called torque addition. Regenerative braking increases the range of electric vehicles and reduces fuel consumption. In the second quadrant, the motor returns energy to the battery through reverse braking, that is, during braking with positive torque. In the fourth quadrant, the battery recharges, but during braking with negative torque. The battery's energy decreases in propulsion mode but increases in regenerative mode.

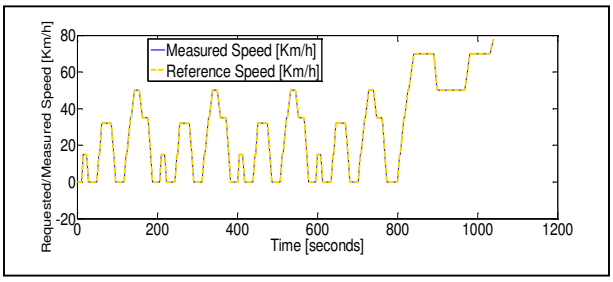


Fig. 8. Comparison between the vehicle speed and the requested speed according to the NEDC cycle.

Figure 1 illustrates the simulation model of an electric vehicle (EV) propulsion system developed based on mathematical equations, represented by each block of the subsystem. From this model, five simulation points were selected and added to output viewers to determine and illustrate the energy flow, performance, and efficiency of two key components of the EV drivetrain: the motor and the battery. The parameters chosen for the simulation work are speed, power, voltage, and battery current. The required power during the simulated driving cycle is calculated based on the power data that the battery can supply, as illustrated in Figure 7. Additionally, in Figure 7, when the torque becomes negative while the speed remains positive, the motor operates in the fourth quadrant and acts as a generator. This operation can be graphically identified either by the negative region on the y-axis of the power curve or by a negative power value.

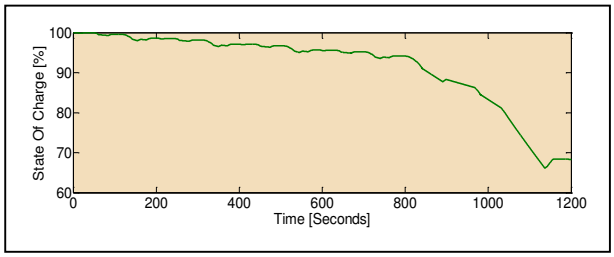


Fig.9. Battery state of charge values based on the NEDC driving cycle.

Figures (6, 7, 10) show the power, voltage, and current drawn from the battery. In these three figures, the battery current curve follows the shape of the motor current and required torque curves. The rise in battery current is therefore directly linked to the increase in torque demand.

From the battery power curve, it is observed that power is drawn from the battery to supply the load in propulsion mode and is returned to the battery in regeneration mode. When the polarity of the voltage and current is identical, the motor operates in standard propulsion mode. However, the motor switches to regenerative braking mode (generator) when the current becomes negative, and the power flows in the opposite direction. Figure. 6 shows the voltage ripple of the battery, that is, the difference between the actual internal voltage of the battery and the voltage calculated from the motor's voltage and current values. The calculated errors will be used by the PI controller, where its modeling is demonstrated to be correct (no negative voltage values). The controller's gain value varies proportionally with the speed demand: an increase in speed demand results in an increase in the gain value. During the modeling, the integrator in the PI controller block is pre-initialized with an initial condition to avoid an algebraic loop error during the simulation. The PI controller compensates for the error by adjusting the selected proportional and integral constants, in order to keep the system in good operating condition.

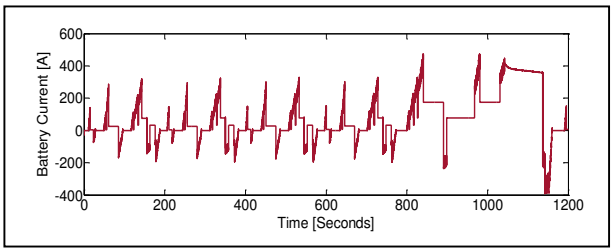


Fig.10. Values of the battery current consumed by the motor during the simulation.

IV. Conclusion

The study in this article focuses on the modeling and simulation of an electric vehicle (EV) in MATLAB-Simulink, with the aim of analyzing the energy flow, performance, and efficiency of the vehicle's drivetrain. In this study, a battery electric vehicle (BEV) is selected, and its components are simulated to analyze the energy flow, performance, and efficiency. Two operating modes are studied during the simulation: motorization and regeneration. The motor's operating mode is determined either by the speed and torque requirements of the road or by the polarity of the motor's current and voltage.

The EV's performance is improved based on the efficiency of the controller used, where its design takes into account the elimination of system errors. In the controller design, we considered maintaining the balance between the input and output power of the battery. A PI controller was also used to compensate for voltage errors. According to the simulation

results, which include the NEDC cycle, it was observed that the vehicle speed and the cycle speed were nearly identical. At this point, it is concluded that the energy consumption data obtained from the model are also accurate.

The battery is recharged using the vehicle's current kinetic energy when the driver removes their foot from the accelerator pedal. According to the driving cycle, the regenerative braking result leads to energy savings. Regenerative braking helps save energy depending on the driving cycle. The vehicle's speed is flexible and can reach significant values within a few seconds. This demonstrates that the driver can provide the necessary acceleration during sudden power demands and rapid acceleration requirements.

The widespread use of electric vehicles will significantly reduce dependence on polluting fossil fuels. Furthermore, supplying electricity from renewable sources will also help limit greenhouse gas emissions.

REFERENCES

[1] Martyushev, N.V. et al, "Mathematical Modeling thePerformance of an Electric VehicleConsidering Various Driving Cycles," Mathematics 2023, 11, 2586. https://doi.org/10.3390/math11112586.

[2] Koech, Alex K. et al, "A review of improvements on electric vehicle battery," Heliyon, Volume 10, Issue 15, e34806.

[3] Yuanjian Zhang et al, "A novel optimal power management strategy for plug-in hybrid electric vehicle with improved adaptability to traffic conditions", Journal of Power Sources. Volume 489, 31 March 2021, 229512.

[4] T.A.T. Mohd et al, " Mathematical modeling and simulation of an electric vehicle, " Journal of Mechanical Engineering and Sciences (JMES) ISSN (Print): 2289-4659; e-ISSN: 2231-8380; Volume 8, pp. 1312-1321, June 2015 © Universiti Malaysia Pahang, Malaysia DOI: http://dx.doi.org/10.15282/jmes.8.2015.6.0128.

[5] Patil, P. (2020), "The Future of Electric Vehicles: A Comprehensive Review of Technological Advancements, Market Trends, and Environmental Impacts," Journal of Artificial Intelligence and Machine Learning in Management, 4(1), 56–68. Retrieved from https://journals.sagescience.org/index.php/jamm/article/view/63.

[6] Shengxi Bai et al, "Overview of energy harvesting and emission reduction technologies in hybrid electric vehicles,"ELSEVIER. Renewable and Sustainable Energy Reviews. Volume 147, September 2021, 111188.

[7] Colin Cole et al, "Modeling, simulation, and applications of longitudinal vehicle dynamics," Vehicle System Dynamics International Journal of Vehicle Mechanics and Mobility," Volume 55, 2017 - Issue 10: State of the Art of the 25th IAVSD Symposium (IAVSD2017, Rockhampton, Australia).

[8] J. Wang and K. Zhang, "Research on modeling and simulation of pure electric vehicle," 2024 7th International Conference on Energy, Electrical and Power Engineering (CEEPE), Yangzhou, China, 2024, pp. 839-844, doi: 10.1109/CEEPE62022.2024.10586462.

[9] Raza, H., Deol, E., Hedge, S., & Tanvir, H. (2024). Model Based Systems Engineering for Sustainable Autonomous Vehicle Design and Development. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 3(4), 188-199. https://doi.org/10.60087/jklst.v3.n4.p188.

[10] X. Huang, "Advisor-based Modeling and Simulation of Hybrid Electric Vehicle," *2024 5th* International Conference on Mechatronics Technology and Intelligent Manufacturing (ICMTIM), Nanjing, China, 2024, pp. 548-553, doi: 10.1109/ICMTIM62047.2024.10629225.

[11] M. D. Rani, D. B. Gundeboyina, H. Chekuri and K. P. Bhimani, "Model Based Design of Electric Vehicle Propulsion System Using Simulink," 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2024, pp. 1-8, doi: 10.1109/INOCON60754.2024.10512037.

[12] Mosleuzzaman , M et al., (2024). "Electric Vehicle Powertrain Design: innovations in electrical engineering," Academic Journal on Innovation, Engineering & Emerging Technology, 1(01), 1–18. https://doi.org/10.69593/ajieet.v1i01.114.

[13] Xiaodan Xu et al." A scalable energy modeling framework for electric vehicles in regional transportation networks," ELSEVIER. Applied Energy. Volume 269, 1 July 2020, 115095.

[14] Lipu, M.S.H. et al. "Artificial Intelligence Approaches for Advanced Battery Management System in Electric Vehicle Applications,". A Statistical Analysis towards Future Research Opportunities. Vehicles 2024, 6, 22–70. https://doi.org/10.3390/ vehicles6010002.

[15] Ryan Van Greunen, Christiaan Oosthuizen, "Data-driven methods for finding coefficients aerodynamic drag and rolling resistance of electric vehicles," Department of Mechanical and Mechatronics Engineering, Tshwane University of Technology, Pretoria 0183, South Africa. World Electr. Veh. J. 2023, 14(6), 134; https://doi.org/10.3390/wevj14060134.

[16] P. K. Prathibha, E. R. Samuel, "Performance Analysis of Electric Car Based on Drag Coefficients and Road Angles," *2022* 2nd International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC), Mathura, India, 2022, pp. 1-6, doi: 10.1109/PARC52418.2022.9726686.

Advanced Virtual Impedance Control Scheme for Power-Sharing Improvement of Three-Phase Paralleled Inverters within a Standalone Microgrid

Camelia Ait Hammouda
LATSI Laboratory, dept. of Renewable Energies
Blida 1 University
Blida 09000, Algeria
aithammouda_camelia@univ-blida.dz

Rafik Bradai
LATSI Laboratory, dept. of electronics
Blida 1 University
Blida 09000, Algeria
bradai_rafik@univ-blida.dz

Rachid Boukenoui
SET Laboratory, dept. of Renewable Energies
Blida 1 University
Blida 09000, Algeria
rachidboukenoui@gmail.com

Abdelhammid Kherbachi
Wind Energy Division
Center of Renewable Energy Development
Bouzareah, Alger, Algeria
a.kherbachi@CDER.com

Ahmed Bendib
SET Laboratory, dept. of electronics
Blida 1 University
Blida 09000, Algeria
ahmeddib28@gmail.com

***Abstract*—The droop control strategy is necessary for the grid-forming operation of Microgrid (MG) to control the power-sharing among parallel-connected Distributed Generations (DGs). However, the main drawback of this control strategy is the poor sharing of the reactive power due to the line impedance mismatch. To overcome this issue, the Virtual Impedance (VI) control concept is adopted as an effective solution to alleviate the impedance mismatch effect thereby improving reactive power sharing. Nevertheless, the sharing of harmonics under nonlinear load operation is the major problem of the existing virtual impedance-based control strategies. This paper proposes an advanced virtual control scheme for improving the reactive power and harmonics sharing among three-phase Voltage Source Inverters (VSIs) forming a standalone MG. The proposed control scheme is designed based on a Multiple Second-Order Generalized Integrator (MSOGI) method, which is introduced to provide accurate estimation of the current harmonics in αβ frame. Therefore, an improvement of the active power, reactive power, and harmonics sharing is expected. For confirming the effectiveness of the proposed VI-based primary control scheme in providing proper power-sharing, a testbed consisting of three DGs-based MG is simulated in MATLAB/Simulink. The results prove the effectiveness of the proposed control scheme for ensuring accurate active and reactive power sharing under linear and nonlinear loads operation.**

Keywords— Droop control, Power-sharing, Virtual Impedance control, MSOGI, Microgrid, Three-phase VSIs.

I. INTRODUCTION

The Microgrid (MG) based on Distributed Energy Resources (DERs) has become an interesting concept for integrating Renewable Energy Sources (RES). Besides, the MG has the ability to operate in both grid-connected and islanded modes [1, 2]. Under islanding mode, the primary controller based on the droop control method is responsible for ensuring load-sharing among Distributed Generation (DG) units and the stability of the frequency and amplitude control within the MG [3, 4]. The droop method is a kind of control strategy adopted to share the active and reactive power between DG units cooperatively, corresponding to the load change by adjusting their voltage and frequency. This control strategy has a major drawback, which is the inability to ensure proper reactive power and harmonics sharing due to the mismatch of the DGs' line impedance. In this regard, the virtual impedance control concept has been adopted to offer the ability to share the current harmonics, improve the reactive power-sharing accuracy, and enhance MG stability [5].

Various approaches have been explored in the literature for three-phase MGs, offering valuable insights into their control and performance enhancement. The classical method involves multiplying the current derivative with an inductance, has been widely adopted in the context of three-phase Voltage Source Inverters (VSIs). This control approach has contributed to enhancing reactive power sharing among DERs.

In [3] VI concept was introduced to ensure power decoupling and sharing challenges by increasing the line impedance between the inverter and common bus. Furthermore, [6] provides a thorough design, analysis and implementation aspects for VI based control for DGs within a MG setting. Notably, adaptive VI control approaches have been brought out, including communication [7], consensus [8] and sliding-based [9] approaches. These innovations aim to to provide accurate harmonic power sharing and voltage harmonic compensation in islanded MGs. On a distributed level, authors in [10] proposed a strategy to achieve robust power sharing and maximize the power transfer through the feeder. The research community has also explored in detail the effects of communication delays and failure on MG performance in [11].

Another concept for the implementation of virtual impedance involves the use of the Second-Order General-Integrator (SOGI) strategy, which has seen further development in [12], and enhancements with the Enhanced SOGI (ESOGI)-based in[13], Moreover, the high order generalized integrator (HOGI) algorithm has gained prominence in addressing power quality issues in [14, 15]. Many studies in the literature provide more analysis of related techniques, such as the TOGI, DOGI, and n-SOGI. These diverse approaches collectively aim to

reduce the output distortions at selected harmonic frequencies caused by current-inherent noise and nonlinear loads, ultimately resulting in improved Total Harmonic Distortion (THD).

From this literature review, it can be noticed that the estimation unit involved in these methods may not be effective in ensuring accurate estimation of the harmonics components, especially in the case of highly distorted current under nonlinear load conditions. This bad estimation may lead to an inaccurate output amount of the VI, and hence, may affect adversely the active, reactive, and harmonic power-sharing accuracy.

In this regard, this paper develops an advanced VI control based on the MSOGI method for three-phase droop-controlled VSIs connected in parallel forming islanded MG. It can provide an accurate estimation of the current fundamental and harmonic components required to compute the output voltage of the VI. Therefore, this accurate estimation may allow the achievement of a precise amount of the VI output needed for reducing the circulating current. Accordingly, it is expected to get a MSOGI-based VICL control scheme that may enhance the active, reactive, and harmonics power-sharing accuracy. A testbed of an MG composed of three paralleled inverters is conducted in MATLAB/Simulink to validate the designed controller under various operation scenarios. The results show high accuracy of the active and reactive power sharing among paralleled VSIs under various operating conditions.

The remainder of this paper is arranged as follows: Section II presents the designed power-sharing control scheme of a three-phase inverter. In section III, the proposed VI controller is described, in addition, to the discussion of the operation principle and mathematical development of the droop control. Section IV highlights the simulation results and discussion of the designed controller. Section V concludes the main findings of the research.

II. Proposed Scheme of Three-Phase VSIs' Power Sharing Control

The structure of the proposed power-sharing control, including the MSOGI-based virtual impedance control unit, intended for the three-phase VSIs is shown in Fig. 1. The power stage of a DG unit involved consists of a DC source supplying an inverter with an LC filter linked to the Point of Common Coupling (PCC) through a line impedance. The proposed controller comprises the droop control unit, the proposed virtual impedance control loop, and the double-loop inner controller.

The inner controller is introduced to ensure the control of the output three-phase voltages of the inverter to their references. Generally, this controller is a double-loop scheme that is composed of an outer voltage and internal current controllers. These control loops handle signals in a stationary reference frame (*dq*). The inner controller provides the duty cycle to the sinusoidal Pulse Width Modulator (PWM), which in turn generates the signal commands of the inverter.

The internal Current Control Loop (CCL) and outer Voltage Control Loop (VCL); in this study, are based on Proportional-Integral (PI) regulators, and their corresponding closed-loop transfer functions are expressed as follows:

$$\begin{cases} G_{v.dq}^{BF}(s) = \dfrac{v_{C.dq}}{v_{C.ref.dq}} = \dfrac{1}{C}\dfrac{k_{p.v}\ s + k_{i.v}}{s^2 + \left(\dfrac{k_{p.v}}{C}\right)s + \dfrac{k_{i.v}}{C}} \\ G_{i.dq}^{BF}(s) = \dfrac{i_{L1.dq}}{i_{L1.ref.dq}} = \dfrac{1}{L_1}\dfrac{k_{p.i}\ s + k_{i.i}}{s^2 + \left(\dfrac{r_1 + k_{p.i}}{L_1}\right)s + \dfrac{k_{i.i}}{L_1}} \end{cases} \quad (1)$$

where $k_{p.i}$, $k_{p.v}$, $k_{i.i}$ and $k_{i.v}$ are the proportional and integral gains of current and voltage controllers, respectively.

By matching (1) with the desired transfer functions given in (2), below, the controllers' gains can be determined.

$$\begin{cases} G_{i.dq}^{FB}(s) = \dfrac{\omega_i^2}{s^2 + 2\zeta_i\omega_i s + \omega_i^2} \\ G_{v.dq}^{FB}(s) = \dfrac{\omega_v^2}{s^2 + 2\zeta_v\omega_v s + \omega_v^2} \end{cases} \quad (2)$$

Note that the parameters $\zeta_i, \omega_i, \zeta_v$, and ω_v are precisely obtained from the transient response overshoot $M_{p.i}$, $M_{p.v}$, and settling time $T_{s.i}$, $T_{p.v}$ expressed by :

$$M_{p.i} = e^{\frac{\zeta_i\pi}{\sqrt{1-\zeta_i^2}}},\ T_{s.i} = \frac{4}{\zeta_i\omega_i},\ M_{p.v} = e^{\frac{\zeta_v\pi}{\sqrt{1-\zeta_v^2}}},\ T_{s.v} = \frac{4}{\zeta_v\omega_v} \quad (3)$$

More details about the inner controller can be found in [16], while the droop control and the proposed Virtual Impedance Control Loop (VICL) are discussed in detail in the following section.

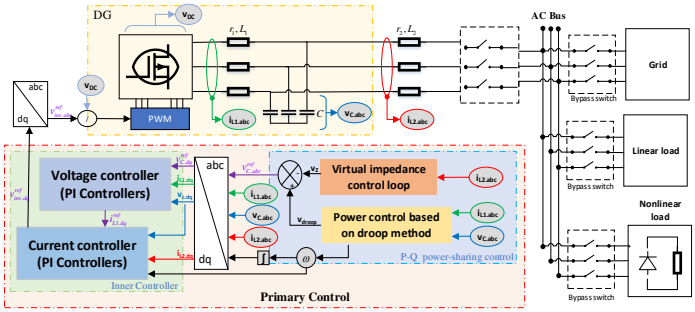


Fig. 1. Block diagram of the primary control.

III. The Proposed VI Control Loop

The proposed virtual control scheme based on the MSOGI method is highlighted in this section. However, before that, the droop control concept and modeling are first presented.

A. *Droop Control unit*

The active and reactive power sharing is often accomplished by the droop control technique, which adopts the calculated active and reactive powers as inputs and produces the output voltage reference of the inverter. The droop control is an effective solution that has been adopted to coordinate the DG units within an MG. One of the major advantages of this control strategy is the ability to enable communication-less power-sharing among parallel-connected DG units. The operation principle of the droop method is based on adjusting the frequency, *f*, and voltage, *E*, of an inverter in order to control the power balance. This concept is inspired by synchronous

generators, which drop their frequency to inject more active power.

As shown in Fig. 2, the droop control unit includes a power calculation block with a low pass filter (LPF), *f*/*P* and *V*/*Q* droop controllers, and a sinusoidal generator. In this control scheme, the calculated average *P* and *Q* powers in the *αβ*-frame based on (4), below, are used by the droop controller to produce the frequency and voltage references ω_i E_i, respectively.

$$\begin{cases} P_i = \left(v_{c.\alpha} \times i_{L2.\alpha} + v_{c.\beta} \times i_{L2.\beta}\right) \\ Q_i = \left(v_{c.\beta} \times i_{L2.\alpha} - v_{c.\alpha} \times i_{L2.\beta}\right) \end{cases} \quad (4)$$

Next, a sinusoidal generator is used for generating the three-phase *abc* references of the inverter output voltage according to the following expression:

$$v^*_{droop}(t) = E_i \sin\left(\omega_i t + \begin{bmatrix} 0 & -\frac{2\pi}{3} & -\frac{2\pi}{3} \end{bmatrix}^T\right) \quad (5)$$

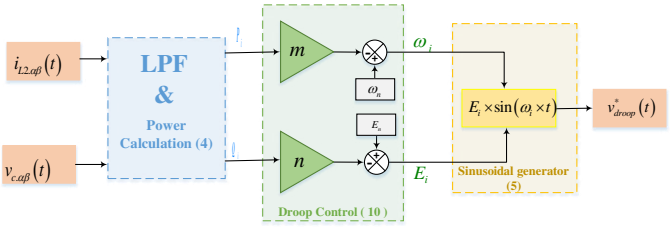


Fig. 2. Block diagram of the droop control unit.

To derive the droop functions, we should analyze the equivalent circuit of two-Voltage Source Inverters (VSIs) connected in parallel to an AC bus via line impedances as given in Fig. 3. In this figure, each inverter stage is modeled as a sinusoidal voltage source with an output impedance in series. According to this figure, the expression of the active *P* and reactive *Q* power of each DG, after some mathematical manipulation, can be obtained as follows:

$$\begin{cases} P_i = \frac{3V_{PCC}}{Z_i}\left(\left(E_i \cos(\varphi_i) - V_{PCC}\right)\cos(\phi_i) + E_i \sin(\varphi_i)\sin(\phi_i)\right) \\ Q_i = \frac{3V_{PCC}}{Z_i}\left(\left(E_i \cos(\varphi_i) - V_{PCC}\right)\sin(\phi_i) - E_i \sin(\varphi_i)\cos(\phi_i)\right) \end{cases} \quad (6)$$

where $E_i\angle\varphi_i$ is the inverter output voltage, $E_{pcc}\angle 0$ is the voltage at the PCC, and $Z_i\angle\phi_i = R_i + jX_i$ is the inverter to the PCC bus impedance, which considers the inverter output impedance and the line impedance of the connection wires.

By assuming that the phase difference between the inverters' output voltages and the PCC, is very small, ($\varphi_i << 1 \rightarrow \sin\varphi_i = \varphi_i$ and $\cos\varphi_i = 0$), and For pure inductive line impedance i.e., $X_i >> R_i$ ($\rightarrow Z_i\angle\phi_i \cong X_i\angle\pi/2$), where X_i and R_i are the reactance and resistance of the inverter output impedance. the expressions of the droop controller can be formulated as follows:

$$\begin{cases} \omega_i = \omega^* - \left(\frac{\omega_f}{s+\omega_f}\right) nP_i \\ E_i = E^* - \left(\frac{\omega_f}{s+\omega_f}\right) mQ_i \end{cases} \quad (10)$$

where ω^* and E^* correspond to the nominal frequency and amplitude output voltage at no load. *n* and *m* are the frequency and amplitude droop gains which can be determined by the maximum frequency and voltage deviations (Δf_{max} and ΔV_{max}) divided by the inverter *P* and *Q* rated power, and the first-order transfer function is of the transfer function of the LPF with a cut-off frequency ω_f.

It is worth mentioning that in this work we considered the case of pure inductive impedance. It can be noticed that in the case of pure inductive impedance, the active power depends on the phase angle , and the reactive power depends on the voltage . The *f*/*P* and *V*/*Q* droop characteristics are presented in Fig. 4.

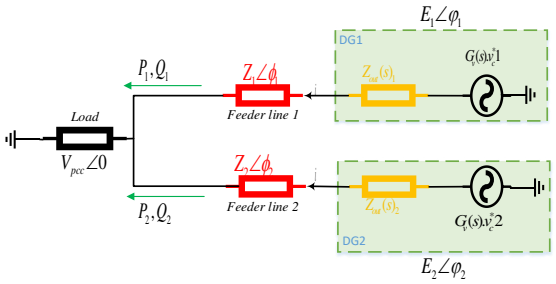


Fig. 3. *Electric circuit for two paralleled VSI*

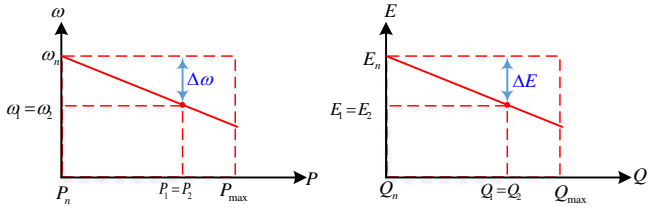


Fig. 4. *f*/*P* and *V*/*Q* characteristics of the droop control.

B. Proposed Virtual Impedance Control Loop

By using the droop method, the inverters' outputs and line impedances can affect the power-sharing accuracy. Also, feeding nonlinear loads creates current harmonics, therefore, a Virtual Impedance Control Loop (VICL) is proposed for enhancing the power-sharing performance. This VICL is introduced to improve the accuracy of the power-sharing and allow the share of current harmonics by normalizing the output impedance of the VSIs. As shown in Fig, 1, this control block considers the output current of the inverter as an input and provides an output voltage to be subtracted from the droop voltage reference.

The structure of the proposed VI intended for a three-phase VSI is depicted in Fig. 5. As seen, it is based on the Multi-Second-Order Generalized Integrator (MSOGI) method, which uses the *α*-version of the output current ($i_{L2.\alpha\text{-}in}$); obtained via *abc*-*αβ* transform; to extract the filtered direct and orthogonal fundamental and their 5th, 7th, and 11th harmonic components ($i_{L2.\alpha\beta\text{-}5,7,11}$). Note that the MSOGI uses the frequency provided by the droop control. Accordingly, the output of the VI, in the *s*-domain, can be derived as follows:

$$v_z(s) = \left[T_{\alpha\beta-abc}\right]\left[r_v\, i_{L2.\alpha\beta} - L_v\omega \times \left(i_{L2.\alpha\beta-1} + 5\, i_{L2.\alpha\beta-5} + 7\, i_{L2.\alpha\beta-7} + 11\, i_{L2.\alpha\beta-11}\right) e^{-j\frac{\pi}{2}}\right] \quad (12)$$

where $T_{\alpha\beta\text{-}abc}$ defines the inverse Clarke transform on the *αβ* components.

These *αβ* voltage components are transferred to *abc* frame and, then, subtracted from the ones generated by the droop

controllers, resulting in the new voltage reference defined as follows:

$$v_{C.ref}(s) = v_{droop.ref}(s) - v_z(s) \quad (13)$$

The structure of the MSOGI used to estimate the current components is presented in Fig. 6. It consists of four SOGI associated in parallel, corresponding to fundamental, 5th, 7th, and 11th harmonics current estimation. The input of a SOGI corresponding to *h* harmonic (*h* = 1; 5; 7; 11) is the sum of all the other current components estimated by the other SOGIs, subtracted from the actual current. In addition, each SOGI is tuned by a frequency resulting from the multiplication of the droop control frequency with respective *h* harmonic. An application of this scheme can be found in [13] . The structure of the SOGI is given in Fig. 7, which considers the sinusoidal single voltage as input and provides in-phase and quadrature-phase components as outputs.

As a result, the MSOGI can ensure the estimation of the current components with high accuracy even under harmonically distorted current in addition to providing a fast dynamic response. This may lead to improving the droop control performance; therefore, an enhancement of the power-sharing accuracy is expected.

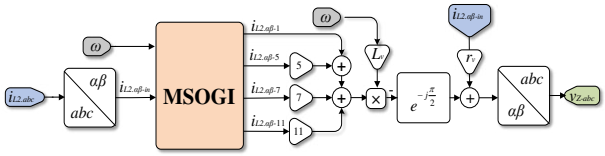


Fig. 5. Structure of the VI control scheme based on MSOGI.

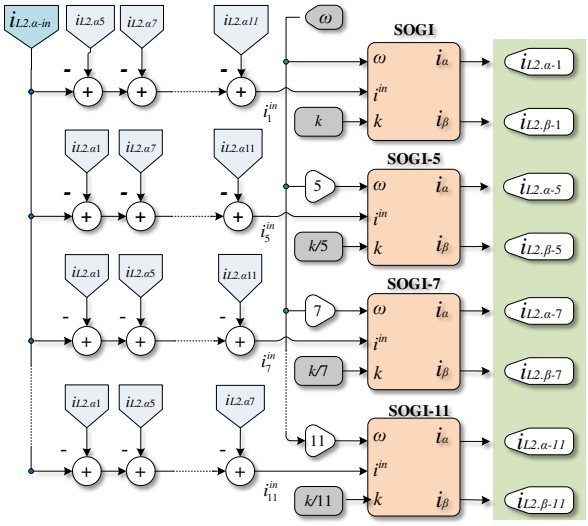


Fig. 6. Multiple-SOGI structure.

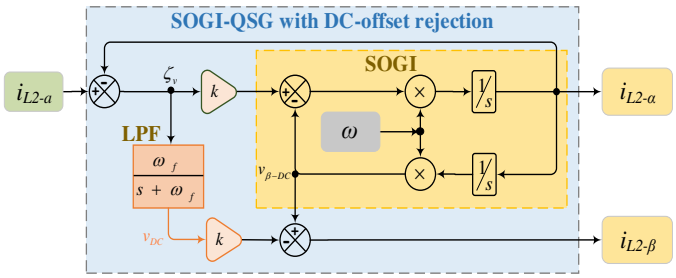


Fig. 7 Structure of the SOGI method.

IV. Simulation Results and Discussion

To assess the effectiveness of the proposed control scheme, an MG model as shown in Fig. 8 is built in MATLAB/Simulink. The controller under study is configured according to the parameters presented in Table I. In this simulation, the performance of the designed controller is tested under various scenarios.

TABLE I: VSIs and Control Parameters

Parameters	Symbol	Unit	Value
Nominal Voltage (RMS)	E_n	V	220
Nominal frequency (f_n)	f_n	Hz	50
Switching frequency	f_s	kHz	20
Simulation frequency	f_e	MHz	1
DC voltage	U_{DC}	V	450
Output filter capacitor	C	µF	23
Output filter inductor	L,r	mH, Ω	2, 1
Line impedance of DG #1	L_1	mH, Ω	1.5, 0.8
Line impedance of DG #2	L_2	mH, Ω	0.5, 0.8
Virtual inductance	L_v	mH	2.7
Virtual Resistance	R_v	Ω	1
P-ω droop	m	rad/(W. s)	0.0005
Q-*V* droop	n	V/Var	0.001
Voltage controller P gain	k_{pv}	µF.rad/s	0.1839
Voltage controller I gain	k_{pi}	mH.rad/s	183.87
Current controller P gain	k_{iv}	mH.rad/s	6.2831

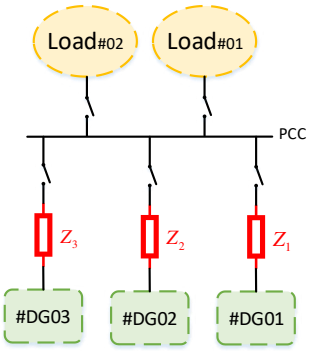


Fig. 8. Simulation network.

Scenario 01: Linear load changes

In this test, the DG units start with no load operation, and then a linear load is added at t = 0.05 s. Next, a second load is connected and disconnected at t = 0.2 s and 0.5 s respectively. The obtained results demonstrating the transient performance of the active power and frequency of DG #1, #2, and #3 are shown in Figs 9 and 10. Fig. 11 displays the plots of the output voltage and current at the PCC.

According to these results, the following remarks can be observed:

- At no-load operation, the active power, reactive power, and currents of DG #1, #2, and #3 are equal to zero, while the frequency and voltage are set to their nominal values. (i.e. 50 Hz and 220 V (RMS))
- When a load is connected, the active and reactive powers of the DGs are equal, which means that the load power is shared between the DG units. From Fig. 9, It can be seen that the

frequency drooped with a significant amount to inject the required active power, while the voltage almost remains unchanged since the reactive power is set to zero. Regarding the voltage and current in the *abc* frame, given in Fig. 11, it can be noticed that proper sinusoidal forms are achieved.

- The same observations are considered when the second load is connected and disconnected.

- High transient responses are achieved during load changes.

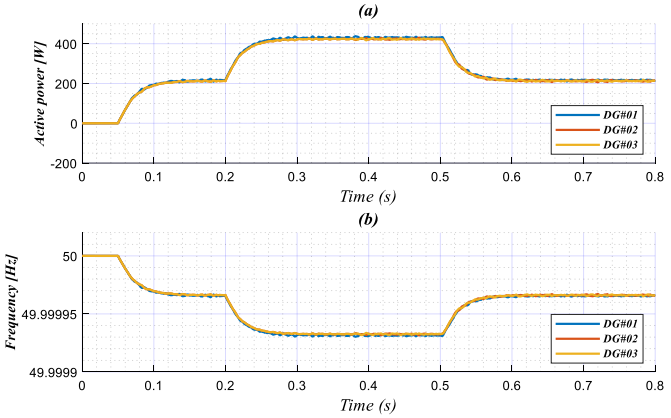


Fig. 9. Obtained results to load change: (a) active power and (b) frequency of VSI #1, #2, and #3.

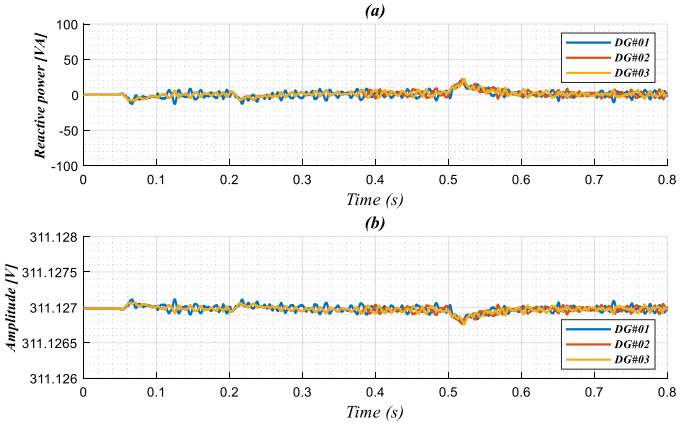


Fig. 10. Obtained results to load change: (a) reactive power and (b) amplitude of VSI #1, #2, and #3.

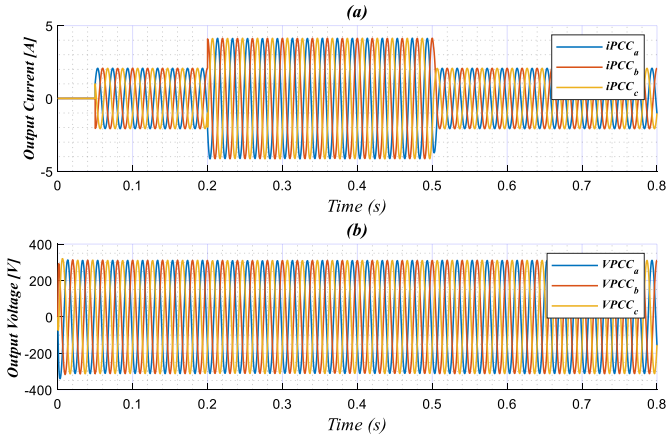


Fig. 11. Obtained results to load change: (a) current and (b) voltage at the PCC.

Scenario 02: Nonlinear load changes

In the following test, the proposed controller is subject to similar working conditions as in the first one, except using nonlinear loads. Fig. 12-14 shows the obtained results, which show the same variables as the first scenario. They reveal that, as for linear load feeding in the first test, the proposed controller is able to distribute equally the active as well as reactive power, ensuring the MG output voltage and current stability.

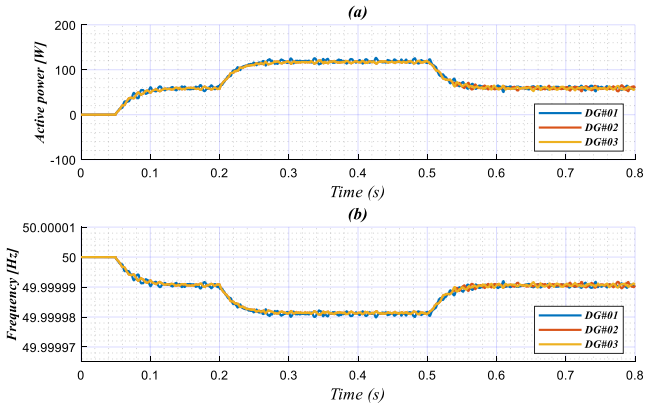


Fig. 12. Obtained results to nonlinear load change: (a) active power and (b) frequency of VSI #1, #2, and #3.

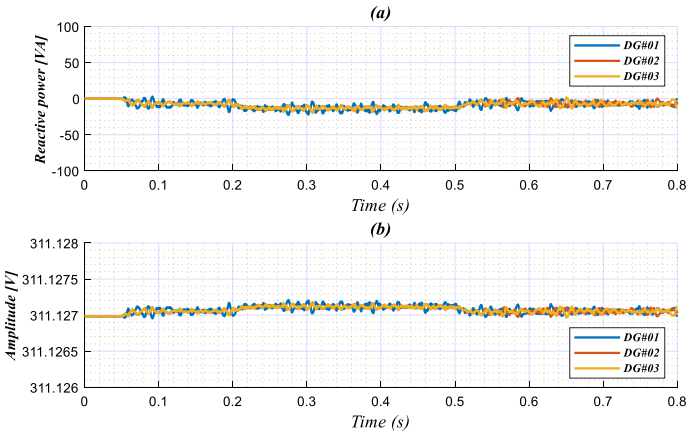


Fig. 13. Obtained results to nonlinear load change: (a) reactive power and (b) amplitude of VSI #1, #2, and #3.

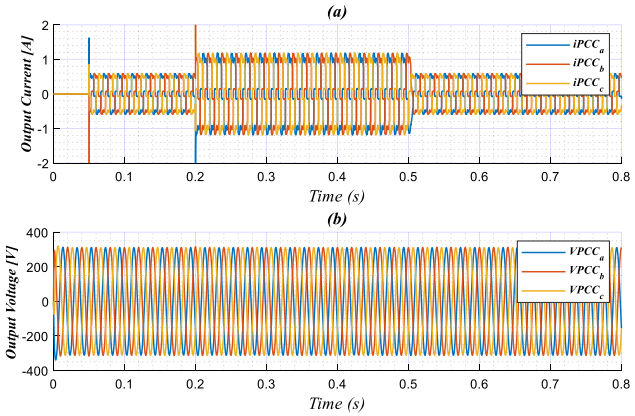


Fig. 14. Obtained results to nonlinear load change: (a) current and (b) voltage at the PCC.

Scenario 03: DG units plug in/out

Finally, the ability of the proposed control to handle the connection/disconnection of the DGs is tested. So, the first DG

is the only one to supply the load from 0.1s. Then, DG #02 and DG#03 are connected at 0.3s and 0.6s respectively. Thereafter, DG#03 is disconnected again at 0.9 s. Figures 15-17 show the obtained results for this test and portray the same variables as the previous scenarios. From these plots, it can be observed that the proposed controller is able to handle the DGs connection and disconnection. In addition, it can be seen that the active and reactive are shared equally after the DGs are connected.

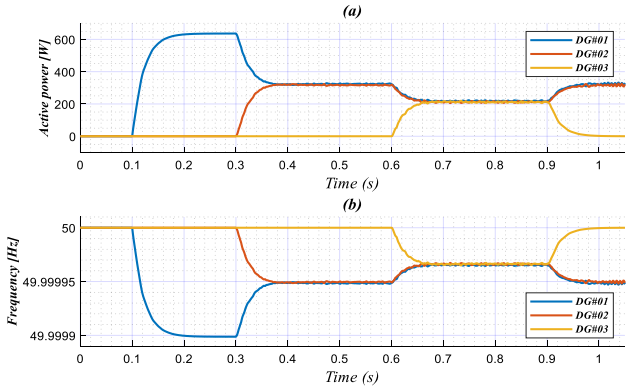


Fig. 15. Obtained results to DG plug in/out.

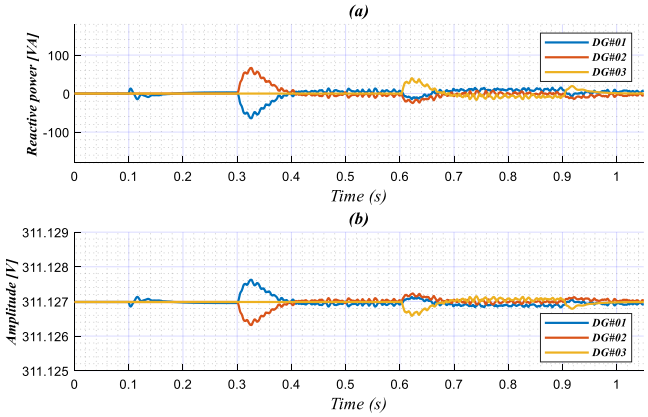


Fig. 16. Obtained results to DG plug in/out.

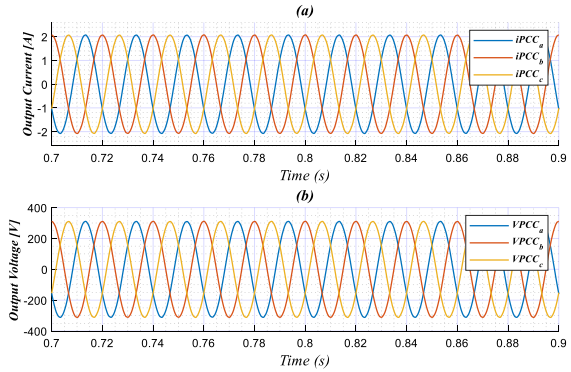


Fig. 17. Obtained results to DG plug-in/out

V. Conclusion

In this paper, a power sharing control with a VI control scheme based on the MSOGI method was proposed. This control scheme is adopted for the aim of improving the power sharing between droop-controlled three-phase VSIs in islanded MG. The principal operation of the droop control was explained, and mathematical modeling was developed. In addition, the structure of the proposed VI control and the corresponding mathematical formulation was provided. Furthermore, the schematic of the MSOGI method was described. A simulation model of an MG with three three-phase VSIs was conducted in Matlab/Simulink to assess the performance of the proposed controller. The obtained results have validated the effectiveness of the proposed controller in achieving accurate active and reactive power sharing between three-phase VSIs under linear and non-linear loads supply, and DG plug-in/out operation as well.

References

[1] J. M. Guerrero, V. Luis Garcia de, J. Matas, M. Castilla, and J. Miret, “Output impedance design of parallel-connected UPS inverters with wireless load-sharing control,” IEEE Transactions on Industrial Electronics, vol. 52, no. 4, pp. 1126-1135, 2005.

[2] N. Nasser, and M. Fazeli, “Buffered-Microgrid Structure for Future Power Networks; a Seamless Microgrid Control,” IEEE Transactions on Smart Grid, vol. 12, no. 1, pp. 131-140, 2021.

[3] J. M. Guerrero, J. C. Vasquez, J. Matas, L. G. d. Vicuna, and M. Castilla, “Hierarchical Control of Droop-Controlled AC and DC Microgrids—A General Approach Toward Standardization,” IEEE Transactions on Industrial Electronics, vol. 58, no. 1, pp. 158-172, 2011.

[4] E. Kabalcı, "Hierarchical Control in Microgrid," pp. 381-401, 2019.

[5] B. Keyvani-Boroujeni, B. Fani, G. Shahgholian, and H. H. Alhelou, “Virtual Impedance-Based Droop Control Scheme to Avoid Power Quality and Stability Problems in VSI-Dominated Microgrids,” IEEE Access, vol. 9, pp. 144999-145011, 2021.

[6] J. He, and Y. W. Li, “Analysis, Design, and Implementation of Virtual Impedance for Power Electronics Interfaced Distributed Generation,” IEEE Transactions on Industry Applications, vol. 47, no. 6, pp. 2525-2538, 2011.

[7] H. Mahmood, D. Michaelson, and J. Jiang, “Accurate Reactive Power Sharing in an Islanded Microgrid Using Adaptive Virtual Impedances,” IEEE Transactions on Power Electronics, vol. 30, no. 3, pp. 1605-1617, 2015.

[8] H. Zhang, S. Kim, Q. Sun, and J. Zhou, “Notice of Removal: Distributed Adaptive Virtual Impedance Control for Accurate Reactive Power Sharing Based on Consensus Control in Microgrids,” IEEE Transactions on Smart Grid, vol. 8, no. 4, pp. 1749-1761, 2017.

[9] S. Chaturvedi, D. Fulwani, and J. M. Guerrero, “Adaptive-SMC Based Output Impedance Shaping in DC Microgrids Affected by Inverter Loads,” IEEE Transactions on Sustainable Energy, vol. 11, no. 4, pp. 2940-2949, 2020.

[10] B. Wei, A. Marzàbal, R. Ruiz, J. M. Guerrero, and J. C. Vasquez, “DAVIC: A New Distributed Adaptive Virtual Impedance Control for Parallel-Connected Voltage Source Inverters in Modular UPS System,” IEEE Transactions on Power Electronics, vol. 34, no. 6, pp. 5953-5968, 2019.

[11] T. V. Hoang, and H. H. Lee, “Virtual Impedance Control Scheme to Compensate for Voltage Harmonics With Accurate Harmonic Power Sharing in Islanded Microgrids,” IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 9, no. 2, pp. 1682-1695, 2021.

[12] X. Xiaofei, J. Yaoqin, H. Changhua, Y. Xumeng, W. Zechen, and G. Benhui, "An enhanced microgrid power sharing strategy based on multiple second order generalized integrators harmonic detection and adaptive virtual impedance." pp. 2343-2347.

[13] C. A. Hammouda, R. Bradai, A. Kherbachi, A. Bendib, R. Boukenoui, and K. Kara, "ESOGI-Based Virtual Impedance Control Scheme with Performance Improvement for Droop-Operated Three-Phase VSIs in Islanded AC Microgrid." pp. 1-6.

[14] Z. Xin, R. Zhao, P. Mattavelli, P. C. Loh, and F. Blaabjerg, “Re-Investigation of Generalized Integrator Based Filters From a First-Order-System Perspective,” IEEE Access, vol. 4, pp. 7131-7144, 2016.

[15] X. Lin, R. Zamora, and C. Baguley, "Droop Control Based on Improved Virtual Impedance in a Stand-Alone Microgrid." pp. 883-888.

[16] C. A. Hammouda, R. Bradai, R. Boukenoui, A. Kherbachi, and A. Bendib, "Modeling and Design of Primary Control’s Inner Loops for Droop-Controlled Three-Phase Inverters within a Microgrid System." pp. 1-6.

Monitoring systems technologies for photovoltaic systems : Overview

Dermouche Reda (Author)
dept. Electrical Engineering And Industrial Computing
National Higher School of Advanced Technologies(ENSTA)
Algiers, Algeria
reda.dermouche@ensta.edu.dz

BOUDALI Khadidja (Author)
dept. Electrical Engineering And Industrial Computing
National Higher School of Advanced Technologies (ENSTA)
Algiers, Algeria
k_boudali@ensta.edu.dz

BOUABDELLAH Hiba (Author)
dept. Electrical Engineering And Industrial Computing
National Higher School of Advanced Technologies(ENSTA)
Algiers, Algeria
h_bouabdellah@ensta.edu.dz

TADJER Sid ahmed (Author)
Faculty of Hydrocarbon and Chemistry
University M'Hamed Bougara, Boumerdes
Algeria
s.tadjer@univ-boumerdes.dz

***Abstract*—The rapid expansion of photovoltaic (PV) installations globally has introduced significant challenges in system control and tracking. This paper provides a comprehensive overview of the current state of monitoring systems technologies for PV systems. It delves into the architecture and functionalities of various monitoring technologies, categorizing them into data processing modules and data transmission protocols. The study explores different implementations, including Arduino, Raspberry Pi, and PLC-based modules, as well as communication protocols such as WiFi, Bluetooth, ZigBee, and LoRa. By leveraging these advanced monitoring technologies, stakeholders can gain valuable insights into system performance, diagnose faults, and implement proactive maintenance strategies. The findings highlight the advantages and potential areas for further research and development in monitoring technologies, ensuring the continued success and sustainability of PV installations in the transition towards clean and renewable energy sources.**
Keywords: Renewable energy, Monitoring, Photovoltaic, IoT
***Index Terms*—Renewable energy, Monitoring, Photovoltaic, Photovoltaic, IOT**

I. Introduction

The efficiency and performance of PV systems are highly influenced by various external factors, such as weather conditions, shading, dust accumulation, and temperature fluctuations. For instance, solar irradiance—the power per unit area received from the Sun—directly impacts the amount of electricity generated. Similarly, shading from trees, buildings, or other obstructions can significantly reduce a system's output. Dust and dirt on solar panels can also obstruct sunlight, diminishing efficiency, while extreme temperatures can affect the semiconductor properties of photovoltaic cells, altering their performance.

Given these variables, continuous monitoring of PV systems is crucial to ensure optimal performance and longevity. Monitoring systems provide real-time data on the functioning of the PV array, helping to identify and diagnose issues swiftly. This can include tracking energy production, detecting faults, and analyzing performance trends over time. By employing sophisticated monitoring technologies, operators can maintain the efficiency of PV systems, predict maintenance needs, and ultimately maximize the return on investment. Furthermore, effective monitoring can aid in the integration of PV systems into broader energy grids, ensuring stable and reliable energy supply.

In order to explore this point, this article aims to provide a detailed overview of monitoring technologies for PV systems, highlighting innovations, advantages, and potential areas for further research and development, structured as follow: In section II we will present the general architecture of a monitoring systems followed by a technologies classification in section III with slight detail for each one, ended by a study of the key challenges facing this technologies in section VI

II. Implementing a Three-Layer IoT Architecture for Advanced PV Monitoring

Layered architecture for IoT applications is typically utilized to provide suitable frameworks for devices, data storage, and processing management. To meet these requirements, various IoT architectures can be deployed at the device, gateway, and cloud layers. While there is no single consensus on IoT architecture, many contributions suggest a three-layer architecture comprising perception, network, and application layers [1].

In the context of photovoltaic (PV) monitoring systems, this three-layer architecture can be effectively implemented to ensure efficient data management and processing. The figure 1 illustrates the three-layer architecture, specifically tailored for PV monitoring technologies.

By integrating these layers, the proposed IoT architecture facilitates efficient data acquisition, management, and utilization, enabling advanced monitoring and optimization of PV systems.

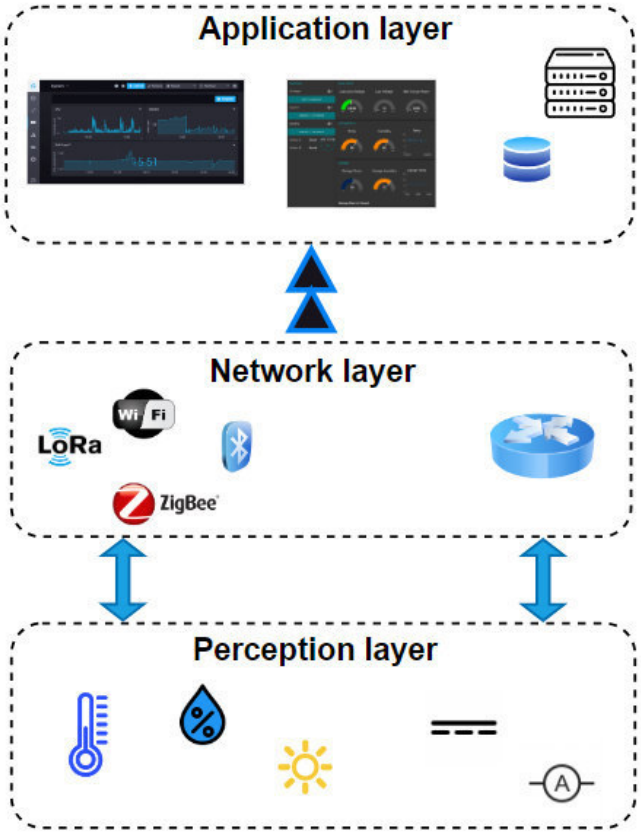


Fig. 1. Three-Layer IoT Architecture for PV Monitoring Systems.

III. Monitoring Technologies

Several and diverse models and system architectures have been proposed for monitoring PV stations using different technologies. Therefore, we chose to classify these systems into two categories: data processing modules and data transmission protocols. illustrated in figure2

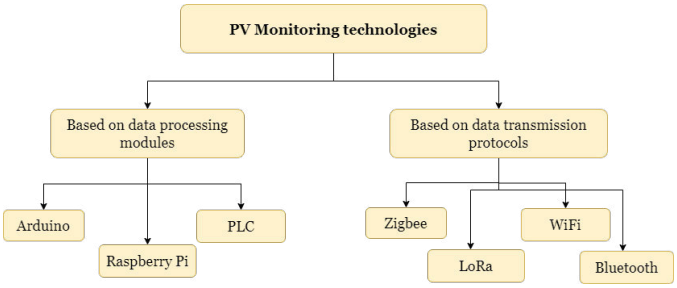


Fig. 2. Classification of PV system monitoring technologies

Data processing modules focus on the collection, storage, and analysis of data from various sensors. These modules are designed to process and temporarily store data before transmitting it to the display and storage systems. On the other hand, data transmission protocols deal with the wireless communication aspects of the monitoring systems, ensuring that data is efficiently transmitted over various distances and conditions.

A. *Data processing modules*

An extensive evaluation of various data processing modules for a solar PV monitoring system has been carried out, offering insights into their implementation, design, specifications and software utilized.

1) **Arduino based module**
The Arduino technology, created by the Ivrea Interaction Design Institute. It has gained popularity as an accessible open-source hardware tool, featuring easy access to hardware design and software coding. Arduino, a low-cost prototyping platform, is based on an ATMEL microcontroller.
JAMIL et al. proposed using arduino to improve the performance of a solar-powered floating photovoltaic system [2]. The Arduino nano is the main controller of the system. Monitoring the voltage, current, and power output of a solar module with a power rating of 10W. Furthermore, the module's monitoring was evaluated under two conditions: on the water's surface (the floating system) and on land (the land system).
The results showed that the floating system had higher power output and efficiency than the land system. As a result, it has met all of its objectives in terms of generating electricity and monitoring system output.

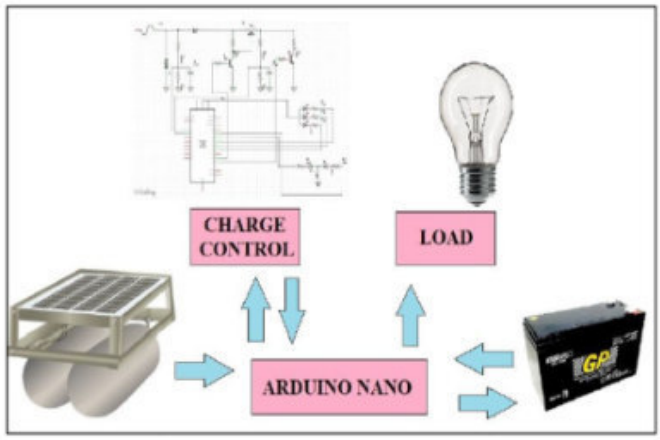


Fig. 3. The complete circuit of the project [2].

Allafi et al [3] created a monitoring system using Reliance SCADA and Arduino Uno on a small photovoltaic (PV) power system to monitor PV current, voltage, battery, and efficiency. The primary objective of the proposed system figure 4 was to assess the MPPT efficiency by analyzing data gathered from the sensors. The data collecting system is comprised of an Arduino controller and sensors. All data is transferred to a PC and displayed on a user interface created by Reliance SCADA. Furthermore, the data is saved to a computer as an Excel file. This enables users and operators to

track the parameters of the PV system in real time. Furthermore, the MODBUS library is installed to the Arduino Uno, allowing communication with Reliance SCADA via USB cable via the MODBUS RTU protocol. The results of the trials reveal that SCADA works in real time and can be efficiently used for monitoring a solar energy system.

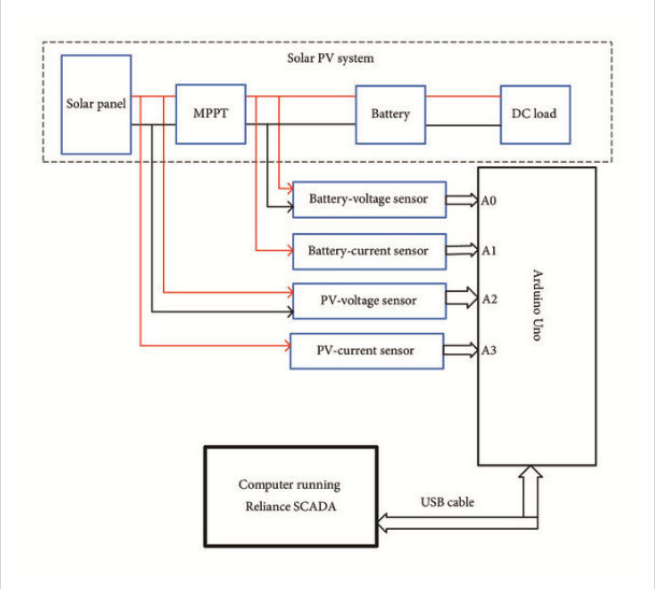


Fig. 4. Design of system proposed [3].

2) **Raspberry Pi based module**

The Raspberry Pi is a low-cost SBC developed by the Raspberry Pi Foundation. It was designed to be a highly flexible and capable computer for a fraction of the cost of a standard PC, allowing anyone to solve issues imaginatively.This can include interactive and autonomous environmental monitoring in real-time [4]. Bikrat et al. [5] created a wireless method for monitoring photovoltaic plants in real time figure 5. The proposal electronic and computer system (software and hardware) is built around six photovoltaic panels and their inverters and is based on the Raspberry Pi3. Each inverter includes a Bluetooth module that communicates with the Raspberry Pi3. This latter is primarily responsible for connecting the inverters to the servers in order to retrieve data and store it directly in an online database. The proposed model might provide measurements of numerous quantities, including current, voltage, and energy as electrical characteristics, meteorological characteristics include illumination and temperature.

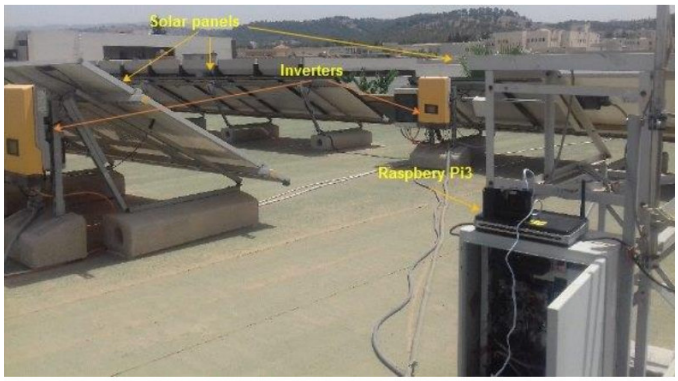


Fig. 5. A wireless prototype for monitoring solar PV systems [5].

Pereira et al. [6] created and implemented the Renewable Energy Monitoring System (REMS) for decentralized renewable energy facilities figure 6, which incorporates Raspberry Pi, multi-user cloud remote monitoring, and the Internet of Things. REMS employs analog and digital sensors. ADCES enables online monitoring of measured variables, while the RPi reads digital sensors. Data is saved in an online server database with a storage capacity of 84.44 MB and 2 GB, respectively. For 391 days, the data was kept at a rate of one sample per minute in a 150-byte format, also available as charts and tables with the free program Web Monitor.

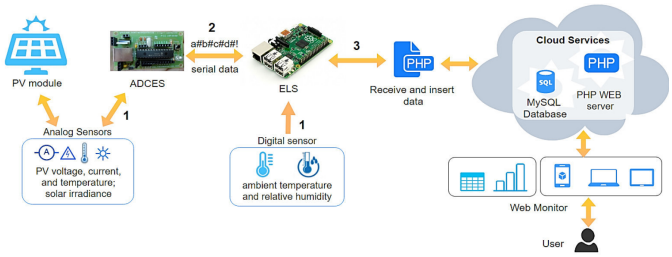


Fig. 6. REMS general diagram [6].

3) **PLC based module**

.

Several remote monitoring methods in terms of renewable energy monitoring such as the suggestion of Kabalci and Kabalci [7], for monitoring solar photovoltaic (PV) systems using power line communication (PLC) technology. to establish a microgrid model developed within the MATLAB/Simulink environment. The system consisted of a DC-AC converter, three solar power plants featuring maximum power point tracking (MPPT) systems, and a multilevel inverter to generate three-phase AC line voltage. Additionally, a 25 km transmission line and PLC modems were incorporated into the setup.

Their approach offered an efficient and economical solution for monitoring solar PV systems within microgrid environments, since power lines carried the generated voltage as well as transferred the power drawn from the

loads. the block diagram of generation, energy conversion, transmission and monitoring stages is illustrated in figure 7.

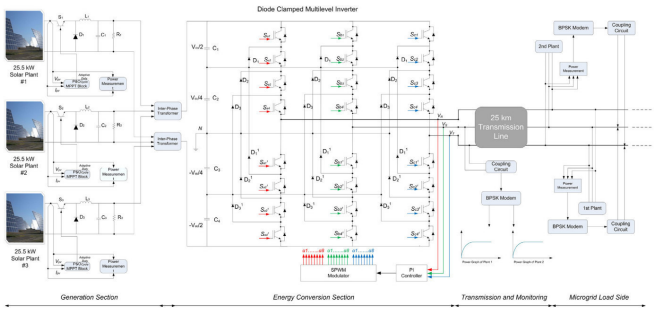


Fig. 7. Block diagram for a solar PV monitoring system implementing the PLC module [7].

Han et al. [8] proposed a PV monitoring system that uses low-cost power line communication (PLC) technology, eliminating the need for additional communication modules. Each PV module is equipped with a PLC module for detailed monitoring, and data from these modules, along with inverter data, are aggregated by a data logger. This integrated system connects to home networks via home plug devices, allowing users to access data conveniently through a smart app. The app provides graphical representations for easy interpretation of the performance metrics. This approach aims to optimize residential PV system efficiency and reduce energy costs. Figure 8 depicts the configuration of a field test for this monitoring system, where each PV module has specifications of 49.4 V, 8.1 A, and a maximum power output of 400 W.

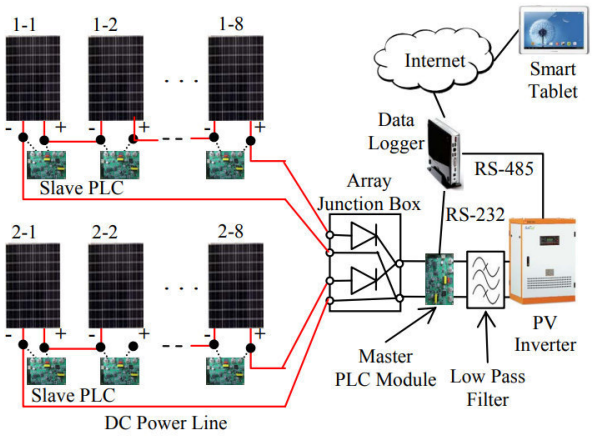


Fig. 8. Configuration of field test for the PV monitoring system [8].

B. Data transmission protocol

1) **WiFi**

WiFi, short for Wireless Fidelity, is a technology that allows electronic devices to exchange data wirelessly over a computer network, including high-speed internet connections. It's based on the IEEE 802.11 family of standards [9].

The ESP8266 controller, was widely used for WiFi remote applications due to its integrated WiFi chip, providing a gateway to the internet. On the other hand, NodeMCU is an open-source firmware and development kit for ESP8266 devices. Based on the Lua scripting language, it offers an easy-to-use platform for programming ESP8266 modules. NodeMCU allows developers to focus on creating IoT applications without needing to delve deeply into the intricacies of low-level programming [10].

Similar to the system introduced by Sutikno et al, it is an IoT-based monitoring system that collects data from sensing units including current, voltage, temperature, and irradiance sensors. This data is sent to the ThingSpeak database, through the esp82 wifi module, providing an online monitoring dashboard. Additionally, this article proposes an additional data storage option on an SD card in case of internet connection failure. The final design of the system is presented in a block diagram in the figure9 below [11].

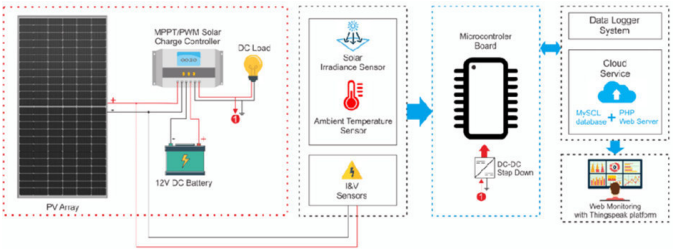


Fig. 9. Block diagram of a proposed IoT-based PV monitoring system [11].

2) **Bluetooth and BLE**

Bluetooth is a wireless technology standard used for exchanging data over short distances, typically between fixed and mobile devices. It operates on the 2.4 to 2.485 GHz frequency band and uses radio waves for communication. It allows devices to connect and communicate with each other without the need for cables. This flexibility and cost-effectiveness make Bluetooth an ideal solution for building IoT ecosystems, enabling seamless connectivity, data exchange, and control of interconnected devices across various domains [12].

BLE stands for Bluetooth Low Energy. It's specifically designed for low-power consumption and short-range data transfer. The range of approximately 100 meters offered by BLE surpasses that of classic Bluetooth by tenfold, while its latency is 15 times shorter. BLE can operate within a transmission power range of 0.01 mW to 10 mW. These characteristics position BLE as a promising technology for IoT applications. The rapid development of the BLE standard by smartphone manufacturers has resulted in its widespread availabil-

ity across most smartphone models. Additionally, the feasibility of this standard has been demonstrated in various applications, including vehicle-to-vehicle communication and wireless sensor networks [13].

Koko Friansa et al. designed a BLE-based wireless sensor network to monitor a system of solar-powered street lamps. It is characterized by its low cost and open platform technology. The proposed WSN utilizes both tree and linear topologies for data transmission. Since BLE technology supports a 2-chain structure with 5 nodes in each one, the system totals 10 nodes, including the master node. Data is sent from one node to another until it reaches the master node, as depicted in the figure below. It is then transmitted to the user's computer using an electric vehicle as an internet gateway [14].

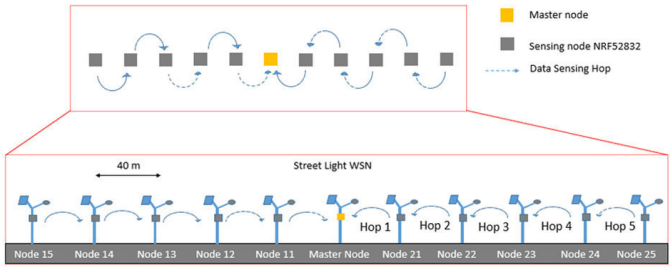


Fig. 10. Proposed WSN topology for battery and PV monitoring [14].

3) **ZigBee**

Zigbee technology has become renowned for its wireless sensor network solutions. It's built upon the IEEE 802.15.4 standard, which defines the physical and MAC layers for low-rate, low-power wireless networks. The Zigbee Alliance, a consortium of companies, develops standards and products for the network layer and offers a framework for the application layer [15].

Farihah et al, developed an online monitoring system for grid-tied photovoltaic system. They proposed a model with Zigbee wireless communication to overcome the limitations of traditional data transmission cables and their additional technical costs. A point-to-point network architecture was employed due to its simplicity of implementation. As shown in the figure below 11 this architecture comprises two logging points: one positioned outdoors to gather meteorological data and another situated indoors, connected to the inverter to collect electrical data and a main base where the PC host is located.

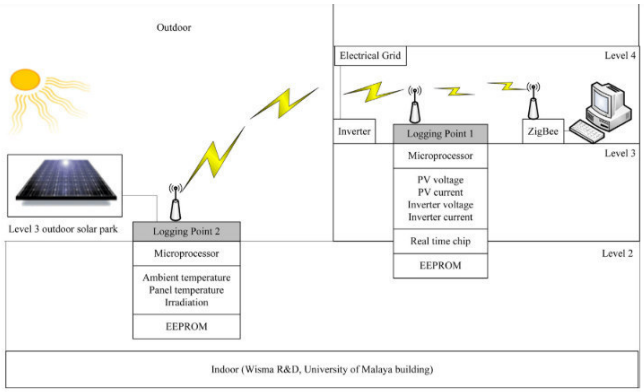


Fig. 11. Overall setup of the grid-connected photovoltaic system [16].

The research details the hardware and software design, and a custom web application for remote data access. The Zigbee technology was chosen for its cost-effectiveness, robustness, and low power consumption, with a transmission range of at least 100 meters. The system was successfully implemented and validated on a 1.25 kWp grid-connected PV system, demonstrating reliable data transmission and effective online monitoring capabilities. The user-friendly web application enabled easy remote access to the monitored data, facilitating prompt evaluation of the system's performance. The study concludes that the Zigbee-based wireless monitoring system is a viable and cost-effective solution for grid-connected PV systems, offering significant advantages over conventional wired systems. Future work is recommended to enhance the system further and explore its application in other renewable energy monitoring scenarios [16].

4) **LoRa/LoRaWAN**

LoRa, which stands for Long Range, is a wireless communication technology designed for long-distance, low-power communication between devices in the Internet of Things (IoT) ecosystem. Operating in unlicensed bands below 1 GHz, LoRa enables reliable communication links over considerable distances. It employs spread spectrum modulation techniques to achieve this, allowing for efficient data transmission while minimizing power consumption [17].

LoRaWAN, standardized by the LoRa Alliance, is a Low Power Wide Area Network (LPWAN) protocol designed to establish the communication protocol and system architecture, whereas LoRa defines the physical layer. It offers different classes of nodes to balance network downlink communication latency with battery lifetime for diverse applications. LoRaWAN uses star-of-star network topology as illustrated in figure 12 [18] It consists of end-devices, gateways, network servers, and application servers. Gateways relay messages between end-devices and the central network server, with end-devices employing single-hop wireless communication

to interact with gateways [19].

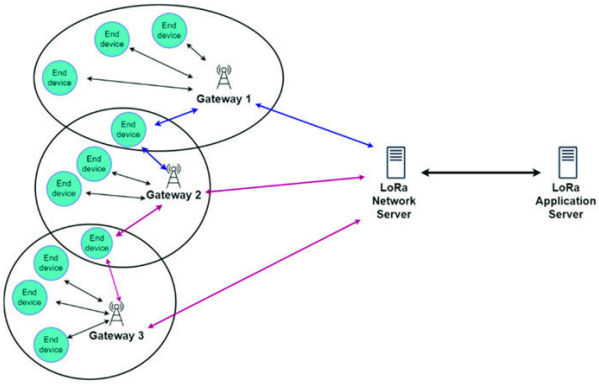


Fig. 12. Star-of-stars LoRaWAN topology [18].

Paredes-Parra et al,proposed a wireless solution that utilizes long-range (LoRa) technology to communicate with remote PV power plants, covering extensive distances while minimizing both power consumption and maintenance efforts. This approach combines low-cost, open-source technology at the sensor layer with a low-power wireless area network (LPWAN) at the communication layer. The proposed network architecture follows typical LoRa network topologies, such as a star configuration comprising end devices, gateways, and a central network server. Sensors selected for monitoring PV installations must adhere to IEC-61724 requirements. They gather PV electrical data and weather parameters to estimate PV modules performance and the operating conditions. These data packets are transmitted from the respective PV power plants to the LoRa gateway and subsequently relayed to the network application via LAN connection. [20]

IV. Main Challenges

Monitoring photovoltaic (PV) systems presents several significant challenges that impact their efficiency, security, range, data processing capability, and overall performance [21].

- Data handling: One of the primary challenge, particularly with the increasing size of utility-scale solar power plants. The sheer volume of data generated overwhelms conventional technologies and data processing modules like Arduino, which, while cost-effective, lack the computational power for complex data acquisition and analysis. Similarly, PLC modules struggle with handling large datasets efficiently. There is a pressing need for more powerful data processing platforms capable of real-time data acquisition and sophisticated analysis.
- Security: is another critical issue in PV monitoring systems, especially in wireless schemes. Data transmitted from sensor nodes to central stations is vulnerable to cyber-attacks and physical tampering if weak security protocols are in place. Ensuring end-to-end encryption, authentication for legitimate users, and protection against unauthorized access are essential measures to safeguard data integrity and system functionality. However, various transmission modules such as ZigBee, Wi-Fi, Bluetooth, and LoRa face unique security challenges, highlighting the need for robust security protocols across all layers of communication.
- Signal interference: poses another significant challenge, affecting data monitoring in solar PV systems. Interference from other communication modes can lead to poor module performance, slow data transfer rates, weak signal strength, and intermittent connections. This interference is evident in technologies like GSM, where signal quality directly impacts service reliability and revenue loss. Even platforms like Raspberry Pi can experience interference from HDMI cables, disrupting Wi-Fi signals and compromising system stability.
- Energy efficiency: remains a critical concern due to the multiple transmitter nodes within PV monitoring systems. The longevity of node batteries directly affects network lifespan and real-time communication capabilities. Developing energy-efficient communication protocols and utilizing multiple transmission paths are crucial strategies to enhance system performance and reliability over time.
- Data transmission range: is a critical issue for monitoring large-scale PV plants. Traditional modules such as Bluetooth, Wi-Fi, and ZigBee have limited ranges (100m to 300m), insufficient for monitoring PV systems spread over kilometers. LoRa, designed for long-range communication, suffers from lower transfer rates and increased Packet Error Rates (PER) with distance, necessitating careful consideration in deployment strategies.
- Environmental impacts: further complicate PV monitoring, with factors like temperature, humidity, irradiance, and dust affecting the performance of data processing and transmission modules. Dust accumulation, for example, can reduce solar panel efficiency, impacting overall PV output. Careful environmental assessment and system placement are essential to mitigate these effects and ensure reliable long-term operation. [21]

V. Conclusion

The research provided a thorough examination of several solar PV monitoring solutions based on the use of different data processing modules and communication protocols. Providing essential information about the hardware and software used in PV monitoring systems.

This review highlights the state-of-the-art developments aimed at supporting sustainable development in the renewable energy sector. Specifically, the advancement of solar PV monitoring systems offers valuable guidance for the industry, encouraging further exploration of IoT-based monitoring systems for large-scale solar PV applications. This review also aids in selecting suitable monitoring technologies to enhance the efficiency, accuracy, and reliability of solar PV systems, fostering ongoing research in this critical area.

REFERENCES

[1] J. M. Paredes-Parra, A. Mateo-Aroca, G. Silvente-Niñirola, M. C. Bueso, and Molina-García, "Pv module monitoring system based on low-cost solutions: Wireless raspberry application and assessment," *Energies*, vol. 11, no. 11, 2018.

[2] N. A. A. Jamil, S. A. Jumaat, S. Salimin, M. N. Abdullah, and A. F. M. Nor, "Performance enhancement of solar powered floating photovoltaic system using arduino approach," *International Journal of Power Electronics and Drive Systems*, vol. 11, pp. 651–657, 2020.

[3] I. Allafi and T. Iqbal, "Low-cost scada system using arduino and reliance scada for a stand-alone photovoltaic system," *Journal of Solar Energy*, vol. 2018, no. 1.

[4] J. W. Jolles, "Broad-scale applications of the raspberry pi: A review and guide for biologists," *Methods in Ecology and Evolution*, vol. 12, no. 9, pp. 1562–1579.

[5] Y. Bikrat, K. Salmi, A. Benlghazi, A. Benali, and D. Moussaid, "A photovoltaic wireless monitoring system," in *2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, pp. 1–5, 2018.

[6] R. I. Pereira, I. M. Dupont, P. C. Carvalho, and S. C. Jucá, "Iot embedded linux system based on raspberry pi applied to real-time cloud monitoring of a decentralized photovoltaic plant," *Measurement*, vol. 114, pp. 286–297, 2018.

[7] Y. Kabalci and E. Kabalci, "Modeling and analysis of a smart grid monitoring system for renewable energy sources," *Solar Energy*, vol. 153, pp. 262–275, 2017.

[8] J. Han, I. Lee, and S.-H. Kim, "User-friendly monitoring system for residential pv system based on low-cost power line communication," *IEEE Transactions on Consumer Electronics*, vol. 61, no. 2, pp. 175–180, 2015.

[9] M. Habaebi, Q. Ashraf, A. Azman, and M. Islam, "Development of wi-fi based home energy monitoring system for green internet of things," vol. 14, pp. 249–256, 01 2016.

[10] Y. S. Parihar, "Internet of things and nodemcu a review of use of nodemcu esp8266 in iot products," vol. 6, p. 1085, 06 2019.

[11] T. Sutikno, H. S. Purnama, A. Pamungkas, A. Fadlil, I. Alsofyani, and M. Jopri, "Internet of things-based photovoltaics parameter monitoring system using nodemcu esp8266," *IJECE*, vol. 11, pp. 5578–5587, December 2021.

[12] J. Zhang, "The application of bluetooth technology in the internet of things," *ACE*, vol. 12, pp. 177–183, 2023.

[13] P. Schneider and F. Xhafa, "Chapter 7 - IoT, edge, cloud architecture and communication protocols: Cloud digital ecosystem and protocols. anomaly detection and complex event processing over IoT data streams," in *Anomaly Detection and Complex Event Processing over IoT Data Streams*, pp. 129–148, Elsevier, 2022.

[14] K. Friansa, A. Iqbal, and M. Rozana, "Design bluetooth low energy for battery and photovoltaic monitoring at solar powered street lamp system," *IOP Conference Series: Earth and Environmental Science*, vol. 537, p. 012028, jul 2020.

[15] A. Zohourian, S. Dadkhah, E. C. P. Neto, H. Mahdikhani, P. K. Danso, H. Molyneaux, and A. A. Ghorbani, "Iot zigbee device security: A comprehensive review," *Internet of Things*, vol. 22, p. 100791, 2023.

[16] F. Shariff, N. A. Rahim, and W. P. Hew, "Zigbee-based data acquisition system for online monitoring of grid-connected photovoltaic system," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1730–1742, 2015.

[17] R. S. Sinha, Y. Wei, and S.-H. Hwang, "A survey on lpwa technology: Lora and nb-iot," *ICT Express*, vol. 3, no. 1, pp. 14–21, 2017.

[18] K. Banti, I. Karampelia, T. Dimakis, A.-A. Boulogeorgos, T. Kyriakidis, and M. Louta, "Lorawan communication protocols: A comprehensive survey under an energy efficiency perspective," *Telecom*, vol. 3, pp. 322–357, 2022.

[19] S.-Y. Wang, Y.-R. Chen, T.-Y. Chen, C.-H. Chang, Y.-H. Cheng, C.-C. Hsu, and Y.-B. Lin, "Performance of lora-based iot applications on campus," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pp. 1–6, 2017.

[20] J. Paredes-Parra, A. García-Sánchez, A. Mateo-Aroca, and A. Molina-García, "An alternative internet-of-things solution based on lora for pv power plants: Data monitoring and management," *Energies*, vol. 12, p. 881, 2019.

[21] S. Ansari, A. Ayob, M. S. H. Lipu, M. H. M. Saad, and A. Hussain, "A review of monitoring technologies for solar pv systems using data processing modules and transmission protocols: Progress, challenges and prospects," *Sustainability*, vol. 13, no. 15, 2021.

Recent Developments in Photovoltaic Energy: Challenges and Opportunities in Achieving Sustainability

Aicha Djalab
Electrical Systems Engineering Department
Faculty of Technology, University M'Hamed Bougara of Boumerdes, Algeria
a.djalab@univ-boumerdes.dz

Belgacem Mecheri
Automation and Electrification of Systems Department
Faculty of Hydrocarbons and Chemistry, University M'Hamed Bougara of Boumerdes, Algeria
mecheri.belgacem@univ-boumerdes.dz

Kacimi Nora
Electrical Systems Engineering Department
Faculty of Technology, University M'Hamed Bougara of Boumerdes, Algeria
n.kacimi@univ-boumerdes.dz

***Abstract*— Photovoltaic energy plays a pivotal role in the global shift toward renewable energy, utilizing sunlight to generate clean and sustainable electricity. This paper focuses on recent advancements in photovoltaic technologies, including floating systems and bifacial panels, new storage technologies, and AI-driven optimization, these innovations are driving significant advancements in efficiency, scalability, and adaptability, reinforcing their critical role in sustainable development. The study concludes with actionable recommendations to accelerate the PV sector's growth and maximize its impact on achieving global sustainability targets, ensuring a cleaner, more resilient energy future.**

Keywords—photovoltaic energy, renewable energy, floating systems, bifacial panels, sustainable development.

I. INTRODUCTION

Humanity today faces major environmental and economic challenges, including climate change and the depletion of natural resources. In this context, renewable energy is considered the ideal solution to achieve a balance between meeting the growing demand for energy and protecting the environment. Solar energy is recognized as the most abundant and environmentally friendly source of renewable energy, offering a pathway to a sustainable energy future. By harnessing the abundant daily output of solar power, solar energy has emerged as one of the fastest-growing and most innovative renewable energy sources, with the installed capacity of photovoltaic systems doubling globally over the past decade [1, 2].

As of 2022, solar energy has maintained its position as the leading force in capacity expansion, witnessing a notable increase of 192 GW (+22%). In 2023, the worldwide photovoltaic (PV) cumulative capacity surged to 1.6 TW, marking a significant increase from the 1.2 TW recorded in 2022 (Figure 1). This growth was fueled by the commissioning of new PV systems, which rose from 407.3 GW to 446 GW [3]. This impressive growth trajectory suggests that within a short span of time, solar energy is on track to become the linchpin of the world's sustainable energy framework.

The significance of photovoltaic energy lies in its flexibility and adaptability to various environments, from large-scale power plants to residential and industrial applications. With continuous technological advancements, these systems have become more efficient and cost-effective, making them accessible to both developing and developed countries.

The purpose of this paper is to shed light on recent innovations that have the potential to significantly enhance the efficiency of photovoltaic energy systems. By exploring advancements such as floating systems, bifacial panels, new storage technologies, and AI-driven optimization, the study emphasizes how these technologies are shaping the future of solar energy. Additionally, it aims to address the key challenges impeding the broader adoption of photovoltaic systems, including high initial costs, reliance on rare materials, and the environmental concerns associated with end-of-life panel recycling. Finally, the paper provides actionable recommendations to improve the performance, cost-effectiveness, and sustainability of photovoltaic systems, ensuring their pivotal role in achieving global renewable energy goals.

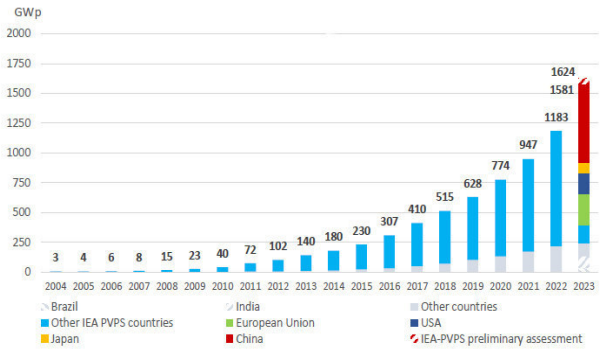


Fig. 1. Evolution of cumulative PV installations [3].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

II. Recent Innovations in Photovoltaic Energy

A. *Floating Systems on Water Surfaces*

With the increasing scarcity of suitable land for solar panel installation, floating systems have emerged as an innovative and effective solution. These systems offer higher efficiency due to the natural cooling effect of water, and they also help minimize water evaporation from reservoirs, (Figure 2 shows a 40 MW FPV power plant in Huainan, China.) [4]. The global installed capacity of floating solar systems reached approximately 4.6 gigawatts in 2022. Deloitte (2022) estimates that this capacity could potentially grow to 13 GWp by 2025. Under a medium growth scenario, it is projected to exceed 20 GWp by 2030 [5].

The primary advantages of floating systems can be outlined as follows:

- Augmented Efficiency: Deploying photovoltaic panels on water surfaces benefits from the cooling effect through evaporation, resulting in lower operating temperatures and more efficient energy output (up to 12%).
- Land Conservation: FPV systems eliminate the need for land, transforming non-commercial water surfaces into economically viable photovoltaic solar power plants, thereby preserving valuable land for agriculture, industry, and tourism.
- Integration with Hydropower Plants: Pairing FPV plants with existing hydropower facilities enhances electricity generation during peak periods and serves as an energy storage solution, utilizing solar energy during the day and hydropower during the night.
- Water Conservation: Covering floating panels minimizes solar energy reaching the water, preventing evaporation and aiding in saving thousands of cubic meters of water annually, particularly in dams, ponds, and reservoirs.
- Water Quality Enhancement: Partially covering basins reduces light penetration, enhancing water quality by preventing biological fouling and algae proliferation, thus reducing water treatment costs.



Fig. 2. The largest FPV plant (40 MWp), Huainan, China [4].

B. *Bifacial Panels*

Bifacial solar panels have demonstrated significant efficiency improvements due to their ability to capture light from both the front and rear sides, thereby increasing energy output. This enhancement is particularly effective in environments with high reflectivity, such as snow-covered or sandy areas. The adoption of bifacial photovoltaic (PV) technology, particularly crystalline silicon (c-Si), is projected to grow significantly, with an expected increase of over 35% in deployment by 2028 [6]. This growth underscores the potential of bifacial panels to become a key player in the renewable energy sector, especially as advancements continue to improve their performance and cost-efficiency.

Bifacial technology is anticipated to become increasingly popular across both commercial and residential solar projects. Figure 3 shows an example of the vertically mounted BPV systems.



Fig. 3. Example of a vertically mounted BPV system on a green roof in Winterthur, Switzerland [7].

C. *Advanced Storage Technologies*

Energy storage systems, particularly batteries, are essential for supporting renewable energy and reducing carbon emissions. They solve the problem of inconsistent energy from sources like solar and wind by storing excess energy for later use. This helps stabilize electricity grids, manage peak demand, and integrate small renewable systems into the network.

New storage technologies, such as lithium-ion batteries and solid-state batteries, have significantly improved the utilization of solar energy by storing surplus power for use during periods of low solar irradiation. Advances in battery energy density have focused on creating lighter, more efficient, and safer designs for various applications. These innovations are pivotal for enhancing energy reliability, especially in off-grid locations [8].

D. *Artificial Intelligence and Data Analytics*

The utilization of artificial intelligence (AI) in analyzing the performance of solar systems and enhancing their efficiency has become increasingly prevalent, where

AI can forecast system failures and optimize the positioning of solar panels to harness the maximum amount of energy. Companies and consumers can benefit from running solar systems more efficiently with artificial intelligence, thereby reducing costs associated with unplanned maintenance and repairs.

III. Discussion

Recent advancements in photovoltaic (PV) technologies, such as floating solar systems, bifacial panels, advanced batteries for energy storage, and AI-driven optimization, are revolutionizing the renewable energy sector. These innovations enhance the efficiency and adaptability of PV systems, enabling them to perform effectively in diverse environments and address global energy demands. Key findings include:

- **Increased Efficiency:** Technologies such as bifacial panels have demonstrated a 20–30% improvement in energy output, particularly in regions with high ground reflectivity, like deserts or snowy areas.
- **Land Optimization:** Floating PV systems have emerged as an effective solution for energy generation in land-constrained areas, with additional environmental benefits like reduced water evaporation.
- **Enhanced Reliability:** AI applications in PV systems have significantly reduced downtime by predicting faults and enabling real-time performance optimization.
- **Energy Storage Advances:** Lithium-ion and solid-state batteries have improved the feasibility of using PV energy as a consistent power source by addressing the intermittency of solar energy.

IV. Challenges

Energy storage systems, particularly batteries, are essential for supporting renewable energy and reducing carbon emissions. They solve the problem of inconsistent energy from sources like solar and wind by storing excess energy for later use. This helps stabilize electricity grids, manage peak demand, and integrate small renewable systems into the network.

Despite these advancements, significant challenges remain:

- **Economic Barriers:** While PV system costs have decreased by over 80% in the last decade, initial capital investment and storage costs remain prohibitive for many regions.
- **Material Dependence:** The reliance on rare and non-renewable materials, such as silver and specific semiconductors, poses long-term sustainability issues.
- **End-of-Life Management:** As PV installations age, managing and recycling outdated panels is an emerging environmental concern.
- **Grid Integration:** High penetration of PV energy can destabilize grids in the absence of adequate storage and management infrastructure.

V. Conclusion

Photovoltaic (PV) energy production stands as a clean, and renewable, energy source that has gained critical prominence in the global energy landscape. This paper spotlighting recent innovations such as floating solar systems, bifacial panels, advanced energy storage, and AI-driven optimization. Together, these developments underscore the critical role of photovoltaic technologies in achieving a cleaner and more reliable energy future.

Despite these opportunities, challenges persist, in order to advance the sustainability of solar energy systems, it is recommended to invest in research and development by focusing on the development of alternative materials to lessen reliance on scarce resources and promoting advancements in recycling technologies for sustainable end-of-life management of PV systems.

References

[1] REN21, 2019. Renewables 2019 Global Status Report, Renewable Energy Policy Network for the 21st Century.

[2] El Hammoumi, A., Chalh, A., Allouhi, A., Motahhir, S., El Ghzizal, A., & Derouich, A. (2021). Design and construction of a test bench to investigate the potential of floating PV systems. Journal of Cleaner Production, 278, 123917.

[3] Report IEA-PVPS T1-42:2024. Strategic PV Analysis and Outreach. Snapshot of Global PV Markets, April 2024.

[4] A. Djalab, Z. Djalab, A. El Hammoumi, G.M. Tina, S. Motahhir, A A. Laouid. "A comprehensive Review of Floating Photovoltaic Systems: Tech Advances, Marine Environmental Influences on Offshore PV Systems, and Economic Feasibility Analysis". Solar Energy 277 (2024) 112711.

[5] The future of floating solar: Drivers and barriers to growth Arnhem, The Netherlands, Accessed, May 26, DNV(2023) .

[6] M. Aghaei, M. Korevaar, P. Babal, H. Ziar "Bifacial Photovoltaic Technology: Recent Advancements, Simulation and Performance Measurement". Chapter Intechopen. June 2022.

[7] A. Garrod, A. Ghosh. "A review of bifacial solar photovoltaic applications". Front. Energy. November 2023. https://doi.org/10.1007/s11708-023-0903-7

[8] G. G. Njema , O . R B. Ouma ,J K. Kibet. "A Review on the Recent Advances in Battery Development and Energy Storage Technologies". Hindawi Journal of Renewable Energy. Volume 2024. https://doi.org/10.1155/2024/2329261

Random Source Coupling to Shielded Cable

Hemza Gueddar
Electrical Engineering Department, University of Jijel
Jijel, Algeria
hamza01gueddar@gmail.com

Sid-Ahmed Touil
Institute of Electrical and Electronic Engineering
University M'hamed Bougara.
Boumerdes, Algeria
s.touil@univ-boumerdes.dz

Nidhal Cherrat
Dept. of Electrical Systems Engineering
ISE University of Boumerdes
Laboratoire d'Automatique de Jijel-LAJ, University MSBY
Boumerdes, Algeria
n.cherrat@univ-boumerdes.dz

***Abstract*— the main purpose of our work is to give a global analysis of the electromagnetic coupling problem in a shielded cable. We will go farther than the deterministic approach which is a classic way of solving this problem towards a probabilistic approach, where the excitation is considered as a random variable. In this work, we use the Chaos Polynomial technique to calculate the probability density function of induced voltages. Where the efficiency and rapidity of this technique are dominated comparing to Monte Carlo technique that takes a lot of time.**

Keywords— ***shielded cable; transmission line; random; Chaos polynomial; Monte-Carlo***

I. INTRODUCTION

The cables of electrical network can be considered as huge antennas that receive radiation from various electromagnetic sources of natural or artificial origins. When the unwanted signals are routed to the lines ends, they affect the proper functioning of any equipment which is connected to them. An electromagnetic disturbance that illuminates a shielded multi-strand electrical cable induces common mode voltages between an internal conductor or a set of internal conductors, the ground reference (shielding) [1]-[2], and a differential mode voltage between the internal conductors providing input and output exit of any equipment if this electromagnetic disturbance is isolated with a high amplitude pulse such as the one which is produced by lightning or a high altitude nuclear explosion [3]-[4]. The common mode voltage is the most critical for the protection of the electronic system because it presents a high risk of failure at the input of the components of all equipment connected to the cable [5]-[6]. This study is to build an analysis methodology for stochastic source coupling to shielded cable, which takes into account the random nature of the environment in which the cable is located. Probabilistic tools and statistics are good candidates to accomplish this task [7]-[8]. They are used to predict the voltage level induced in the ends of the cable. The computational inefficiency of traditional tools implemented in many circuit simulators, such as Monte Carlo (MC) [9]-[10], has generated a great deal of interest in stochastic techniques for circuit modeling and, in particular, in the design of electrical network. Based on the Polynomial Chaos (PC) those have been proposed for statistical efficiency analysis of transmission line circuits. PC seeks a responses approximation of the stochastic transmission line in terms of an expansion in orthogonal polynomials [11]. The expansion coefficients can be calculated using different approaches. In this work we propose the regression method for calculating the polynomial coefficients of chaos, directly supplying relevant statistical information such as value and variance. In addition, the PC Extension (PCE) can be more generally used as a cheap computer [12], but still macro accurate model to extract other statistical properties like higher order statistical moments(for example, asymmetry and flattening) or distribution functions.

II. EXCITATION SOURCE COUPLING TO SHIELDED CABLE BY DETERMINISTIC APPROACH

We consider a shielded cable located at height h above a ground of finite conductivity and excited by a voltage generator as shown in Fig.1. Where the induced voltages and currents along the cable can be calculated using the formalism of a multi-conductor line, which is expressed in frequency as following [1]:

$$\begin{cases} \dfrac{d}{dx}[U(x)]+[Z][I(x)]=[0] \\ \dfrac{d}{dx}[I(x)]+[Y][U(x)]=[0] \end{cases} \quad (1)$$

Where: x is space variables, [Z] and [Y] are successively the impedance and admittance matrices for the shielded cable that had been deduced by Ametani [5]. By taking into account the relation developed in the Appendix 1 $[\overline{Z}]+jw[L]$.The equation (1) will be given as:

$$\frac{d}{dx}[U(x)]+jw[L][I(x)]+[\overline{Z}][I(x)]=[0] \quad (2)$$

Where $[L]$ and $[\overline{Z}]$ are defined in the Appendix 1

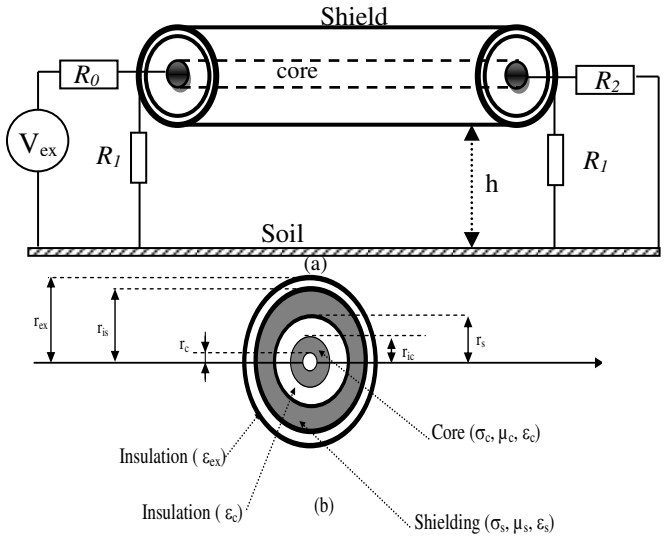


Fig.1. (a) Shielded cable over ground plane with excitation (b) Cross-section of a Shielded cable

A. Temporal Coupling Equations

The coupling equations which are given above (1) can be converted into time domain by the following expressions [4]:

$$\frac{\partial}{\partial x}\left[u(x,t)\right]+\left[L\right]\frac{\partial}{\partial t}\left[i(x,t)\right]+\int_0^t\left[\zeta_s(t-\tau)\right]\times\frac{\partial}{\partial\tau}\left[i(x,\tau)\right]d\tau + \int_0^t\left[\zeta_{sh}(t-\tau)\right]\times\frac{\partial}{\partial\tau}\left[i(x,\tau)\right]d\tau=\left[0\right] \quad (3)$$

$$\frac{\partial}{\partial x}\left[i(x,t)\right]+\left[C\right]\frac{\partial}{\partial t}\left[u(x,t)\right]=\left[0\right] \quad (4)$$

Where:

$\left[\zeta_s\right]$: Resistance matrix due to the finite ground conductivity;

$\left[\zeta_{sh}\right]$: Resistance matrix due to finite shielding conductivity [4].

Where

$$\begin{cases}\left[\zeta_s\right]=\begin{bmatrix}\zeta_{sii}(t) & \zeta_{sij}(t)\\ \zeta_{sji}(t) & \zeta_{sii}(t)\end{bmatrix}\\ \left[\zeta_{sh}\right]=\begin{bmatrix}2\zeta_0(t)-2\zeta_m(t) & \zeta_0(t)-\zeta_m(t)\\ \zeta_0(t)-\zeta_m(t) & \zeta_0(t)\end{bmatrix}\end{cases} \quad (5)$$

$$\zeta_{sii}(t)\cong\min\left\{\frac{1}{2\pi h}\sqrt{\frac{\mu_0}{\varepsilon_0\varepsilon_{rs}}},\frac{\mu_0}{\pi\tau_{sii}}\times\begin{bmatrix}\frac{1}{2\sqrt{\pi}}\sqrt{\frac{\tau_{sii}}{t}}+\frac{1}{4}\exp\left(\tau_{sii}/t\right)\times\\ erfc\left(\sqrt{\frac{\tau_{sii}}{t}}\right)-\frac{1}{4}\end{bmatrix}\right\} \quad (a5)$$

Where: $\tau_{sii}=h^2\mu_0\sigma_{sii}$ and *erfc* the complimentary error function defines as following

$$erfc(x)\cong\frac{2}{\pi}\int_\infty^\infty\exp\left(-\tau^2\right)dt=1-\frac{2\exp\left(-x^2\right)}{\pi}\sum_{n=0}^{\infty}a_n x^{2n+1} \quad (b.5)$$

B. The Coupling Equations Solution

In frequency, the coupling equations solution by the chain transition matrix is well detailed in [1]. In temporal, it is resolved by using the so-called numerical Finite Difference Time Domain (FDTD) [2]-[4]. Using the discrete notations below:

$$\left[u_k^n\right]=\left[u\left[(k-1)\Delta x,n\Delta t\right]\right] \quad (6)$$

$$\left[i_k^n\right]=\left[i\left[\left(k-\frac{1}{2}\right)\right]\Delta x,\left(n+\frac{1}{2}\right)\Delta t\right] \quad (7)$$

Where the two indices "*n*" and "*k*" indicate respectively [6] time and space. The spatiotemporal discretization alternates a voltage and current nodes which are separated by $\Delta x/2$ in space and $\Delta t/2$ in time; where both nodes ends are voltage nodes. The coupling equations (3) and (4) turn into:

$$\left(\frac{[L]}{\Delta t}+\frac{3}{2}[\zeta_s(\Delta t)]+\frac{1}{2}[\zeta_{sh}(\Delta t)]+\frac{IMAT}{2\Delta t}\right)\left[i_k^{n+1}\right]=\left(\frac{[L]}{\Delta t}+\frac{1}{2}[\zeta_s(\Delta t)]+\frac{1}{2}[\zeta_{sh}(\Delta t)]\right)\left[i_k^n\right] + \left([\zeta_s(\Delta t)]+\frac{IMAT}{2\Delta t}\right)\left[i_k^{n-1}\right]-\frac{\left[u_{k+1}^{n+1}\right]-\left[u_k^{n+1}\right]}{\Delta x}-\frac{1}{2}\sum_{j=0}^{n-1}\left\{\left[\zeta_s((n-j)\Delta t)\right]+\left[\zeta_{sh}((n-j)\Delta t)\right] + \left[\zeta_s((n+1-j)\Delta t)\right]+\left[\zeta_{sh}((n+1-j)\Delta t)\right]\right]\times\left(\left[i_k^{j+1}\right]-\left[i_k^j\right]\right)\right\} \quad (8)$$

For *k=1,2…,kmax*

kmax: the number of space steps

$$\left[\frac{[C]}{\Delta t}\right]\left[u_k^{n+1}\right]=\left\{\left[\frac{[C]}{\Delta t}\right]\left[u_k^n\right]-\frac{\left(\left[i_k^n\right]-\left[i_{k-1}^n\right]\right)}{\Delta x}\right\} \quad (9)$$

For*k=2,3,…, kmax*

"*IMAT*" is the following matrix:

$$IMAT=\begin{bmatrix}2\int_0^{\Delta t}\zeta_0(t)dt-2\int_0^{\Delta t}\zeta_m(t)dt & \int_0^{\Delta t}\zeta_0(t)dt-\int_0^{\Delta t}\zeta_m(t)dt\\ \int_0^{\Delta t}\zeta_0(t)dt-\int_0^{\Delta t}\zeta_m(t)dt & \int_0^{\Delta t}\zeta_0(t)dt\end{bmatrix} =\begin{bmatrix}2In_0-2In_m & In_0-In_m\\ In_o-In_m & In_0\end{bmatrix} \quad (10)$$

Both integrals In_0 and In_m are performed after an approximation of the transient impedances modified clean and mutual $\zeta_0(t)$ and $\zeta_m(t)$ to lift the singularity at t = 0 from the proper transient impedance $\zeta_0(t)$, the latter is approximated by a finite sum as follows

$$\zeta_0(t)=K_\delta.\delta(t)+R_{dc}+\sum_{k=1}^{N_0}\zeta_{0K}(t) \quad (11)$$

Where:

N_0: finite number of the sum terms, $\delta(t)$ is Dirac function and K_δ is a constant given in:

$$K_\delta=\frac{2\mu_b d}{\pi^2}\sum_{k=N_{0+1}}^{\infty}\frac{1}{K^2}=\frac{2\mu_b d}{\pi^2}\left(\frac{\pi^2}{6}-\sum_{k=N_{0+1}}^{\infty}\frac{1}{K^2}\right) \quad (12)$$

The modified transient mutual impedance is also approximated by a finite sum:

$$\zeta_m(t)\cong\begin{cases}0 & for\ t=0\\ R_{dc}+\sum_{k=1}^{Nm}\zeta_{mk}(t) & for\ t>0\end{cases} \quad (13)$$

In this case, N_m only terms of the sum are considered [6]. By using the approximate expressions (12) and (13) modified transient impedances, we compute analytically the two integrals "Ini_0" and "Ini_m", we are getting

$$In_0=K_\delta+R_{dc}\Delta t+\sum_{k=1}^{N_0}2R_{dc}\left(-\frac{\tau_{sh}}{k^2\pi^2}\right)\left[e^{-k^2\pi^2\frac{\Delta t}{\tau_s}}-1\right] \quad (14)$$

$$In_m=R_{dc}\Delta t+\sum_{k=1}^{N_m}2R_{dc}(-1)^k\left(-\frac{\tau_{sh}}{k^2\pi^2}\right)\left[e^{-k^2\pi^2\frac{\Delta t}{\tau_s}}-1\right] \quad (15)$$

At both ends of the cable, and for resistive terminations [R1] and [R2], replace "k" with "1"; then by "kmax + 1"equation (9).

For k = 1

$$\left(\frac{[C]}{\Delta t}\right).\left[u^{n+1}(0)\right]-\frac{\left[i^{n+1}(0)\right]}{\Delta x}=\left(\frac{[C]}{\Delta t}\right).\left[u^{n}(0)\right]-\frac{\left[i_1^n\right]}{\Delta x/2}+\frac{\left[i^n(0)\right]}{\Delta x} \quad (16)$$

For k=kmax+1

$$\left(\frac{[C]}{\Delta t}\right).\left[u^{n+1}(\ell)\right]-\frac{\left[i^{n+1}(\ell)\right]}{\Delta x}=\left(\frac{[C]}{\Delta t}\right).\left[u^{n}(\ell)\right]-\frac{\left[i_{k\max}^n\right]}{\Delta x/2}+\frac{\left[i^n(\ell)\right]}{\Delta x} \quad (17)$$

The stability criterion requires that the following condition be met:

$$\Delta t \le \Delta x/\upsilon p \quad (18)$$

Where: υp is the wave propagation speed.

III. EXCITATION SOURCE COUPLING TO SHIELDED CABLE BY STOCHASTIC APPROACH

The underlying idea of PC is instead to represent stochastic unknowns (in this case, voltages and currents) as PCE [7].

$$U(t,\xi)\cong u_0(t)\phi_0(\xi)+u_1(t)\phi_1(\xi)+u_2(t)\phi_2(\xi)+... \quad (19)$$

Where $u_0(t)$, $u_1(t)$, $u_2(t)$,… are the yet-to-be-determined PC coefficients and ϕ_0 , ϕ_1 , ϕ_2 ,.. are polynomials of increasing degree in the standardized variable ξ . They are orthonormal according to the inner product defined as

$$\langle\phi_m,\phi_n\rangle=\int_{-\infty}^{+\infty}\phi_m(\xi)\phi_n(\xi)W(\xi)d\xi=\delta_{mn} \quad (20)$$

meaning that the above integral is non-zero only when m=n. The function $W(\xi)$ appearing in the integrand is the probability density function (PDF) , which is defined by the statistical model ascribed to the random parameters in the problem. For the sake of illustration, this article focuses on the common case of Gaussian variability [8]-[9], for which the above PDF. The corresponding orthonormal polynomials that satisfy (19) are the normalized Hermite polynomials, the first three being.

$$\begin{cases}\phi_0=1\\ \phi_1=\xi\\ \phi_2=\left(\xi^2-1\right)\sqrt{2}\end{cases} \quad (21)$$

In this section, we use the regression technique to calculate the chaos polynomial coefficients by applying it to the modeling of random excitation source variations in multi-conductor transmission lines [10]-[11]. The frequency domain of the transmission line equation is written as:

$$\begin{cases}\dfrac{d}{dx}[U(x,\xi)]=-[Z(\xi)][I(x,\xi)]\\ \dfrac{d}{dx}[I(x,\xi)]=-[Y(\xi)][U(x,\xi)]\end{cases} \quad (22)$$

Where $\xi=[\xi_1,...,\xi_n]$ is a multivariate set of n independent random variables that affect the excitation source [11]. So the voltages *U* and the currents *I* are random and dependent on ξ In short, the idea here is to approximate the dependence on the random variables ξ which means the expansion in a truncated series of proper multivariate orthogonal polynomials as following

$$X(..,\xi)\approx\sum_{j=0}^{p-1}X_j(...)\phi_j(\xi) \quad (23)$$

Where X (…) stands for $\tilde{U}(x,\xi),\tilde{I}(x,\xi)$ and they X_j are the respective coefficients. The series has a finite number of terms P-1, while $\{\phi_j\}$ defines an orthogonal polynomial basis, whose choice relies on the distribution of the random variables ξ_i . So using Hermit polynomials is the optimal choice for the Gaussian [12]

After solution equation in (22) and the application of the regression approach (23) which is written as

$$\begin{cases}\tilde{U}(x,\xi)\approx\sum_{j=0}^{p-1}\alpha_j\phi_j\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\\ \tilde{I}(x,\xi)\approx\sum_{j=0}^{p-1}\beta_j\phi_j\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\end{cases} \quad (24)$$

Where the expansion coefficients α_j and β_j can be computed with regression approach, using the least squares method, the quadratic deviation ΔU between the random [8].

$\tilde{U}$: Solution approach by PCE, $\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)$: Constituent plan of experiences

Variables U and $\tilde{U}$ is minimized with respect to the coefficients $\{\alpha_k,k=0,...,p-1\}$

$$\Delta U=\sum_{i=1}^{n}\left(U^{(i)}-\tilde{U}^{(i)}\right)^2 \quad (25)$$

This is equivalent to minimizing the coefficients $\{\alpha_k,k=0,...,p-1\}$ the expression:

$$\Delta U=\sum_{i=1}^{n}\left(U^{(i)}-\sum_{j=0}^{p-1}\alpha_j\phi_j\left(\left\{\xi_k^{(i)}\right\}\right)_{k=1}^{M}\right)^2 \quad (26)$$

The minimum is obtained for α_j verifying values

$$\frac{\Delta U}{\Delta U_y}=0 \quad \forall y\in\{0,...,p-1\} \quad (27)$$

Where

$$\sum_{i=1}^{n}\phi_y\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\left(U^{(i)}-\sum_{j=0}^{p-1}\alpha_j\phi_j\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\right)=0 \quad (28)$$

$$\forall y=\in\{0,...,p-1\}$$

Finally the system that we must solve:

$$\sum_{j=0}^{p-1}\alpha_j\left[\sum_{i=1}^{n}\phi_y\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\phi_j\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\right]=\sum_{i=1}^{n}U^{(i)}\phi_y\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right) \quad (29)$$

$$\forall y=\in\{0,...,p-1\}$$

We can write it also as following:

$$\begin{bmatrix} \sum_{i=1}^{n}\phi_0\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\phi_0\left(\left\{\xi_{k=1}^{M}\right\}\right)\cdots & \sum_{i=1}^{n}\phi_0\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\phi_{P-1}\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right) \\ \vdots & \ddots \\ \sum_{i=1}^{n}\phi_{P-1}\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\phi_0\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\cdots & \sum_{i=1}^{n}\phi_{P-1}\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\phi_{P-1}\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right) \end{bmatrix}\times \begin{pmatrix}\alpha_0\\ \vdots\\ \alpha_{P-1}\end{pmatrix}=\begin{pmatrix}\sum_{i=0}^{n}U^{(i)}\phi_0\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\\ \vdots\\ \sum_{i=0}^{n}U^{(i)}\phi_{P-1}\left(\left\{\xi_k^{(i)}\right\}_{k=1}^{M}\right)\end{pmatrix} \quad (30)$$

By solving the system (30), we find the coefficients $[\alpha_0,...\alpha_{P-1}]$, so can solve equation (24).

IV. SIMULATION

A. *Simulation on Deterministic Model*

To validate our temporal approach, we compare between the coupling formalism in temporal and the one in frequency. In addition, we use the Inverse Fourier Transform (IFFT).We consider an aerial cable with an outer raduis r_{ext} = 2.475mm, located at a height h=10m and a length ℓ = 1km above a ground with a fixed conductivity 0.01S/m. The cable is excited by a bi-exponential generator which is given as below and shown in Fig.1.

$$V_{excit}(t)=V_0(\exp(-\alpha t)-\exp(-\beta t))$$

With the parameters values:

$V_0=10^6 v$, $\alpha=0{,}00364 \quad .10^6 S^{-1}$ and $\beta=0{,}6522 \quad .10^6 S^{-1}$.

We calculate induced voltage at both ends of the core and shielding. Where we fix all the geometric and material parameters values of the cable which are conductor, insulating, conductivity, permittivity and permeability of the layers as following:

For the core: r_c=0.45 mm, r_{ic}=1.457 mm, $\sigma_c=1^8 s/m$, $\varepsilon_{rc}=1.8$, $\varepsilon_{ric}=1.8$ and $\mu_r=1$

For the shielding: r_{sh}=1.775 mm, r_{ish}=2.457 mm, $\sigma_c=0.58^8 s/m$, $\varepsilon_{rsh}=1$, $\varepsilon_{rish}=1.8$ and $\mu_r=1$

The results are shown in Fig.2 and Fig.3.

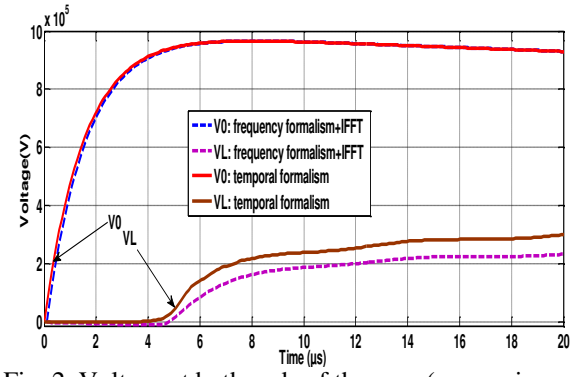


Fig. 2. Voltage at both ends of the core (comparison between temporal and frequency formalism + IFFT)

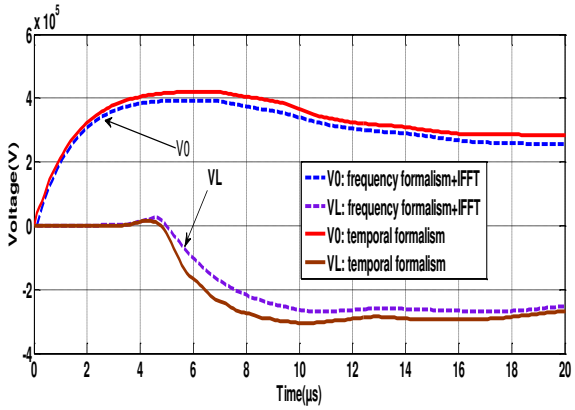


Fig. 3. The induced voltage at both ends of the shielding (comparison between temporal and frequency formalism + IFFT)

In short, our aim from calculating the induced voltage in the cable is to know the voltage distribution along the core and shielding of the cable, using two different ways; temporal and frequency formalism +IFFT and gives very close of calculating the induced voltage.

B. *Simulation on Stochastic Model*

In this case, we impose that the time is constant and fixe its value at $t=2\mu s$. The cable is excited by a bi-exponential generator with random parameters values, where the amplitudeV0, the constant β and the constant α are in the range:

$$V_0=[0{,}9.10^6 \quad 1{,}1.10^6],\ \beta=[0{,}64.10^6 \quad 0{,}66.10^6]$$
$$\alpha=[0{,}0034.10^6 \quad 0{,}0037.10^6]$$

We calculate the induced voltage PDF at the ends of the core and shielding at $t=2\mu s$, using two different ways: the Monte Carlo (MC) and Chaos Polynomial (PC) methods where the results are represented in fig.4 and fig.5 as

below:

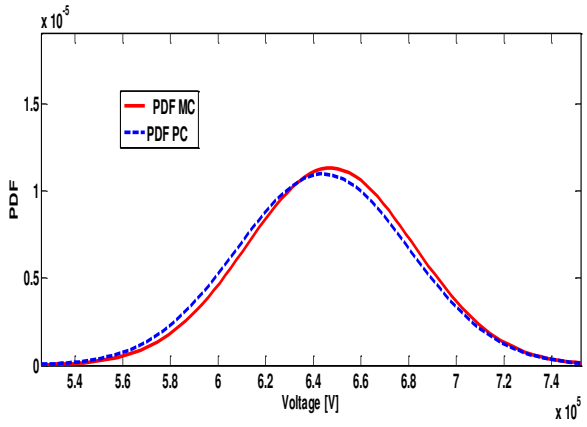


Fig. 4. PDF of voltage at the core end L = 0.

Table I. The statistical parameters of voltage at the core end L=0.

Number of simulations / Statistic parameters	166 375 MC simulations	3 PC simulations (P=1)
Mean $(\mu)[V]$	647174	644104
Variance $(\sigma^2)[V]$	1,2367.10^9	1,31618.10^9
Standard deviation $(\sigma)[V]$	35166.8	36279.1
Simulation time	1h	1sec

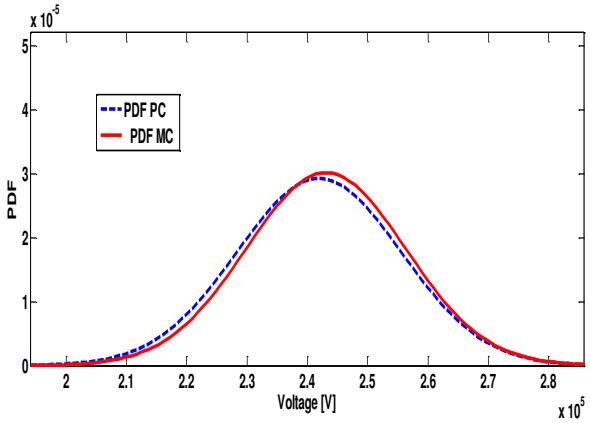


Fig. 5. PDF of induced voltage at the shielding end L = 0.

Table II. The statistical parameters of induced voltage at the shielding end L=0.

Numbre of simulations / Statistic parameters	166 375 MC simulations	3 PC simulations (P=3)
Mean $(\mu)[V]$	243090	241937
Variance $(\sigma^2)[V]$	1,74485.10^8	1,85698.10^8
Standard deviation $(\sigma)[V]$	13209.3	13627.1
Simulation time	1h	1sec

The Monte Carlo method used 166 375 random choice for the excitation generator parameters values, where The Chaos Polynomial method used only 3 random choices for that. So we notice that there is a good coherence between MC and PC methods by regression simulations which is appreciated.

V. CONCLUSION

In this work, we discussed the excitation source coupling problem for a shielded cable. using both deterministic and stochastic method, the challenge point of this method is to model the coupling where we calculated the induced voltage at the cable ends, and the statistical model in which we used two different methods: (MC) and (PC), we conclude that the PC method has an efficiency in the practical applications which are dealing with the parameters that is characterized by a high variability in uniform statistical distributions.

APPENDIX 1

For a shielded cable, located above lossy ground, using Kirchhoff's laws (current and voltage) and the concept of coupled lines the matrix $[\bar{Z}]$ is given by A. Ametani [5] in the following form:

$$[\bar{Z}] = [Z_0] + [Z_i] \qquad (1.A)$$

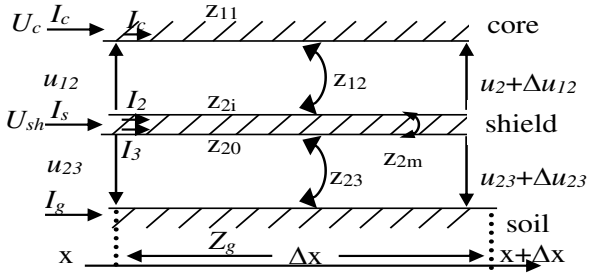


Fig.A.1: Equivalent circuit for impedances of a shielded cable

Where [Zi] expresses the self-impedance matrix of a shielded cable

$$[Z_i] = \begin{bmatrix} Z_{cc} & Z_{cs} \\ Z_{sc} & Z_{ss} \end{bmatrix} \qquad (2.A)$$

The elements of the matrix [Zi] are defined as follows:
Z_{cc}: core self-impedance
Z_{ss} : sheath self-impedance
Z_{cs}: mutual impedance between the core andsheath
Where:

$$Z_{cc} = (z_{11} + z_{12} + z_{2i}) - 2z_{2m}$$
$$Z_{ss} = (z_{20} + z_{23})$$
$$Z_{cs} = (z_{20} + z_{23}) - 2z_{2m}$$
$$z_{cs} = (z_{20} + z_{23}) - 2Z_{2m}$$

z_{11}: internal impedance of core outer surface
z_{12}: core outer insulator impedance
z_{2i}: internal impedance of sheath inner surface
z_{2m}: sheath mutual impedance
z_{20}: internal impedance of sheath outer surface
z_{23}: sheath outer insulator impedance

At this stage we propose to write the matrix [Zi] as following:

$$[Z_i] = [Z_{sh}] = [Z_L] \qquad (3.A)$$

$$[Z_L]=\begin{bmatrix} z_{12}+z_{23} & z_{13} \\ z_{23} & z_{33} \end{bmatrix} \quad (4.A)$$

$$[Z_{sh}]=\begin{bmatrix} z_{2i}+z_{20}+2z_{2m} & z_{20}+z_{2m} \\ z_{20}+z_{2m} & z_{20} \end{bmatrix} \quad (5.A)$$

We note $[Z_{sh}]$ as impedance matrix of the shielding.

For the matrix $[Z_0]$ called the impedance of the ground return, it is defined as below:

$$[Z_0]=\begin{bmatrix} Z_0 & Z_0 \\ Z_0 & Z_0 \end{bmatrix} \quad (6.A)$$

$$[Z_0]=[Z_s]=\begin{bmatrix} Z_g & Z_g \\ Z_g & Z_g \end{bmatrix} \quad (7.A)$$

Where Z_g is the ground impedance that takes into account the effect of finite conductivity of the ground, whose general expression is given by Sunde [4];

$$Z_{gij}=\frac{j\omega\mu_0}{\pi}\int_0^{\infty}\frac{e^{-(h_i+h_j).x}}{\sqrt{x^2+\gamma_s^2}+x}.\cos(r_{ij}.x)dx \quad (8.A)$$

And Z_e is the self-external impedance of the circuit formed by the cable and the perfect soil. In the case of an overhead cable with an outer radius r located at a height h of the ground, this impedance is expressed as following:

$$Z_e=\frac{j\omega\mu_0}{2\pi}\ln\left(\frac{2h}{r_{ext}}\right)=j\omega L_e \quad (9.A)$$

which gives us the matrix [Z], the following expression

$$[Z]=[Z_{sh}]+[Z_L]+[Z_s]+[Z_e]=[\overline{Z}]+jw[L] \quad (10.A)$$

By defining $[\overline{Z}]$

$$[\overline{Z}]=[Z_{sh}]+[Z_s] \quad (11.A)$$

And the matrix [L]:

$$[L]=\frac{([z_e]+[z_L])}{jw}=\begin{bmatrix} l_e+l_{12}+l_{23} & l_e+l_{23} \\ l_e+l_{23} & l_e+l_{23} \end{bmatrix} \quad (12.A)$$

Whose elements are independent of frequency and well-defined analytically.

REFERENCES

[1] C. R. Paul, "Analysis of Multiconductor Transmission Lines", John Wiley & Sons, Inc. New York, 1994.

[2] F. M. Teshe, "On the inclusion of loss in time domain solutions of electromagnetic interaction problems", IEEE Trans. EMC 32(1) 1990.

[3] M. Feliziani and F. Maradi, "Field analysis of penetrable conductive shields by the finite difference time-domain method with impedance network boundary conditions (INBCs) ", IEEE Transactions on Electromagnetic Compatibility, vol.41, No.4, pp. 307-319, November 1999.

[4] F. Rachidi, S. L. Loyka, CA. Nucci and M. Ianoz, "Anew expression for ground transient resistance matrix elements of Multiconductor overhead transmission lines", Journal of Electric Power System Research, vol. 65, No. 1, pp.41-46, 2003.

[5] A. Ametani, "A General Formulation of Impedance and Admittance of cables", IEEE Trans on Power Apparatus and systems, vol.PAS-99, N°.3, May/June 1980.

[6] S. Kaouche, S. Mezoued, B. Nekhoul, K. Kerroum and K. El Khamlichi Drissi, "Induced Disturbance in Power Network by Lightning", International Symposium on Electromagnetic compatibility: EMC Europe Barcelona, Spain, pp. 935-940, September 4-8, 2006

[7] S. Stievano, P. Manfredi, and F. G. Canavero, "Stochastic analysis of multiconductor cables and interconnects," IEEE Trans. Electromagn. Compat, vol. 53, no. 2, pp. 501–507, May 2011.

[8] M. Berveiller, "Eléments finis stochostiques: approaches intrusive et non intrusive pour des analyses de fiabilite", PHD Thesis 2005.

[9] P. Manfredi et al, "Stochastic Transmission Line Analysis via Polynomial Chaos Methods: an Overview". IEEE Electromagnetic Compatibility Magazine, vol. 6 n. 3, pp. 77-84, November 2017.

[10] A. Papoulis, "Probability, Random Variables and Stochastic Processes". New York: McGraw-Hill, 1991.

[11] S. Shiran, B. Reiser, H. Cory, "A probabilistic model for the evaluation of coupling between transmission lines", IEEE Trans.on EMC, vol. 35, No.3, Aug. 1993.

[12] R. Trinchero, P. Manfredi et al, "Machine learning for the Performance Assessment of High Speed Links, " IEEE Transaction on Electromagnetic Compatibility. February 2018.

Impact of High Penetration Photovoltaic Generation On Distribution System Voltage Regulation and Stability

Zouhir Boumous
LEER Laboratory Department of Electrical Engineering
University of Souk-Ahras
Souk-Ahras, Algeria
zohir.boumous@univ-soukahras.dz

Samira Boumous
LEER Laboratory, Department of Electrical Engineering
University of Souk-Ahras
Souk-Ahras, Algeria
samira.boumous@univ-soukahras.dz

Med Mounir Abbes
LEER Laboratory, Department of Electrical Engineering
University of Souk-Ahras
Souk-Ahras, Algeria
m.abbes@unvi-soukahras.dz

Rayane Yasmine Mahfoudi
LEER Laboratory, Department of Electrical Engineering
University of Souk-Ahras
Souk-Ahras, Algeria
r.mahfoudi@univ-soukahras.dz

Hadia Belhouchet
LEER Laboratory, Department of Electrical Engineering
University of Souk-Ahras
Souk-Ahras, Algeria
ha.belhouchet@univ-soukahras.dz

***Abstract*—This paper investigates the effects of integrating large-scale photovoltaic (PV) generation, with penetration levels up to 50%, on voltage regulation and stability in distribution systems. Voltage profiles are analyzed using power-flow calculations over a 24-hour period with varying load conditions. Steady-state voltage stability is evaluated at different times of the day through a custom continuation power-flow method, employing demand as the continuation parameter up to maximum load conditions. The load-flow analysis for both voltage regulation and stability is performed using the forward/backward sweep method. A secant predictor technique is implemented to estimate node voltages, followed by corrections using a load-flow solver. Three models of PV interface inverters are utilized in the study, incorporating comprehensive environmental data. These inverters regulate node voltages by leveraging available capacity for reactive power support. Various potential voltage collapse scenarios are assessed across different times of the day. The developed methods and models are applied to a 33-bus radial distribution feeder with high PV penetration. Results show that PV inverters enhance voltage profiles, improve system reliability, and reduce the switching frequency of on-load tap-changing transformers, there by extending their operational lifespan.**

***Index Terms*—PV system, IEEE test system, Voltage profile, Regulation, Stability.**

I. INTRODUCTION

Voltage Regulation in Distribution Networks Voltage regulation is essential for reliable power distribution. Traditional methods, like on-load tap-changing (OLTC) transformers and capacitor banks, are designed for unidirectional power flow. However, PV systems introduce bidirectional flow, which complies with traditional controls. Authors in [1-2] , highlight how high PV penetration increases voltage fluctuations and OLTC wear, calling for PV inverters to assist with reactive power. PV Inverter-Based Solutions PV inverters can regulate voltage through reactive power control. Researchers in [3] found that this support improves voltage profiles and reduces the workload on traditional equipment, enhancing reliability and extending component lifespan. Steady-State Voltage Stability Analysis This aspect ensures voltage remains stable under normal conditions and disturbances. Research by [4] using continuation power-flow (CPF) methods showed that while PV helps support voltage, excessive penetration without control can push systems toward instability. Load-Flow Analysis Techniques Accurate load-flow analysis, such as the forward/backward sweep method noted by [5], is vital for understanding PV impacts. Predictive techniques like the secant predictor enhance node voltage accuracy in dynamic conditions. Case Studies on High PV Penetration Real-world studies, such as [6] on a 33-bus system, show that advanced PV inverters with reactive power capabilities can improve voltage regulation and avert collapses during peak generation. Our study aims to evaluate how PV systems, which generate electrical power from solar energy, can contribute to reactive power in distribution systems. Reactive power is essential for maintaining voltage levels within safe operating ranges, which

directly impacts the stability and reliability of the power grid.

II. MATHEMATICAL MODELING

$$P_i^k = (1 + k_{pi}\tau)p_{io} \quad (1)$$

$$Q_i^k = (1 + k_{qi}\tau)Q_{io} \quad (2)$$

P_{io}, Q_{io} the active and reactive powers at the base load, and it can be considered initially equal zero

P_i^k, Q_i^k the active and reactive powers at point k

K_i the load multiplier at bus i.

τ the continuation parameter factor .

PV's output power can be computed using equations (3) through (5). This model has taken into account how the temperature of the surrounding environment and solar radiation affect the output power. The following are the Maximum Power Point (MPP) equations:

$$P_{pv} = V_{mmp} \times I_{mmp} \quad (3)$$

$$V_{mmp} = V_{mmp,ref} \times P_{v,oc}(T_c - T_{c,ref}) \quad (4)$$

$$I_{mmp} = I_{mmp,ref} \times I_{sc,ref}(T_c - T_{c,ref}) \quad (5)$$

Where:

P_{pv} is the panel's power

V_{mmp} is the potential voltage

$V_{mmp,ref}$ is V_{mmp} at standard condition

I_{mmp} and $I_{sc,ref}$ is the panel's current and short circuit current at standard condition respectively

$P_{v,oc}$ is the open circuit (V/°c) temperature coefficients.

$T_{c,ref}$ is the temperature of the panel at standard operational conditions that is equal to 250°c and T_c is the operational temperature of the panel that is calculated as follows:

$$T_{C(t)} = T_{a(t)} + \frac{NOCT-20}{800} \cdot G_T \quad (6)$$

Where, $T_{a(t)}$ is the ambient temperature (°C),

III. THE STUDY OF POWER SYSTEM

The IEEE 33-bus test system (figure 1) is a standard benchmark widely used in power system research to analyze and test algorithms and methodologies for distribution network studies, including load flow analysis, optimal power flow (OPF), and voltage stability. It represents a medium-sized, radial distribution system typically consisting of 33 buses (or nodes) and 32 branches (lines), which interconnect the buses [7]. The network is designed to mimic a typical urban or industrial electrical distribution system where power flows from a central source, usually represented by a voltage source at bus 1, to various loads located at different buses. The system's voltage levels are typically set at 12.66 kV, which corresponds to a medium-voltage distribution network, and the total load demand is about 3.715 MW, with a reactive power demand of around 2.300 MVAr. The buses in this system model a range of points in the network, from substations to consumer connections, and the network structure is radial, meaning there are no loops, and power flows outward from the central source[8-9]. The test system is mainly used for power flow studies, where the distribution of active and reactive power across the system is analyzed, as well as the voltage profile at each bus. This helps determine whether the network can operate efficiently and within safe limits. The IEEE 33-bus system is also employed in optimal power flow studies, where the objective is to minimize costs, and losses, or improve voltage regulation while respecting system constraints[10]. The branches between the buses are modeled with typical resistance and reactance values that represent real-world power lines. Though the base version of the IEEE 33-bus system does not typically include distributed generation, it can be modified to incorporate renewable energy sources such as solar or wind power, and energy storage devices like batteries, which are increasingly relevant in modern power systems. Additionally, this test system can be used to study fault conditions, reliability, stability, and control strategies such as voltage regulation and demand-side management. Researchers and engineers use power system simulation software like MATLAB/Simulink, DIgSILENT PowerFactory, and OpenDSS to analyze the IEEE 33-bus system, enabling them to test new algorithms and validate results for practical applications. Given its simplicity and well-established configuration, the IEEE 33-bus system remains a key reference model for evaluating the performance of various power system techniques and control strategies, especially in the context of integrating renewable energy and improving grid stability[11]. The study assesses the impact of introducing photovoltaic generators on three different buses, numbers 22, 25, and 33.

IV. RESULTS AND DISCUSSION

Figure (2) illustrates the daily variation in active power demand (in kW) across 24 hours for different consumer types: commercial, industrial, urban, and light. The commercial sector, represented by the blue bars, shows the highest power

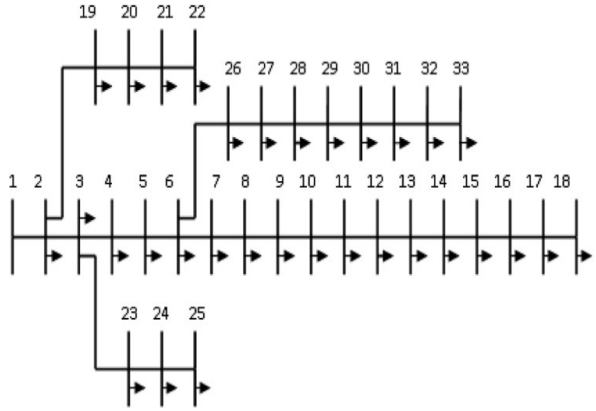


Fig. 1. A single-line diagram of the IEEE 33-bus test distribution network

consumption, especially during business hours (morning to early evening). Industrial power usage, shown in orange, also contributes significantly during the daytime, with some fluctuations. Urban (yellow) and light power usage remain lower and more constant throughout the day, indicating residential and public area demands. Overall, power consumption peaks during midday to early evening, driven by commercial and industrial activities, and decreases during the night when only urban and lighting needs persist.

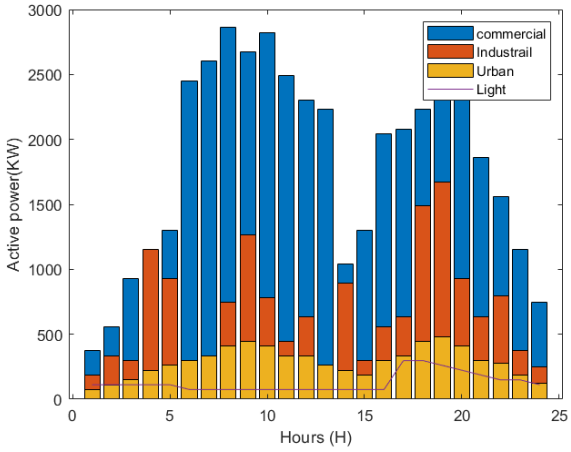


Fig. 2. Active power in different load type

Figure (3) presents the total power load (in kW) over a 24-hour period, showing a typical daily pattern of energy consumption. The load is lowest during the early morning hours (0-5 hours), then rises sharply in the morning (5-9 hours) as commercial and industrial activities start, reaching a peak around midday (9-12 hours). There is a slight dip in the early afternoon (12-15 hours), potentially due to reduced industrial operations during lunch hours. The load surges again in the late afternoon to evening (15-20 hours), indicating increased residential and continued commercial usage, before dropping off significantly at night (20-24 hours) as activity subsides. The dual peaks in the late morning and evening reflect typical

work and residential energy demands.

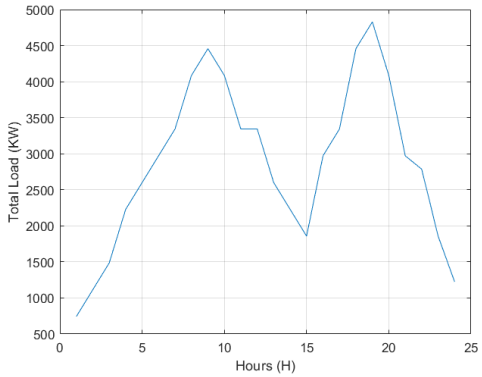


Fig. 3. Total load during the day

The pie chart in Figure (4) represents the distribution of total power load among different sectors: commercial, industrial, urban, and light. The largest share is held by the commercial load, depicted in dark blue, indicating that it is the most significant contributor to overall power consumption. The industrial load, shown in light blue, is the next major contributor, although it is notably smaller than the commercial portion. The urban load, represented in green, occupies a smaller fraction, while the light load, shown in yellow, takes up the smallest segment of the chart. This distribution highlights the dominance of commercial energy use, followed by industrial usage, with urban and light loads making up much smaller shares.

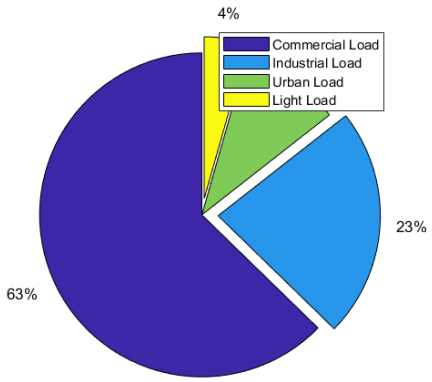


Fig. 4. Percentage of loads in the power system

The graph in Figure (5) shows two line graphs representing the active and reactive power of a photovoltaic (PV) system at bus 22 over a 24-hour period. In the top graph, the active power in kW rises sharply in the early morning, peaking around midday to early afternoon, and then decreases in the evening, indicating PV generation correlating with sunlight availability. The bottom graph shows the reactive power in kVAr, which

fluctuates more than the active power. It starts with a positive value, decreases to negative values around midday, and then returns to positive in the evening. This variation in reactive power reflects the changing power factor and voltage support provided by the PV system throughout the day.

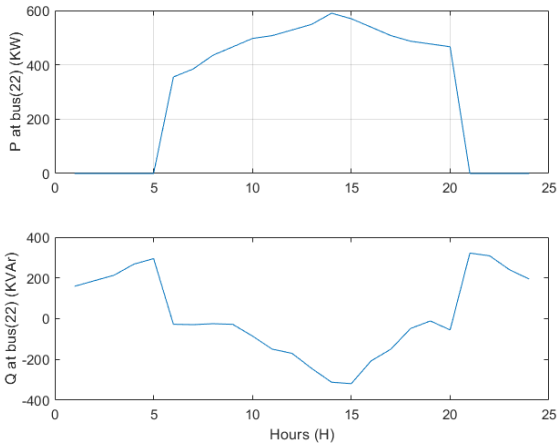


Fig. 5. Active and reactive power in Bus number 22

Figures (6) and (7) show two line graphs of active and reactive power for a photovoltaic (PV) system at bus 25 and bus 33 over 24 hours. In the top graph, the active power (in kW) follows a similar pattern to typical PV generation: it rises sharply in the morning, peaks around midday, and drops to zero in the evening as sunlight decreases. The bottom graph displays the reactive power (in kVAr), which has significant fluctuations throughout the day, reaching its highest values in the morning and late afternoon, with a notable dip around midday. These variations suggest dynamic adjustments in reactive power to manage voltage stability and power factors in response to changes in PV generation.

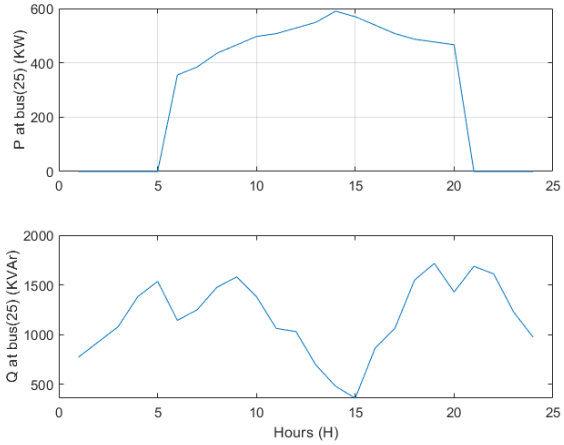


Fig. 6. Active and reactive power in Bus number 25

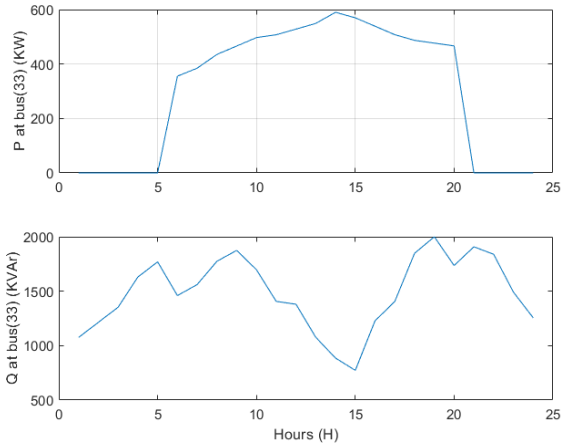


Fig. 7. Active and reactive power in Bus number 33

In Figure (8) we present the voltage profiles across various buses in a power distribution network, measured in per unit (pu) over a 24-hour period, for two scenarios: without photovoltaic (PV) support (top) and with PV support (bottom). In both scenarios, voltage levels drop around midday, likely due to peak demand. In the ”Without PV” scenario, the voltage at some buses drops to around 0.85 pu during midday, showing a significant decrease from the nominal voltage level. Most buses remain between 0.85 pu and 0.95 pu during this dip, with recovery to around 0.95–1.0 pu during other hours. In the ”With PV” scenario, the midday voltage dip is less severe, with the lowest values reaching around 0.9 pu instead of 0.85 pu. This indicates that PV support helps maintain voltage stability by keeping buses closer to nominal voltage (0.95–1.0 pu) throughout the day. The PV system reduces the depth of the voltage drop, improving the overall voltage profile and enhancing network reliability during high-demand periods.

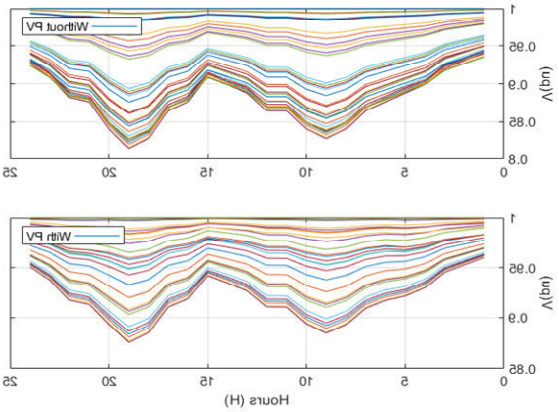


Fig. 8. Voltage evolution in different Bus: a) without PV, b) With PV

Figure (9) illustrates voltage deviation over 24 hours for a distribution system with and without PV (photovoltaic) generation. The solid blue line, representing the system without

PV, shows higher voltage deviations, peaking around mid-morning and late evening with values reaching up to 0.7. In contrast, the dashed red line, showing the system with PV, has consistently lower deviations, peaking below 0.4 during similar times. This indicates that the integration of PV generation significantly reduces voltage deviations, resulting in a more stable voltage profile throughout the day, particularly during peak periods.

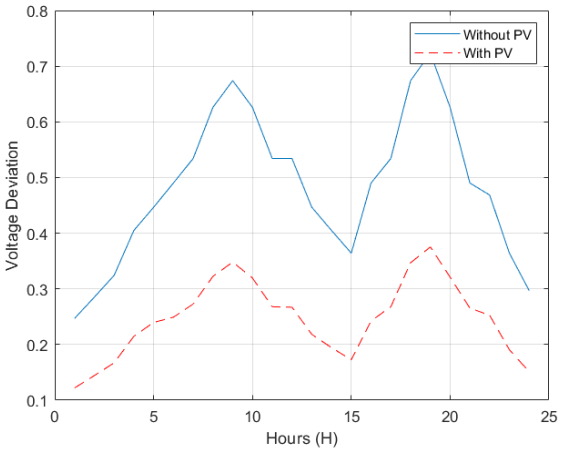


Fig. 9. Voltage deviation without PV and With PV

V. CONCLUSION

The study examined voltage stability in distribution systems with high levels of PV generation that include reactive power support. Case studies under various operation modes and environmental conditions showed that reactive power support enhances voltage regulation and creates a smoother voltage profile that stays within acceptable limits, even with changing demand. The findings highlight that voltage stability improves with reactive power support compared to unity power factor operation, resulting in a higher maximum loading capacity. Increasing the inverter rating beyond the PV module's real power capacity allows the power conditioning unit to provide reactive power when needed. A 10% increase in the inverter's capacity benefits the system during peak PV generation, helping maintain voltage within the desired range, though it only slightly impacts the maximum loading condition. Reactive power support significantly enhances voltage stability in distribution systems with high PV penetration by providing a mechanism to counteract voltage fluctuations that occur due to varying power outputs from PV sources. When PV generation rapidly changes due to environmental conditions like cloud cover, reactive power support helps to stabilize the voltage by compensating for the sudden imbalances. It does this by either absorbing or injecting reactive power as needed, which keeps voltage levels within acceptable limits and prevents sharp deviations that could lead to instability or voltage collapse. Additionally, reactive power support helps improve the system's maximum loading capacity, allowing the distribution network to handle greater power flows without compromising voltage stability. This is particularly important during peak generation times, such as midday when PV output is highest, as it ensures that voltage remains within safe operating limits, even under high generation conditions. By maintaining a smooth voltage profile and supporting the power system during demand variations, reactive power support contributes to the overall reliability and resilience of the grid, making it essential for systems with significant renewable energy integration.

REFERENCES

[1] R. Nandi, M. Tripathy, and C. P. Gupta, "Coordination of BESS and PV system with bidirectional power control strategy in AC microgrid," Sustainable Energy, Grids and Networks, vol. 34, p. 101029, 2023. doi: 10.1016/j.segan.2023.101029.

[2] J. Hu et al., "A coordinated control of hybrid AC/DC microgrids with PV-wind-battery under variable generation and load conditions," Int. J. Electr. Power Energy Syst., 2019.

[3] K. P. Bharti, H. Ashfaq, R. Kumar, and R. Singh, "Designing a Bidirectional Power Flow Control Mechanism for Integrated EVs in PV-Based Grid Systems Supporting Onboard AC Charging," Sustainability, vol. 16, no. 20, p. 8791, 2024. doi: 10.3390/su16208791.

[4] N. Dharmala et al., "Win-win transportation strategies for India: Linking air pollution and climate mitigation," Energy Clim. Chang., vol. 3, p. 100072, 2022.

[5] A. Gupta, H. O. Bansal, P. Jaiswal, and R. Kumar, "Modeling and Analysis of a V2G Scheme: A Concept in Smart Grid," in Proc. 2020 Int. Conf. Emerging Trends Commun., Control and Computing (ICONC3), Sikar, India, 2020, pp. 1–6.

[6] F. Jain and S. Bhullar, "Operating Modes of Grid Integrated PV-Solar Based Electric Vehicle Charging System—A Comprehensive Review," e-Prime- Adv. Electr. Eng. Electron. Energy, vol. 8, p. 100519, 2024.

[7] H. Rashidizadeh-Kermani, H. R. Najafi, A. Anvari-Moghaddam, and J. M. Guerrero, "Optimal decision-making strategy of an electric vehicle aggregator in short-term electricity markets," Energies, vol. 11, no. 9, p. 2413, 2018.

[8] S. H. Dolatabadi, M. Ghorbanian, P. Siano, and N. D. Hatziargyriou, "An Enhanced IEEE 33 Bus Benchmark Test System for Distribution System Studies," IEEE Trans. Power Syst., vol. 36, no. 3, pp. 2565–2572, May 2021. doi: 10.1109/TPWRS.2020.3038030.

[9] M. Ghorbanian, S. H. Dolatabadi, and P. Siano, "Game theory-based energy-management method considering autonomous demand response and distributed generation interactions in smart distribution systems," IEEE Syst. J., vol. 15, no. 1, pp. 905–914, Mar. 2021.

[10] D. Feroldi and P. Rullo, "Optimal Operation for the IEEE 33 Bus Benchmark Test System With Energy Storage," in Proc. 2021 IEEE URUCON, Montevideo, Uruguay, 2021, pp. 1–5. doi: 10.1109/URUCON53396.2021.9647175.

[11] I. B. Sperstad and M. Korpa˚s, "Energy storage

scheduling in distribution systems considering wind and photovoltaic generation uncertainties," Energies, vol. 12, no. 7, p. 1231, 2019.

Modeling and Simulation of an Intelligent Energy Management System for Hybrid Renewable Energy Microgrids

Chakib Guebli, Boubakeur Rouabah, Mohamed Abdelbasset Mahboub, Bilal Benarabi

Electronics And Communications Department
Faculty Of New Information and Communication Technologies
Kasdi Merbah University, Ouargla, Algeria

guebli.chakib@univ-ouargla.dz

Abstract - **This paper presents the modeling and simulation of an energy management system (EMS) for a hybrid renewable energy system consisting of a photovoltaic (PV) array, a hydrogen fuel cell, and a battery storage system, with the goal of efficiently managing the power flow between these sources to supply a microgrid load. The EMS utilizes data such as load demand, PV irradiance, temperature, and battery state-of-charge to determine the optimal source to meet the load requirements. The PV array is connected to a DC-DC converter with a maximum power point tracking (MPPT) controller, the fuel cell is coupled with a boost DC-DC converter regulated by a proportional-integral-derivative (PID) controller, and a bidirectional DC-DC converter with a PID controller is used to charge and discharge the battery bank. An intelligent switching function is developed to manage the power flow between the sources, with the battery used when PV output is insufficient and the fuel cell activated as a backup if the battery state-of-charge drops below 15% and PV power is still unavailable. The entire hybrid system is modeled and simulated using MATLAB/Simulink, and the results demonstrate the effectiveness of the proposed EMS in coordinating the operation of the PV, fuel cell, and battery to reliably power the microgrid load while maximizing the utilization of renewable energy sources.**

Index Terms – Energy management system, Hybrid renewable energy system, Photovoltaic system, fuel cell system, Battery energy storage.

I. INTRODUCTION

The transition towards sustainable and renewable energy sources has become a global priority in recent years, driven by concerns over climate change, environmental impacts, and the depletion of fossil fuel reserves. One promising approach to addressing these challenges is the development of hybrid renewable energy systems, which integrate multiple renewable energy sources and energy storage technologies to provide reliable and efficient power supply [1].

In this context, this paper presents the modeling and simulation of an energy management system (EMS) for a hybrid renewable energy system consisting of a photovoltaic (PV) array, a hydrogen fuel cell, and a battery storage system. The objective of the EMS is to efficiently manage the power flow between these sources to supply a microgrid load, while maximizing the utilization of renewable energy and ensuring a reliable and stable power supply.

The EMS developed in this study employs various power electronic converters and control strategies to coordinate the operation of the PV array, fuel cell, and battery storage system. Specifically, the PV array is connected to a DC-DC converter with a maximum power point tracking (MPPT) controller, the fuel cell is coupled with a boost DC-DC converter regulated by a proportional-integral-derivative (PID) controller, and a bidirectional DC-DC converter with a PID controller is used to charge and discharge the battery bank [1], [2].

Moreover, an intelligent switching function is implemented within the EMS to manage the power flow between the sources, prioritizing the use of renewable energy (PV) and the battery storage system, and activating the fuel cell as a backup when the PV output is insufficient and the battery state-of-charge drops below a critical level.

The proposed EMS is modeled and simulated using MATLAB/Simulink, and the results demonstrate the effectiveness of the system in coordinating the operation of the PV, fuel cell, and battery storage to reliably power the microgrid load while maximizing the utilization of renewable energy sources.

II. LITERATURE REVIEW

The growing interest in hybrid renewable energy systems has motivated extensive research in the field of energy management systems (EMSs) for these applications. Several studies have investigated the modeling, control, and optimization of EMSs for various configurations of renewable energy sources and energy storage systems [3].

A. MPPT Algorithms

Power extraction from photovoltaic is optimized by MPPT algorithms like P&O, IncCond, and PSO. When PV systems were evaluated between P&O and IncCond, it was discovered that IncCond had greater efficiency and quicker convergence. Enhanced power generation for PV turbines by using a PSO-based MPPT algorithm. To investigate their use in microgrid systems and integration with other control mechanisms, more research is necessary [6].

B. PID Control Strategies

For effective battery management in micro-grid systems, PID control is frequently used. It maximizes energy efficiency and prolongs battery life by controlling battery charging and discharging rates in response to load demand. PID battery management techniques have been developed in earlier studies, improving load balance and battery longevity [5]. To

further understand PID tuning methods and how they interact with other control elements, more research is needed.

C. Integration of Fuel Cells

Fuel cell integration with microgrids has drawn a lot of interest. outlined a novel control method that makes use of a fuel cell coupled to a Proportional-Integral (PI) controller on a boost DC-DC converter. Their system guarantees steady functioning and effectively controls power flow. The study emphasizes how crucial control tactics are for making the best use of fuel cell resources [4].

D. EMS Algorithms

Microgrids utilize advanced control algorithms to efficiently manage and transition between various energy sources. The recent studies [2], [3] introduced a novel algorithm that integrates inputs from photovoltaic (PV) systems, fuel cells, and battery storage into an innovative energy management system. This integrated approach enables intelligent coordination and optimization among the different distributed energy resources to effectively power complex applications.

III. System Overview

A. The Microgrid System's Description

Photovoltaic (PV) panels, a fuel cell, a load, and a battery are the various energy sources that make up the suggested system. Creating an Energy Management System (EMS) that efficiently makes use of various energy sources to power the load and recharge the battery is the goal. Utilizing PV power as the main energy source is prioritized by the EMS. To extract the most power possible from the PV panels, it integrates the Maximum Power Point Tracking (MPPT) algorithm. When PV power is not enough to satisfy load demand, the system automatically switches to using the battery as a backup power source by draining it [2], [7].

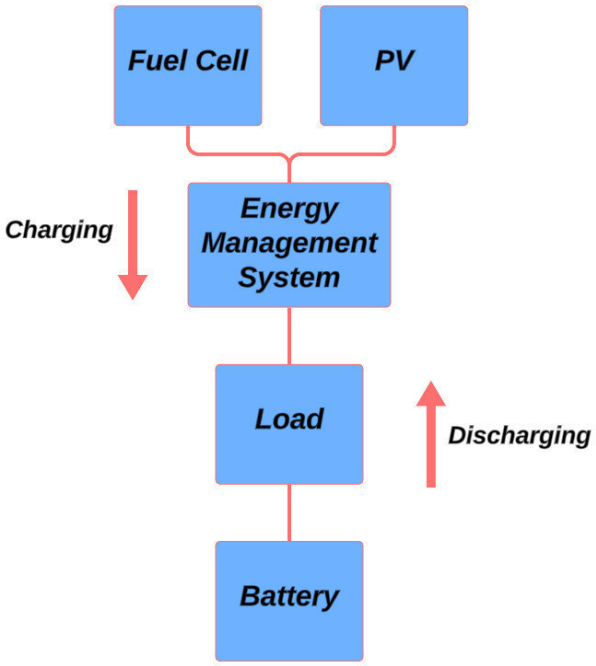


Fig. 1. Integrated Energy Management System Schema for PV-WT Hybrid System

When the PV panels are not producing enough power and the battery is less than 15% charged, the EMS triggers the fuel cell to operate as a backup power source. In order to maintain stable voltage levels, a Proportional-Integral-Derivative (PID) controller is used to control the voltage of the DC-DC converter in the fuel cell system.

The fundamental goal of this system is to prioritize photovoltaic power as the main energy source. The EMS uses the battery to augment the power source if the PV power is not enough. The fuel cell is turned on as a backup power source when the battery is low and stays on until the PV power is once again adequate.

This method guarantees the effective and wise use of the energy sources at hand, giving the load a steady and dependable power supply. The PV system's MPPT algorithm and the fuel cell DC-DC converter's PID controller work together to maximize the efficiency of their respective energy sources.

B. Elements and Their Functions

Figure 2 below provides a more thorough description of these elements.

- Photovoltaic effect: Transform solar energy into electrical energy using PV panels.
- A DC-DC converter with MPPT applies Maximum Power Transfer.
- To get the most power out of the PV, use Maximum Power Point Tracking.
- Fuel cells use an electrochemical process to produce DC electricity.
- DC-DC boost converter with PI controller controls the output voltage and raises the fuel cell's voltage to the necessary level.
- The system's energy storage is provided by the battery bank.
- The battery can be charged and discharged using a bidirectional DC-DC converter with PI controller, which also controls the voltage during these processes.
- The intelligent function of IGBT switches regulates the flow of power between the sources and the load.
- The intelligent control algorithm prioritizes the PV, uses the battery for discharge, and activates the fuel cell as a backup energy source, depending on the system conditions.
- To keep the system's voltage levels constant, PID controllers control the DC-DC converters' voltage.
- The electrical demand that must be satisfied is represented by the load.
- The platform used to implement and simulate the complete system is called Simulink.

C. Algorithmic Distribution and Management of Power

A control approach for regulating the production and distribution of power in a microgrid system is shown in Figure 2. Photovoltaic (PV) panels, fuel cells, and a battery energy storage system make up the microgrid. The goal is to maintain the battery's state of charge (SOC) and maximize the use of renewable energy sources while making sure the microgrid can supply the required amount of power.

The flowchart illustrates a decision-making procedure that ascertains the best course of action in light of the microgrid's demand and available power generation. It describes what should happen when PV generation surpasses demand, when battery discharge is feasible to fulfill demand, and when fuel cell backup is triggered when PV is insufficient and battery level is less than 15%.

PV generating is given priority. PV generation is used first when it is adequate to meet the demand. The extra electricity is utilized to charge the battery if PV generation surpasses demand. The battery is discharged to augment the power supply when photovoltaic generation is not enough to meet demand. The fuel cell is turned on as a backup until PV generation and battery power are restored if the battery level is less than 15% and there is no PV generation available.

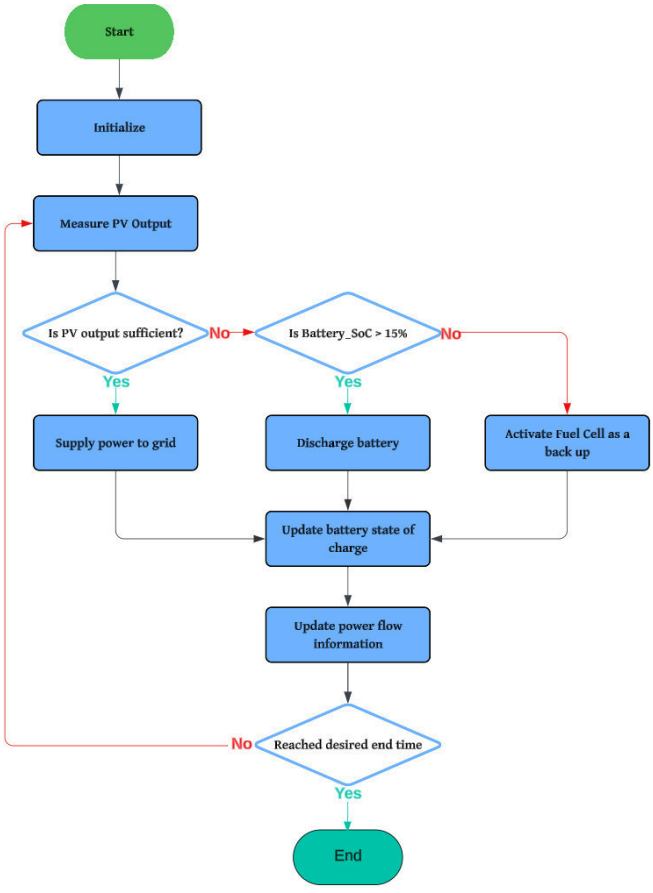


Fig. 2. Smart Microgrid Power Management Algorithmic Flowchart

IV. Results and Discussion

A. PV and Battery Operation

The PV system's operation is illustrated graphically in Figure 4, which also features the PV array and MPPT DC-DC converter. This provides important details about power generation and converter efficacy in the renewable energy configuration.

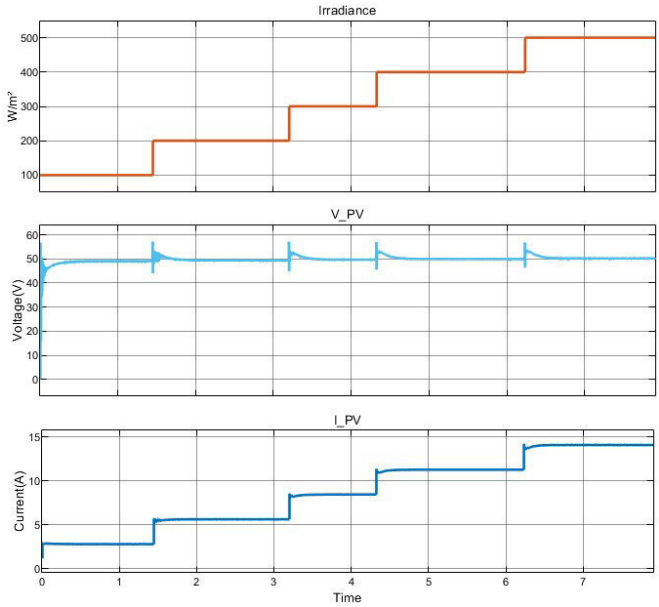


Fig. 3. PV Performance

B. Fuel Cell and Battery Operation

The performance of the fuel cell system is shown in great detail in Figure 5. This figure highlights important variables, such as Vfc and Ifc, and shows how the fuel cell can function as a backup power supply. This fuel cell device has the ability to feed the grid and charge the battery autonomously.

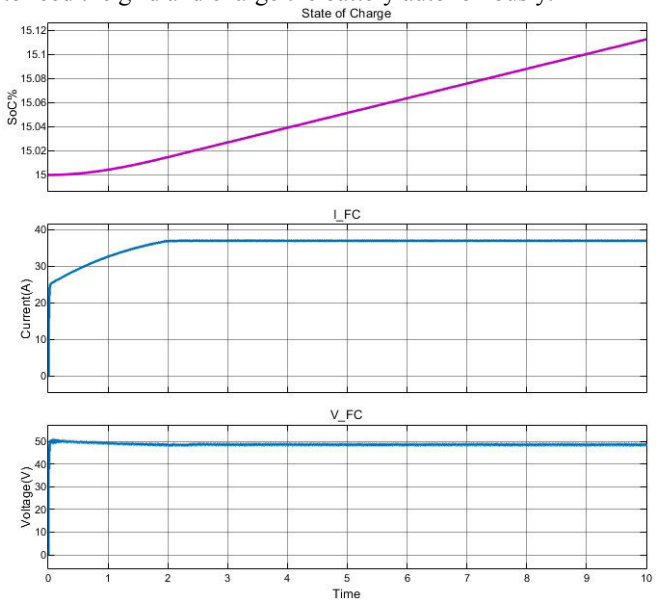


Fig. 4. Fuel Cell Performance

C. Performance of Multi-Source Management

After assessing the operation of the battery multi-source, fuel cell, and PV systems, we now turn our attention to the energy management system. An illustration of the system is shown in Figure 6, which also highlights the clever switching method made possible by IGBTs. We can see how power is allocated and switched between the sources, such Pwt and Pfc, in this picture. The system effectively controls power flow and makes sure that resources are used to their fullest potential. This thorough comprehension of the energy management system highlights its capacity to provide a dependable and effective power source.

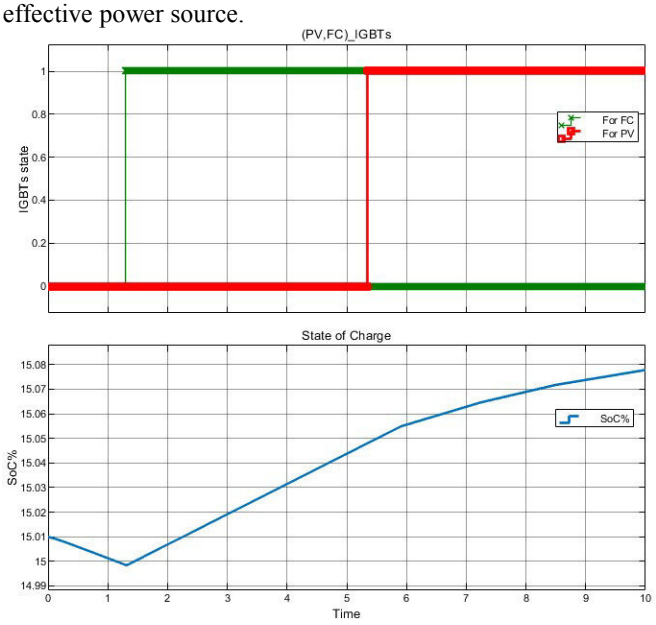


Fig. 5. Performance of the Battery and EMS

In addition to the data, Figure 6 shows the IGBT status for both the fuel cell and the PV, as well as the battery's state of charge (SoC). The energy management system uses intelligent battery draining to satisfy demand when the PV system's power generation is insufficient or nonexistent. Additionally, the energy management system automatically turns on the fuel cell as a backup power source when the battery's state of charge falls below 15%.

The functioning of the energy management system is comprehensively and comprehensively illustrated in the picture. This involves the PV and fuel cell switching on-the-fly depending on the available electricity and battery level. This all-encompassing viewpoint highlights the system's ability to efficiently control energy flow and ensure a steady supply of power.

V. CONCLUSIONS

The examination of a hybrid energy system with PV, fuel cell, and battery components was the main focus of this study. The study provided important insights into the energy management and power dynamics of the system through comprehensive performance assessments. Insulated Gate Bipolar Transistors (IGBTs) facilitated an intelligent switching mechanism that was essential to the effective management and synchronization of power distribution among the system's constituent parts. This creative method made sure that all of the resources were used to their fullest potential, resulting in a dependable and effective power source.

When faced with situations including inadequate PV power or low battery charge, the energy management system proved to be effective. In these situations, the fuel cell was immediately turned on by the system as a backup, guaranteeing end users would always have power. The study offers a thorough comprehension of the multi-source energy system's synergistic operation and performance, emphasizing its potential for dependable and effective power management in a range of applications.

By providing insights to optimize the functioning of multi-source energy systems, this work adds to the body of knowledge in this field. The results can direct the development and implementation of analogous systems, enabling dependable and sustainable power solutions. Potential avenues for future research and development could encompass refining energy management techniques, investigating sophisticated control systems, and incorporating extra renewable energy sources. Such initiatives will promote ecologically responsible and efficient energy systems. The results indicate advancements in dependable and environmentally friendly power infrastructures by enabling stakeholders to make knowledgeable decisions and put creative ideas toward sustainable energy solutions into practice.

REFERENCES

[1] W. Bai, M. R. Abedi, and K. Y. Lee, 'Distributed generation system control strategies with PV and fuel cell in microgrid operation', Control Engineering Practice, vol. 53, pp. 184–193, 2016.

[2] B. Benlahbib et al., 'Experimental investigation of power management and control of a PV/wind/fuel cell/battery hybrid energy system microgrid', International Journal of Hydrogen Energy, vol. 45, no. 53, pp. 29110–29122, 2020.

[3] Y. Han, W. Chen, Q. Li, H. Yang, F. Zare, and Y. Zheng, 'Two-level energy management strategy for PV-Fuel cell-battery-based DC microgrid', International Journal of Hydrogen Energy, vol. 44, no. 35, pp. 19395–19404, 2019.

[4] A. H. Tariq, S. A. A. Kazmi, M. Hassan, S. A. M. Ali, and M. Anwar, 'Analysis of fuel cell integration with hybrid microgrid systems for clean energy: A comparative review', *International Journal of Hydrogen Energy*, vol. 52, pp. 1005–1034, 2024.

[5] Y. Belkhier, A. Oubelaid, and R. N. Shaw, 'Hybrid power management and control of fuel cells-battery energy storage system in hybrid electric vehicle under three different modes', *Energy Storage*, vol. 6, no. 1, p. e511, 2024.

[6] Z. Ishrat, A. K. Gupta, and S. Nayak, 'A comprehensive review of MPPT techniques based on ML applicable for maximum power in solar power systems', *Journal of Renewable Energy and Environment*, vol. 11, no. 1, pp. 28–37, 2024.

[7] A. Neffati, M. Guemri, S. Caux, and M. Fadel, 'Energy management strategies for multi source systems', *Electric Power Systems Research*, vol. 102, pp. 42–49, 2013.

Study And simulation of Packed U Cells 5 Inverter For Photovoltaic Application.

Fateh Khelifi
Dept of Process Automation & Electrification
Faculty of Hydrocarbons and Chemistry,
University of boumerdes
boumerdes, Algeria
Fa.khelifi@univ-boumerdes.dz

Aimad Boudouda
Dept of of Electrical Systems Engineering
Faculty of technology
University of boumerdes
boumerdes, Algeria
a.boudouda@univ-boumerdes.dz

Tadjer Sid ahmed
Dept of Process Automation & Electrification
Faculty of Hydrocarbons and Chemistry
University of boumerdes
boumerdes, Algeria
s.tadjer@univ-boumerdes.dz

Abstract´ the direct exploitation of solar energy through sensors is based on two distinct technologies: solar thermal energy, which produces heat, and solar photovoltaic energy, which generates electricity. To optimize the efficiency of the conversion and management of this energy, static converters play a crucial role.

In the literature, several types of multi-level converters are presented, such as multi-cell inverters, those with floating diodes and those with NPC structure. In our work, we focus on the PUC (Packed U Cell) type voltage inverter, which is used to inject energy from a photovoltaic (PV) source into the electricity grid. This topology makes it possible to produce a quasi-sinusoidal voltage and reduce harmonics thanks to the high number of voltage levels it generates.

In this article we will present the simulation results using Matlab Simulink software of the 5-level PUC inverter with a grid-connected photovoltaic generator.

Keywords´ photovoltaic systems, power converter, Packed U Cells inverter, Maximum Power Point Tracking

I. Introduction

Solar photovoltaic energy is a promising solution to meet growing energy demand while reducing greenhouse gas emissions. However, integrating this energy into electricity networks represents a complex challenge. PUC type multi-level inverters offer innovative solutions to facilitate this integration.

II.Photovoltaic system

Photovoltaic energy production systems connected to a network (figure 1.1) are a result of the trend towards decentralization of the electricity network. Energy is produced closer to the places of consumption. Grid-connected systems reduce the need to increase the capacity of transmission and distribution lines. It produces its own electricity and routes its excess energy to the grid, from which it draws supplies as needed; these transfers eliminate the need to purchase and maintain a battery [1].

It is still possible to use these systems to serve as backup power when a network outage occurs.

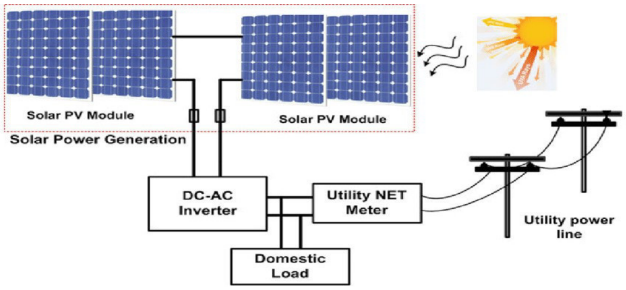


Fig.1 : Photovoltaic systems connected to the grid.

II.1.Characteristics of a PV module

A photovoltaic module is made up of photovoltaic cells that convert solar energy into electricity. Its main characteristics are its peak power (Wp), which corresponds to the maximum power it can deliver under standard test conditions (1000 W/m2, 25éC). The module is also characterized by its current-voltage (I-V) curve, which defines the possible electrical configurations of the cell under a given illumination.

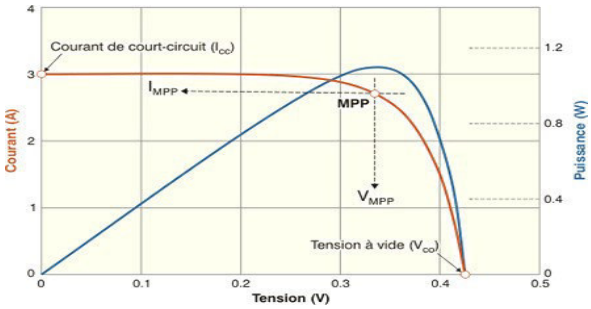


Fig.2. Electrical characteristics of photovoltaic cells and modules

II.2. PV simulation

The LG Electronics Inc. LG385N2C-G4 module has 72 monocrystalline silicon solar cells connected in series. LG Electronics Inc. LG385N2C-G4 can produce a maximum power of 385 watts at 40 volts [2].

This allowed us to determine the power and current as a function of the voltage of the module studied for an illuminance of 1000 W/m2 and a temperature T=25é. The electrical characteristics of the LG Electronics Inc. LG385N2C-G4 Photovoltaic module under standard test conditions are shown in Table 1.

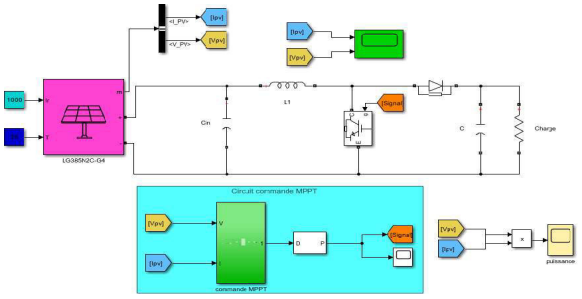


Fig. 3. Simulink block of a PV module controlled by MPPT.

Table1. Parametres of LG Electronics Inc. LG385N2C-G4

Sizes	Values
Standard illuminance G	1000W/m2
Temperature T	25
Maximum peak power Pm	385 W
Optimal voltage V m	40.1 V
Optimal current Im	9.61 A
Short circuit voltage V co	48.9 V
Short circuit current Isc	10.16 A

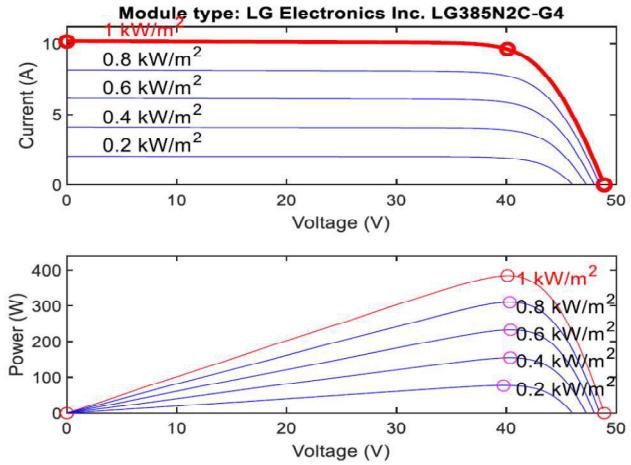
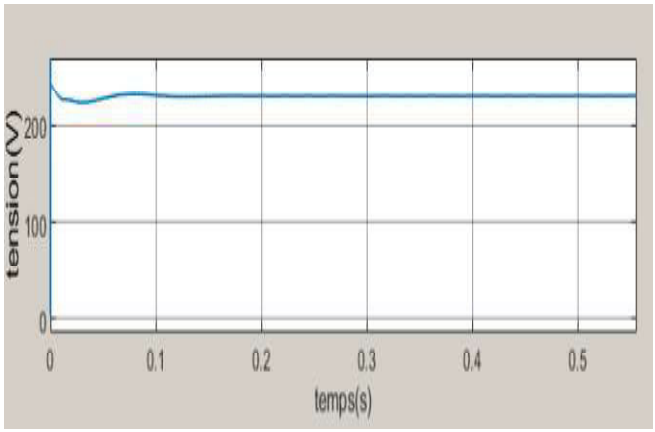


Fig. 4. Influence of variation in illumination on the characteristic (I-V) and (P-V).



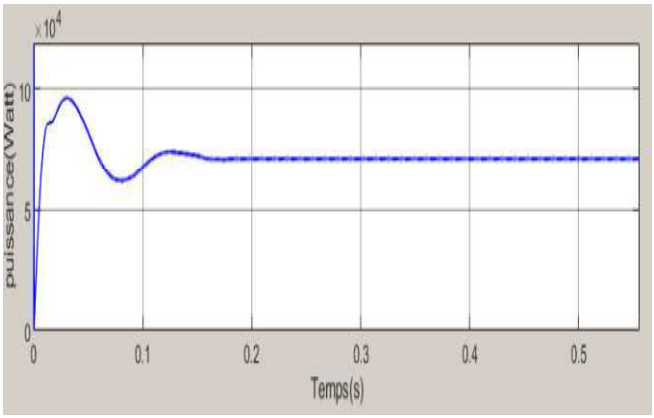


Fig. 5. The maximum power and voltage at the PV output after applying MPPT.

III. The single-phase PUC inverter

The Packed U Cell (PUC) multilevel converter is an advanced topology that evolved from serial multi-cell converters, also known as Flying Capacitor Multi-level Inverter (FCMI), developed in the 1990s. This topology uses cells in the shape of a "U" arranged in a cascade, each cell being composed of two switches and a capacitor. Powered by a direct voltage source, this converter can produce a voltage at 5 levels or more [3].

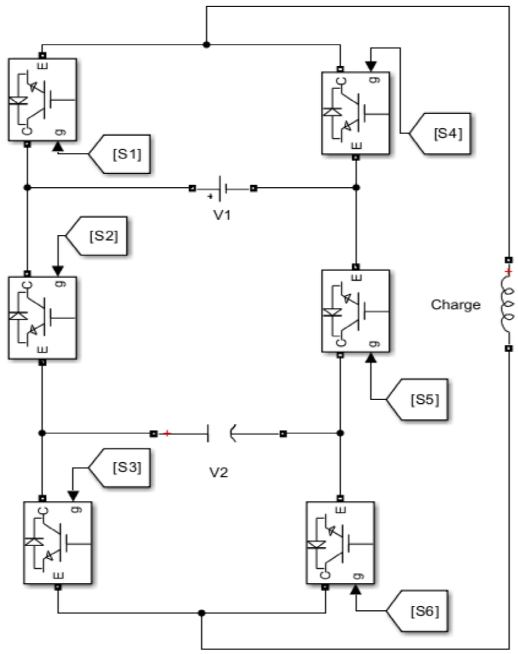


Fig. 6. Five-level PUC converter.

III.1. Command sequences

The PUC5 inverter uses eight different switching states to generate a five-level output voltage. In states 2 and 7, the capacitor is charged by connecting it in series with the DC source and the load, thereby maintaining the capacitor voltage at half the DC source voltage. The other switching states are used to generate the different output voltage levels, ranging from -2E to +2E.

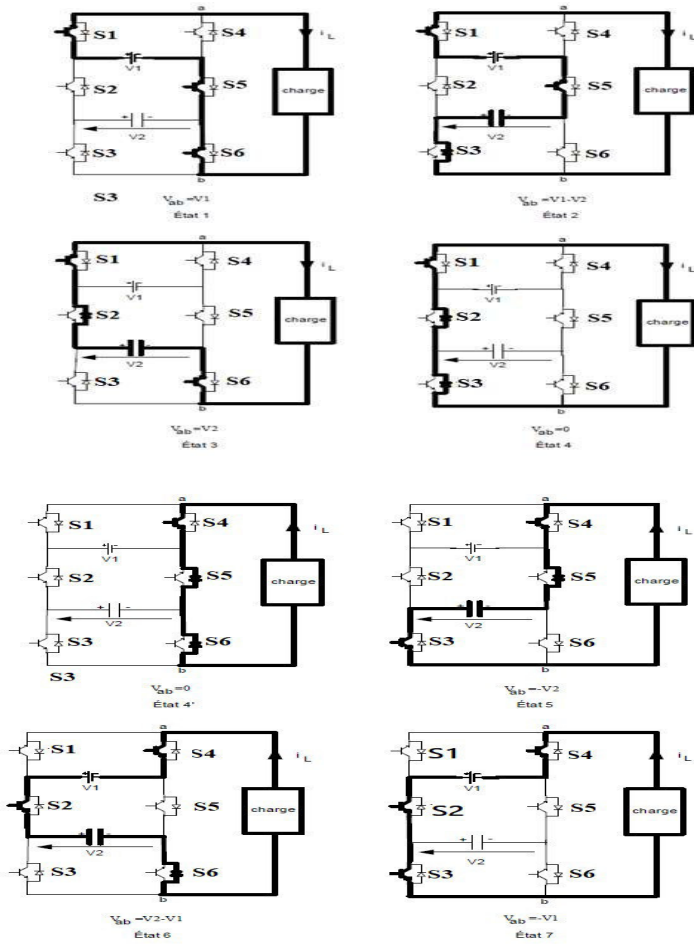


Fig.7. five-level PUC converter operating sequences.

Table2. Inverter switching table

States	S1	S2	S3	Output voltage	VI
1	1	0	0	V1	+2E
2	1	0	1	V1-V2	+E
3	1	1	0	V2	+E
4	1	1	1	0	0
5	0	0	0	0	0
6	0	0	1	-V2	-E
7	0	1	0	V2-V1	-E
8	0	1	1	-V1	-2E

III.2. The PUC Inverter PUC Sinus Triangle PWM Control has five levels

The PUC5 inverter control technique allows generating a five-level output voltage waveform using redundant switching states, while keeping the capacitor voltage fixed at half the DC source voltage thanks to judicious balancing of the charges and discharges of the capacitor. Specifically, six switching states are used to produce three voltage levels (-E, 0, +E), thereby creating redundant paths to flow current through the load and generate the five output voltage levels (0 , ěE, ě2E) [4].

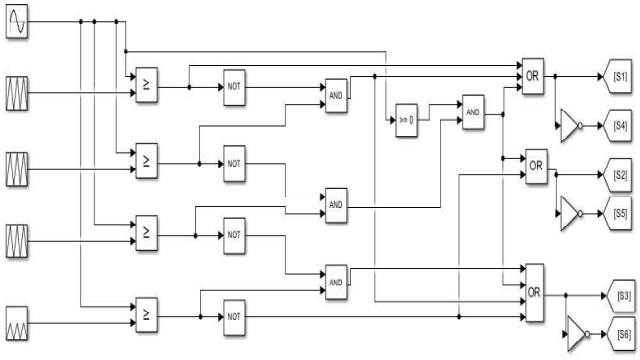


Fig.8. PUC5 inverter PWM control block

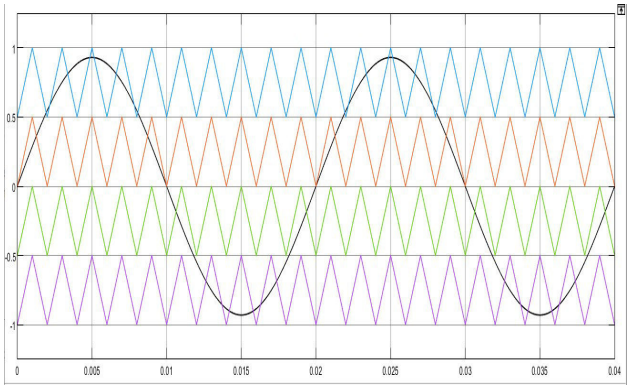


Fig.9. reference signal and the four carriers

III.3. Simulation of five-level PUC inverter

Table .3. Five-level PUC inverter simulation parameters

Voltage V1	200V
Capacitor capacity	3060 uF
Load resistance	60 ohm
Load inductance	40 mH

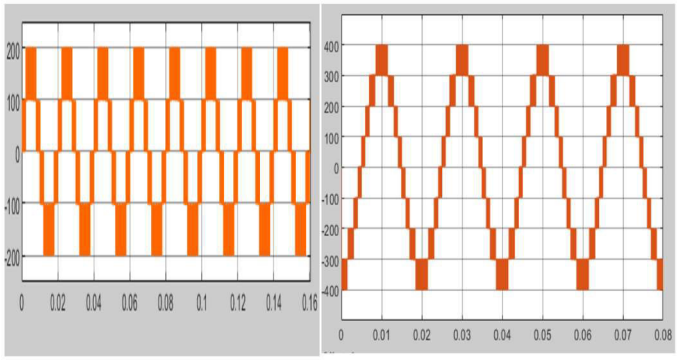


Fig.10. Simple and compound voltages for r= 1, Mf = 200

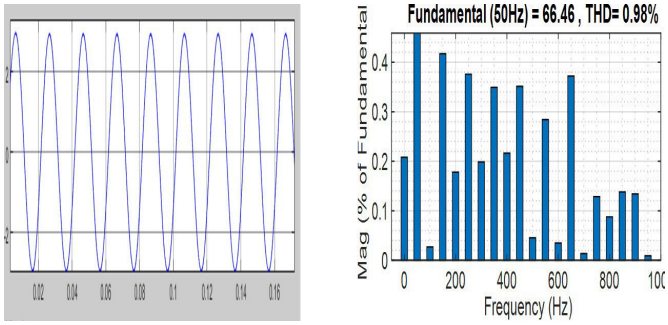


Fig. 11.The current waveform and its THD for r=1, Mf=200

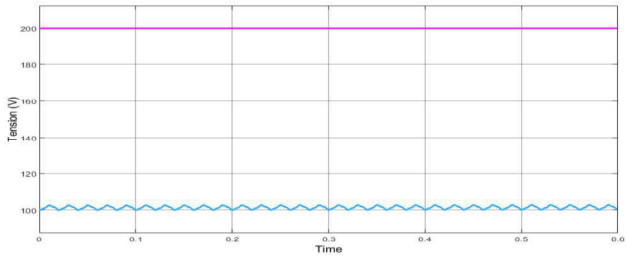


Fig.12. Source voltage (V1) and capacitor voltage (V2)

Table.4. The THD% values of PUC inverter current at five levels.

Rate of modulation	mf=100 (fp=5000Hz)			mf=200 (fp=10000Hz)		
Index of regulation	r=0.6	r=0.8	r=1	r=0.6	r=0.8	r=1
THD % of current	1.92%	1.32%	0.98%	1.18%	0.73%	0.59%

1. Interpretations of the results

The optimal modulation index allows maximum use of the available voltage, which produces an output waveform very close to the reference waveform.

The THD is at its minimum because the harmonic components are reduced to a minimum for this modulation frequency.

By setting the modulation rate at 100 (with a carrier frequency of 5000 Hz) and varying the regulation index from 0.6 to 1, we note:

ð By increasing the regulation index r from 0.6 to 1, the number of compound voltage levels increases from 7 to 9.

ð The amplitude of the current increases and approaches a sinusoidal form, which reduces its THD from 1.92% to 0.98%

By increasing the modulation rate to 200 while maintaining the same regulation index values, we see that the current THD values are smaller than those obtained with a modulation rate of 100.

The amplitude of the capacitor voltage (V2) is equal to half the amplitude of the setpoint (V1).

IV. Photovoltaic system connected to the electricity network:

On our setup, we interconnected 5 LG385N2C-G4 PV modules in series to increase the voltage and 66 in parallel to increase the current, thus forming an efficient photovoltaic system. This system is controlled by a boost chopper regulated by an MPPT P&O algorithm, allowing energy production to be continuously optimized by following the maximum power point. The generated energy is then stored in a storage battery for later use, ensuring continuous power supply even in the absence of sunlight [5].

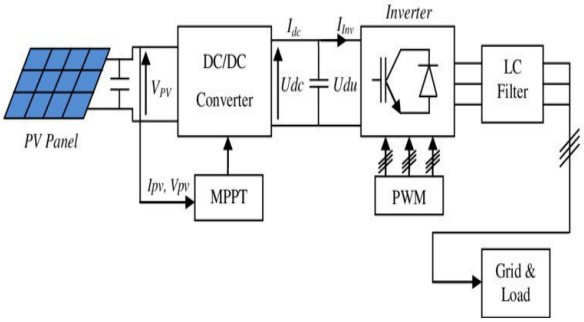


Fig.13. Synoptic diagram of a photovoltaic system connected to the electricity network.

IV.1. Simulation of the five-level PUC inverter connected to the closed-loop grid:

The five-level PUC converter is made up of six switches (fig.6) Each switch can only have two states, namely a state where it is blocked and another where it is saturated, which makes it possible to reach (23) eight states including two redundant. Table.2 shows the five voltage levels produced depending on the switch states.

V.1.1. Regulation loop:

The control of the PUC5 inverter for connection to the grid is based on a controller specially designed to regulate the amplitude and phase shift of the injected current, thus allowing active and reactive energy to be exchanged optimally with the grid [6].

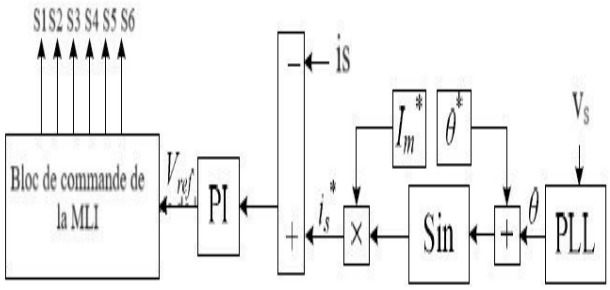


Fig.14. PUC5 network connected with controller

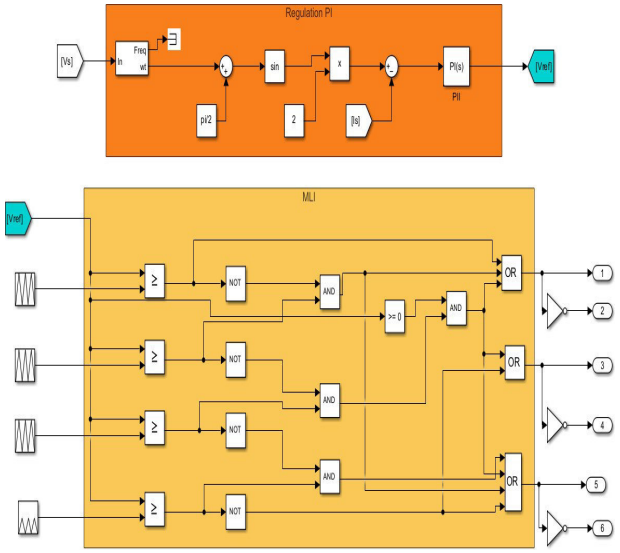


Fig.15. Closed loop PWM control block

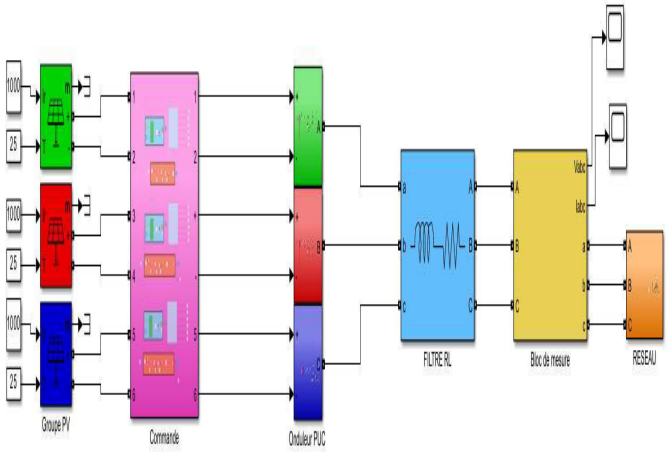


Fig.16. The five-level three-phase PUC inverter in grid-connected PV systems

IV.2. Simulation result

We simulated a three-phase five-level grid-connected PUC-type inverter controlled by closed-loop PWM with a PI regulator. The simulation parameters are shown in the table above.

Table.5. grid-connected PUC inverter simulation parameters

Network voltage	230/400V
Network frequency	50Hz
Resistance (ŕ(ĵ) of filter	10 ohm
Inductance (ĺ(ĵ) of filter	79.6 mH
Bus capacitor	3060 uF
Switching frequency Carriers	5 KHz
Switching frequency of the chopper	2.5 Khz

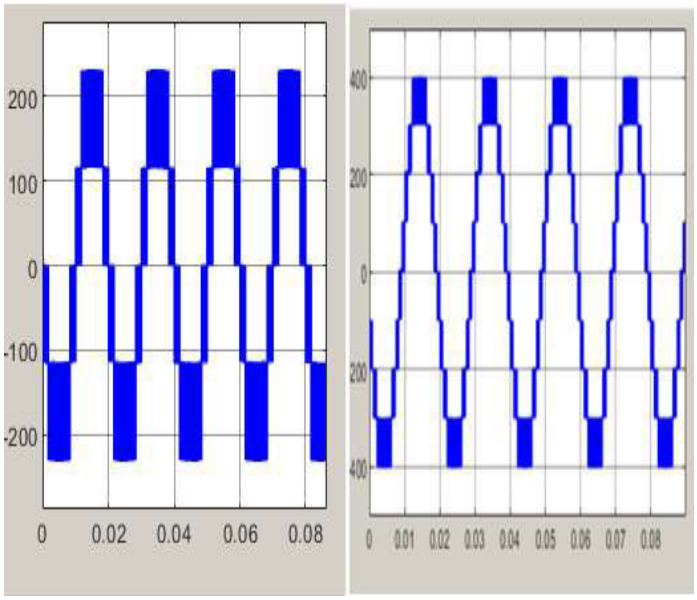


Fig.17. simple and compound voltages of PUC5

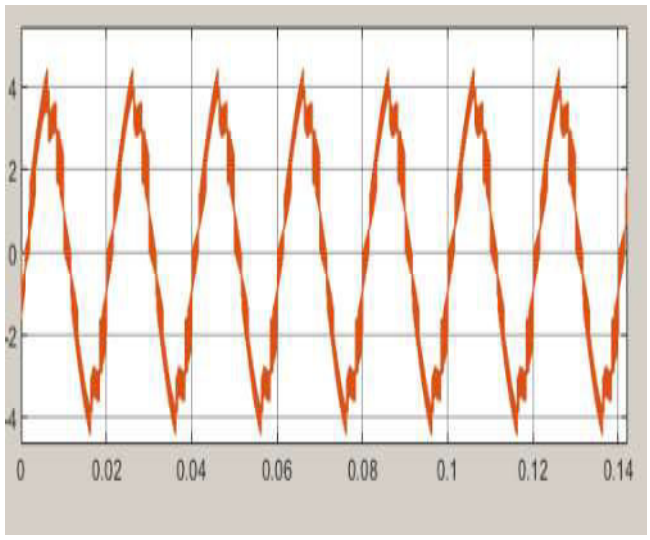


Fig.18. The current of phase Ia Before filtering

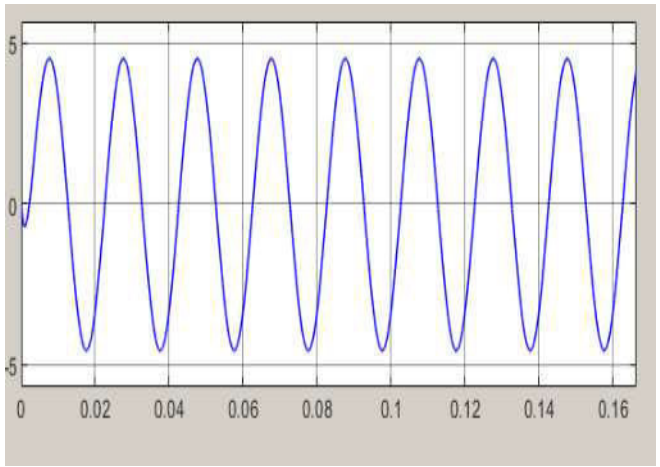


Fig.19. The current of phase Ia After filtering

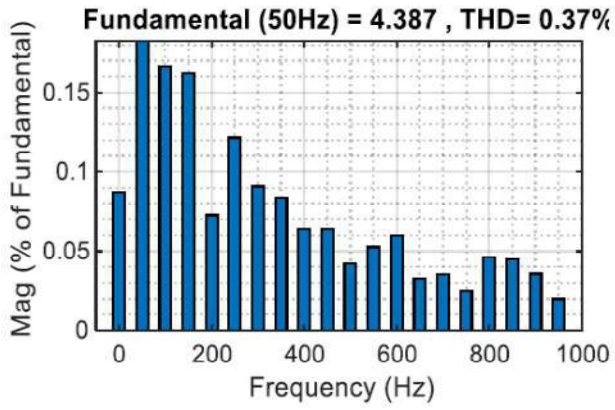


Fig.20.. Current THD

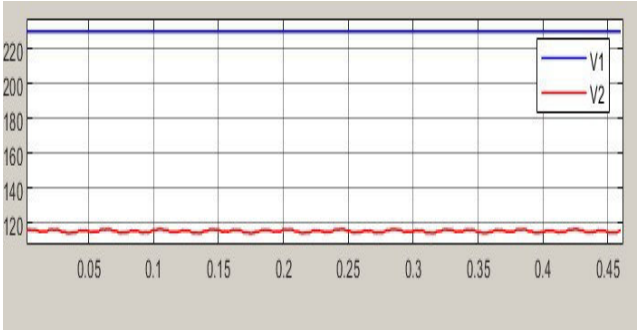


Fig.21. The source voltage (V 1) and the capacitor voltage (V 2).

Interpretations

The simple voltage carries the five levels: +2E, +E, 0, -E, -2E.

- The waveform of the current injected into the network, close to a sinusoid and with a low THD of 0.37%.

- The voltage of the capacitor (V 2) is maintained at half that of the setpoint (V 1).

Conclusion

We simulated the five-level PUC inverter, controlled by Pulse Width Modulation (Sine Triangle PWM), each time modifying the frequency modulation (MF) rate and the regulation index (r). The results show a simultaneous improvement in simple and compound voltages. Increasing the modulation rate and varying the regulation index allow an improvement in the shape of the current and a reduction in harmonic distortions. Filtering is used to adjust the waveform to approach the sinusoidal.

References

[1] N. T. Z, ĭ Lammouchi, IT Modeling and simulation of a photovoltaic system Master˜s thesis in Electrical Engineering, university of Echahid Hama Lakhder -D ˜el-oued ŏ ŏ. 2015.

[2] A. Chebana, ĭ Γ Control of a photovoltaic system connected to the network Master Thesis in Electrotechnics, university of Mohamed Khider -Biskra ŏ. 2014.

[3] A. Labouret et M. Villoz, ĭ Photovoltaic solar energy ŏ ŏ, The professional's manual, Edition DUNOD, aoát 2003.

[4] H. Vahedi et K. Al-Haddad, ĭ Real-time implementation of a seven-level packed u-cell inverter with a low-switching-frequency voltage regulatorŏ ŏ, IEEE Transactions on Power Electronics, vol. 31, no 8, p. 5967-5973, 2016.

[5] P.-A. L. H. Vahedi et K. Al-Haddad, ĭ Sensor-Less Five-Level Packed U-Cell (PUC5) Inverter Operating in Stand-Alone and Grid-Connected Modes ŏ, IEEE Transactions on Industrial Informatics, vol. 12, no 1, p. 361-370, February 2016.

[6] H. Vahedi et K. Al-Haddad, ĭ Real-Time Implementation of a Seven-Level Packed U-Cell Inverter with a Low-Switching-Frequency Voltage Regulator ŏ, IEEE Transactions on Power Electronics, vol. 31, no 8, p. 5967-5973, aoát 2016.

Control of a Doubly-Fed Induction Generator Under Fault Condition

Djamila Abdeslame
Department of Electrical Systems Engineering
University M'Hamed Bougaara
Boumerdes, Algeria.
d.abdeslame@univ-boumerdes.dz

Mokrane DEHMAS
Laboratory Signals and Systems,
Institute of Electrical and Electronic Engineering,
University M'Hamed Bougaara
Boumerdes, Algeria.
mdehmas@univ-boumerdes.dz

Abstract— **This work employs the maximum power point tracking (MPPT) method for maximum power extraction. Two control configurations have been proposed and examined. The first controller lacks a power loop, whereas the second one incorporates a power loop along with two supplementary proportional-integrator regulators. A cascade comprising two double-stage pulse-width modulation (PWM) converters is associated with the generator.**

The simulations, conducted using MATLAB, illustrate the behaviour of the employed dual-feed induction machine and focus on parameters characterising the machine, including active and reactive power and electromagnetic torque. The analysis considers the absence and presence of disturbances (fault event) with the two controllers. It was observed that the controller with a power loop demonstrated qualitative results, as the assumed disturbances did not significantly impact the machine parameters throughout the fault duration.

Keywords— DFIG, Vector control, MPPT control, Indirect field oriented control (IFOC).

I. INTRODUCTION

Injection of wind-generated electricity into power grids causes disturbances to the grid, and inversely, because of the wind speed irregularity. These disturbances include voltage drops in the power network.

Before the introduction of standards governing wind energy generation, the protection plan was based on the disconnection of the wind farm from the power grid in case of fault event, such as a voltage drop at the connection between the farm and the power grid to less than 0.85 of Un [1-3].

It is essential for wind farms to be equipped with fault detection systems and have the ability to disconnect in voltage drop events. This issue has led to the introduction of new solutions to be applied in voltage drop situations. The connection of variable-speed wind turbines (DFIM and SM structures) to the grid has become much smoother due to the use of power electronics systems and networks.

II. WIND SPEED MODELING

The dynamic properties of the wind are fundamental to the study of the whole energy conversion system, since under optimum conditions the wind power depends on the cube of the wind speed.

a. *Model of the turbine under study*

Knowing the turbine speed, the aerodynamic torque is therefore expressed by:

$$C_{aer} = \frac{1}{2} C_p \, \rho S \, V^3 \frac{1}{\Omega_t} \quad (1)$$

Based on the above equations, the dynamic model of the turbine can be illustrated by the diagram of Fig.1:

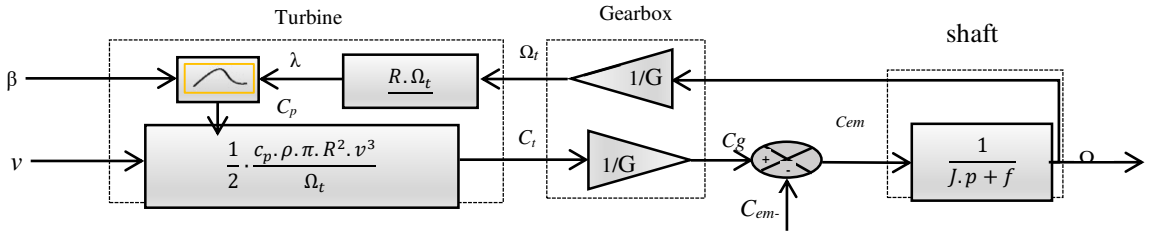


Fig. (1). Block diagram of the wind turbine

The simulations were carried out using the MATLAB/SIMULINK simulator [4]. The wind average speed is set to 7.7 m/s. Fig. (2) to (5) show the wind profile of the considered site, the rotor speed and the power coefficient for different pitch angles and the speed multiplier gain, respectivel

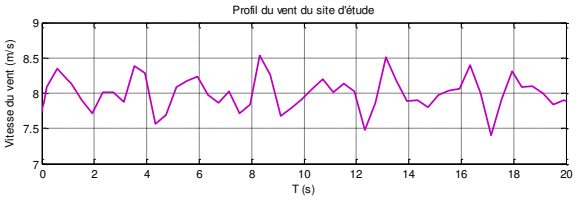


Fig 2. Wind profile of the considered site

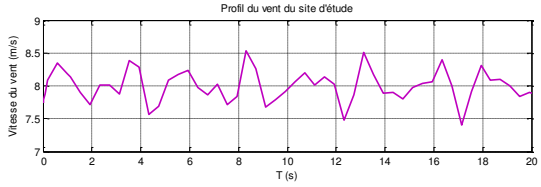


Fig 2. Wind profile of the considered site

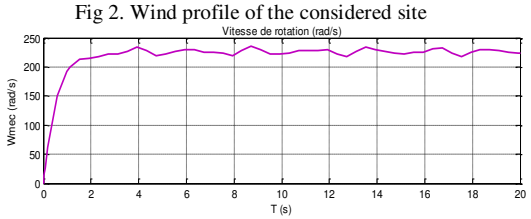


Fig . 3. Rotor speed

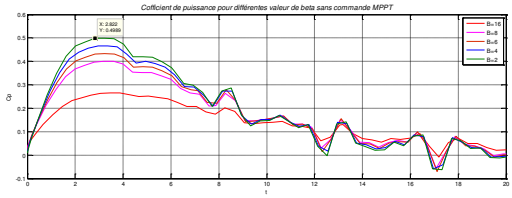


Fig. 4: Power coefficient for different values of the stall angle without MPPT control and for G=5.4

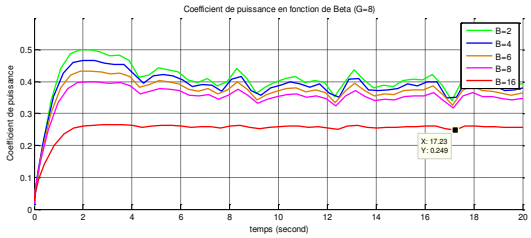


Fig. 5. Power coefficient for different pitch angles without MPPT control and for G=8

The simulation results show that rotational speed match with the wind speed variation. This shows the impact of the variation of the mechanical speed as a function of the wind speed on the generated electrical power. Fig.(4) and (5) show the variation of the power coefficient Cp as a function of the pitch angle for different multiplier gains. The results show that for a pitch angle, β = 2, Cp reaches 0.48 at 3s for G = 5.4, after which Cp decreases. However, for G = 8 and β = 2, Cp reaches its maximum level, and decreases for higher pitch angles.

III. TURBINE CONTROL USING MPPT METHOD

The most commonly used control system for conventional wind turbines directly connected to the power grid is mainly based on maximum power point technology (MPPT) [5], without centralized monitoring. They need to be disconnected from the grid when a fault occurs. In order to extract a maximum generated power, it is necessary to set the optimal wind speed ratio, λoptim, to the maximum power coefficient Cp_max.

Figure (6) illustrates the MPPT bloc diagram without the wind speed measurement.

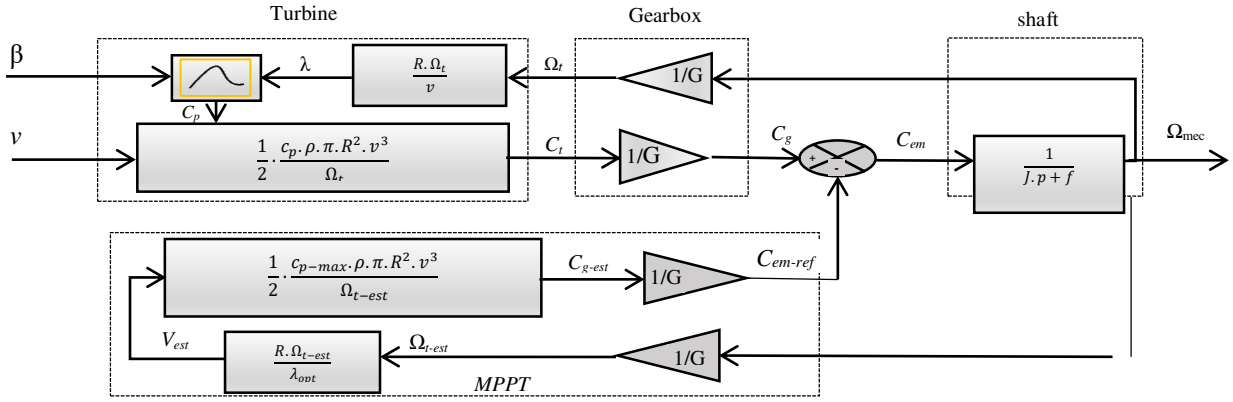


Fig. 6. MPPT bloc diagram without the wind speed measurement

Using the MPPT strategy for the whole system, and for the same wind profile, the rotor speed and the power coefficient are shown in figures (7) and (8) respectively.

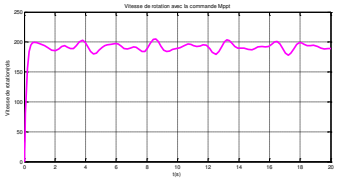


Fig. 7. Rotor speed using MPPT control

.

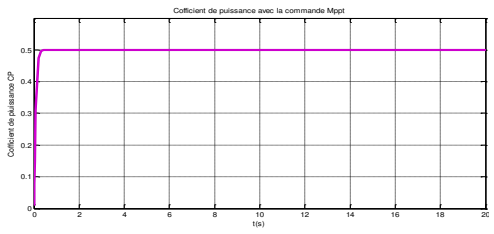


Fig. 8. Power coefficient using MPPT control

Figure (9) shows that the power coefficient converges to its maximum theoretical level. This factor is obtained for a fixed pitch angle β of 2°, which yields an optimum λ of 9.2. We can see that the Cp curve reaches its maximum value with and without wind variation denoting that the MPPT control is reliable. This control method converges to the optimum conditions by measuring the rotation speed, without using a wind speed sensor, and simplifies significantly the maximum power search algorithm.

IV. MADA modeling and control

After having studied the mechanical part of the wind turbine, we will now focus on the DFIFG operation in a generator mode as well as the power converters.

Variable speed wind farms (DFIG and SM structures) are being more performant with power electronics based interfaces. The main feature of the widely used DFIG structure is that it consists of an asynchronous machine with a wound rotor, and can deliver active power to the grid via the stator and the rotor. When directly connected to the grid, the rotational speed must remain practically constant in order to keep the machine close to the synchronous speed. This limitation reduces the wind turbine efficiency at high wind speeds [6]. To overcome this problem, a solution consists on adding a static converter between the machine and the grid. The device can then operate at a variable speed because the voltage is rectified whatever is the machine speed, and an inverter on the grid side matches the device and the grid frequencies.

This configuration allows achieving good efficiency over a wide range of wind speeds at a reasonable cost, as the power converters small in size [7]. On the other hand, the connection to the grid presents a significant challenge. The electrical grid is not stable and it is usually subject to disturbances. The latter can generate fault currents in the machine and, ultimately, damage the converters.

In this context, we are going to summerize the system operation and focus on the sensitivity to disturbances which usually cause faults in the network, in particular, short-circuits and voltage dips in the power grid. A voltage dip is a sudden drop in voltage amplitude lasting between half a period and one minute [8]. It is characterized by its depth (ΔV) and time duration (Δt).

- Indirect control

Is so called because it does not use a flux control loop and, therefore, does not require a sensor. In this control, the flux vector is estimated from the measurement of

- Voltage equations

The equation relating the electromagnetic torque to the stator flux and the rotor current is [9] :

$$C_{em} = p\frac{M}{L_s}\left(\varphi_{ds}.I_{qs} - \varphi_{qs}.I_{ds}\right) \quad (2)$$

Where p is the number of pairs of poles in the machine.

- Vector control of a double-feed induction machine

Flux-oriented control applied to electrical machines is used to obtain the desired operating mode by positioning the current vectors and the resulting flux vectors, [9].

By choosing a d-q two-phase reference frame related to the stator rotating field and aligning the stator flux vector φ_s with the d-axis, we can write:

$$\varphi_{ds} = \varphi_s \qquad \text{et } \varphi_{qs} = 0 \quad (3)$$

By choosing a d-q two-phase reference frame related to the stator rotating field and aligning the stator flux vector φs with the d-axis, we can write:

Equation (29) then becomes

$$C_{em} = P\frac{M}{L_s} I_{qr}\varphi_{ds} \quad (4)$$

- Relation between stator power and rotor currents

The stator active and reactive powers of an asynchronous machine in one of the two-phase references are expressed by:

$$\begin{cases} P_s = -V_s\dfrac{M}{L_s}I_{qr} \\ Q_s = \dfrac{V_s^2}{w_s L_s} - V_s\dfrac{M}{L_s}I_{dr} \end{cases} \quad (5)$$

The systems of equations allow illustrating the network to be controlled as shown in figure 9.

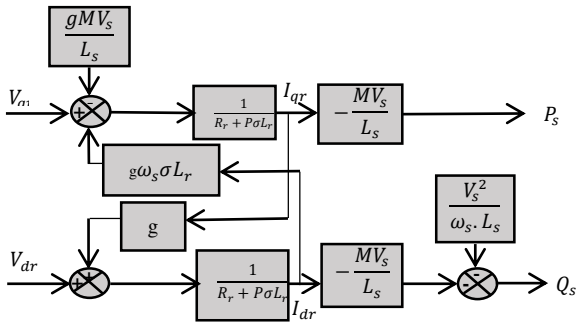


Fig.9. MADA block diagram

We observe that power and voltage quantities are related by first-order transfer equations. It is also noticed that we can use vector control because the control can be achieved independently along each axis. In order to maintain a uniform power factor on the stator; and thus optimize the power transfer to the grid, the reactive power setpoint is kept at zero. On the

rotor currents and speed, based on the machine rotor circuit equations.

a) Indirect control without power loop

Keeping the assumptions about the network stability, we obtain the control system illustrated in figure (10) where we have a rotor current control loop whose setpoints are deduced from the power levels assigned to the machine.

The rotor currents iqr and idr, which are the images of the stator active power Ps and the stator reactive power Qs respectively, follow their reference levels.

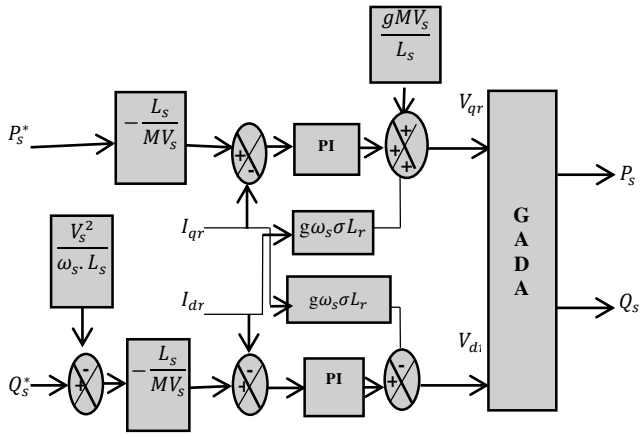


Fig. 10. Block diagram of indirect control without power loop

The simulation results for Indirect Control without a powerloop are shown in the figures below.

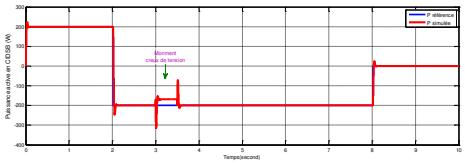


Fig.11. Active power

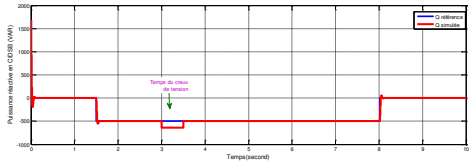


Fig.12. Reactive power in CIDSBP

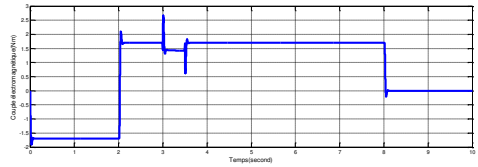


Fig.13. Electromagnetic torque

other hand, the active power setpoint must achieve optimal wind turbine power coefficient. The method, known as the indirect method, consists in taking into account the coupling terms and compensating them by using a system consisting of two loops to control the rotor powers and currents.

The obtained results show clearly that the Indirect Control method is more performant since it achieves perfect decoupling and a good fault tracking. When the fault is present, the error between the estimated and the reference quantities is visible because the power loop is missing. Instead, the rotor current remains almost invariant during the fault duration.Concerning the decoupling between the direct and the quadrature axes, it is observed that this does not appear in the reactive power graph. However, the influence of the reactive power variation is visible in the active power graph. Still, this influence is less significant in Indirect Control which does not have a power loop.

b) Indirect control with power loops

In order to improve the controller performance, we are going to incorporate an additional control loop at the power level in order to eliminate the static error while preserving the dynamics of the system. This results in the block diagram shown in Figure (14), which shows the two control loops for each axis; one controlling the current, and the other, the power.

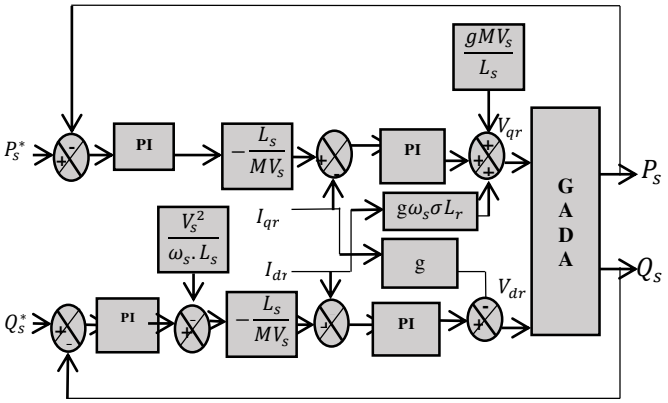


Fig. 14. Block diagram of indirect control with power loop

The simulated results for Indirect Control with Power loops are depicted in figures 15-17.

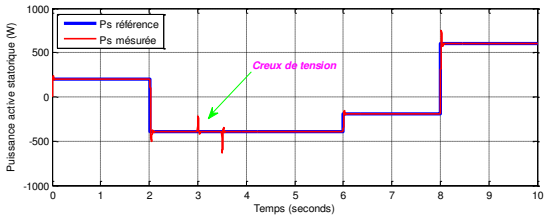


Fig. 15.Active power in CIDABP

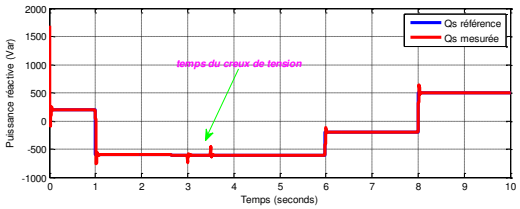


Fig .16. Reactive power in CIDABP

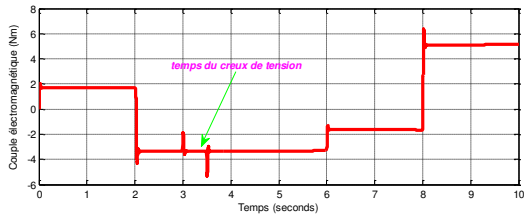


Fig.17. Electromagnetic torque in CIDABP

The obtained results show clearly the significant performance of the PI controller. Indeed, the active and the reactive powers follow their respective reference levels as illustrated in Figure 15 and 16, respectively, even in the presence of a fault. Furthermore, the rotor current and the electromagnetic torque do not vary during the fault event. Similar results are observed with indirect control without a power loop since the impact of the P step is much less significant with a power loop.

V. Conclusion

A brief description of the main steps involved in wind to electrical power conversion is first presented in this paper. The generated electrical power and its injection to a power grid network are also discussed along with the relevant limitations dealing, mainly, with fault events. The latter usually consists of voltage disturbances in the power grid system. Solutions based on power electronics interfaces have been proposed. These ones need to be driven by suitable and performant controllers.

In this work, the MPPT method for maximum power extraction is used. Two control configurations have been suggested and investigated. The first controller does not include a power loop, whereas a power loop is inserted in the second one along with two additional proportional-integrator regulators. A cascade consisting of two double-stage PWM converters is associated to the generator.

The simulations, carried-out using the MATLAB software, describe the behavior of the used dual-feed induction machine and concern parameters characterizing the machine such as active and reactive power and the electromagnetic torque. The analysis assumed absence and presence of disturbances (fault event), using the two controllers. It has been observed that the controller with a power loop exhibits qualitative results since the assumed disturbances have not induced any significant impact on the machine parameters throughout the fault duration.

As a further scope, it is suggested to use other controller types and achieve a performance comparative study with the investigated configuration.

References

[1] Heier, Siegfried. Grid integration of wind energy: onshore and offshore conversion systems. John Wiley & Sons, 2014.

[2] Ling, Y., & Cai, X. (2013). Rotor current dynamics of doubly fed induction generators during grid voltage dip and rise. International Journal of Electrical Power & Energy Systems, 44(1), 17-24.

[3] Najafi, Hamid Reza, and Farshad Dastyar. "Dynamic maximum available power of fixed-speed wind turbine at islanding operation." International Journal of Electrical Power & Energy Systems 47 (2013): 147-156.

[4] Environnement Matlab simulink , logiciel de calcul . https://www.mathworks.com/help/matlab/.

[5] E. Koutroulis and K. Kalaitzakis, "Design of a maximum power tracking system for wind-energy-conversion applications," IEEE transactions on industrial electronics, vol. 53, pp. 486-494, 2006.

[6] F. Poitiers, "Etude et commande de génératrices asynchrones pour l'utilisation de l'énergie éolienne-machine asynchrone a cage autonome-machine asynchrone a double alimentation reliée au réseau," Université de Nantes, 2003.

[7] D . Defontaines, "Etude et simulation de la MADA", Thèse de doctorat, École de technologie supérieure, 2012.

[8] J. C. Smith, G. Hensley, and L. Ray, "IEEE recommended practice for monitoring electric power quality," IEEE Std, pp. 1159-1995, 1995.

[9] A. Boyette, "Contrôle-commande d'un générateur asynchrone à double alimentation avec système de stockage pour la production éolienne," Université Henri Poincaré-Nancy I, 2006.

Adaptive Asynchronous Compressed Spectrum Sensing in Cooperative Networks

1st Hadj Abdelkader BENZATER
Laboratoire télécommunications
Ecole Militaire Polytechnique
Algiers, Algeria
d_BENZATER.HADJ_ABDELKADER@emp.mdn.dz

2nd Nacerredine LASSAMI
Laboratoire traitement du signal
Ecole Militaire Polytechnique
Algiers, Algeria

3rd Djamal TEGUIG
Laboratoire télécommunications
Ecole Militaire Polytechnique
Algiers, Algeria

4th Hamza Zeraoula
Laboratoire systemes lasers
Ecole Militaire Polytechnique
Algiers, Algeria

***Abstract*—This paper investigates a selective user cooperation strategy for cooperative compressed spectrum sensing in cognitive radio networks. By estimating the signal-to-noise ratio of compressed measurements for multiple secondary users, the proposed approach dynamically selects high-SNR users to participate in cooperative detection. This selection reduces the impact of noise from low-SNR users, significantly enhancing the quality and accuracy of sparse signal recovery. The methodology is validated through both simulations and an SDR-based implementation, demonstrating notable improvements in recovery accuracy and spectrum sensing performance, particularly in challenging, noisy environments.**

***Index Terms*—Cooperative Compressed Spectrum Sensing, Selective User Cooperation, Signal-to-Noise Ratio, Sparse Signal Recovery, SDR, Cognitive Radio Networks.**

I. Introduction

In modern communication systems, accurately estimating the Signal-to-Noise Ratio (SNR) is crucial for optimizing performance, particularly in environments characterized by noise and interference. This paper addresses the estimation of SNR for compressed vectors in cooperative sensing frameworks, where multiple nodes collaborate to improve signal detection and estimation. Using compressed sensing (CS) techniques, we aim to develop methods capable of identifying and eliminating the worst-performing channels, thereby improving overall system efficiency. Recent advances in deep learning and CS provide a robust foundation for this research, allowing for more accurate SNR estimations under varying conditions. CS has emerged as a powerful tool for acquiring and processing sparse or compressible signals with significantly fewer samples than traditional methods. In cooperative sensing scenarios, multiple nodes collaborate to enhance performance by sharing information and leveraging diversity. However, the presence of channels with low SNR can degrade the overall performance of the system.

In the context of cooperative compressed spectrum sensing (CCSS), several works have explored the estimation of SNR to enhance detection performance. These studies typically focus on improving the accuracy of SNR estimates, which are crucial for adaptive decision making in CS algorithms. Using techniques such as statistical modeling or machine learning, these approaches aim to optimize detection thresholds and improve the robustness of spectrum sensing in low-SNR environments.

The paper [1] presents a channel estimation technique for massive MIMO systems that minimizes pilot overhead by integrating compressive sampling matching with sparsity adaptive matching pursuit. Enhances normalized mean square error (NMSE) performance at high SNR while lowering computational complexity compared to traditional methods. The study [2] presents a method for estimating the unknown noise variance in CS problems, where measurements are often contaminated by additive noise. Accurately estimating the noise variance is important because it helps to improve the design of signal reconstruction algorithms. Authors in [3] propose low-complexity blind estimators for noise power, signal power, SNR, and MSE in multi-antenna mmWave systems, utilizing the sample median and sparsity in the beamspace representation of channels for efficient estimation.

In [4], the authors propose a new approach for primary user (PU) compressed detection in cognitive radio networks using an ensemble of machine learning classifiers, improving accuracy and robustness in multipath fading and hidden terminal scenarios.

Similar projects have already been completed. In this work [5], we examined the effectiveness of the effect of fusion rules on cooperative networks for CSS. The AND rule has the lowest false alarm rate P_f and a relatively high detection probability P_d for small networks, while the vote rule has the highest P_d for acceptable P_f according to **IEEE 802.22** cognitive radio standard. In addition, a virtualization strategy was proposed to improve SNR and size sensitivity performance. Another study [6] explores the use of the RMP recovery algorithm in conjunction with the AND rule that exhibits the lowest P_f for support tracking. The current literary works performed better under the suggested structure.

Both works in [5], [6] improve the detection performance of dense networks and require precise synchronization between channels for coherent integration. We are mostly interested in small networks for faisability and manageability purposes. To reduce complexity and improve performance at the same time, we proceed to a quick SNR estimation to sort channels and eliminate the worst one from the cooperation process. The estimation does not need to be highly precise since the exact SNR value does not matter too much. Estimating the SNR from two measurement periods can be a useful approach to isolate noise and improve sparse recovery in compressed sensing. This document outlines the method for estimating the SNR of two periods of measurements by subtracting them and isolating the noise component.

This method requires neither channel synchronization nor a large number of secondary users, with noticeable improvements even in small network configurations.

II. Compressive sensing

Compressed Sensing (CS) is an advanced signal processing method designed to reconstruct sparse or compressible signals using only a limited set of linear measurements. The key concept behind CS is its ability to take advantage of the signal's natural sparsity or compressibility when represented in an appropriate basis.

A sparse signal, which contains only a small number of non-zero components in a given basis Ψ^{-1} [7], can be compressed using the following expression:

$$y_{M\times 1} = \Phi x + b = \Phi_{M\times N}\Psi^{-1}_{N\times N} s_{N\times 1} + b_{M\times 1} \quad (1)$$

In this context, y represents the measurement vector, Φ is the sampling matrix, x is the signal sampled at the Nyquist rate, Ψ is the sparsifying matrix, b accounts for the noise, and s denotes the sparse vector. The compression is achieved through the sensing matrix $A = \Phi\Psi^{-1}$.

The goal is to reduce the size of the signal, achieving a sub-Nyquist sampling rate where $M \ll N$. Each measurement in the compressed vector represents a weighted sum of the components of the original signal. In the absence of noise, equation (1) becomes underdetermined, making it impossible to recover x or s directly from y. However, by exploiting the sparsity of s, characterized by its support $supp(s) = k \ll N$, it becomes feasible to reconstruct the signal. The sparsest solution can be found by solving:

$$\mathbf{s}^0 = \text{argmin}_{\mathbf{s}} |\mathbf{s}|_0 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{As} \quad (2)$$

Although solving for $|\mathbf{s}|_0$ is NP-hard, it can be approximated through convex relaxation by minimizing $|\mathbf{s}|_1$ [8], which is computationally feasible. The number of required measurements depends on the signal sparsity, the Nyquist rate, and the properties of the sampling matrix. Some common examples of measurement requirements include:

$$\begin{cases} M_{min} \simeq 1.32Klog(N/K+4)+3 & \mathcal{N} \text{ and } \mathcal{B} \\ M_{min} \simeq 1.7Klog(N/K+1)+1 & AIC \end{cases} \quad (3)$$

One of two requirements must be satisfied for optimal and robust signal recovery: low mutual coherence, which is the greatest correlation between the columns of the sensing matrix, or the Restricted Isometry Property (RIP) [7], [9]. Numerous sampling matrices, including Toeplitz matrices, circulant matrices, and random matrices, have been presented [10], [11]. Although random matrices are easy to construct, they require a lot of computing power and memory to store data. Toeplitz and circulant matrices, on the other hand, are structured matrices that require fewer measurements for signal sampling and are computationally efficient.

There are various types of sparse signal recovery techniques, including deep learning, convex relaxation, greedy algorithms, and Bayesian approaches [10]. Convex relaxation is based on linear programming, or one-norm minimization, which provides excellent precision at the expense of computational complexity. The Augmented Lagrange Method (ALM), Gradient Descent, and Least Squares are popular methods [12]. Greedy algorithms, which iteratively choose the most advantageous components based on the measurements, provide a simpler alternative to these methods due to their considerable complexity.

Although it breaks down the signal iteratively, the fundamental Matching Pursuit (MP) method may not be very accurate. By making sure the chosen components are orthogonal to the residual, Orthogonal Matching Pursuit (OMP) enhances MP and increases accuracy while controlling overfitting [13]. OMP still needs to be carefully adjusted and is susceptible to noise.

For wideband spectrum sensing in Cognitive Radio Networks (CRNs), Fast Matching Pursuit (FMP) is used in [14]. It achieves effective signal reconstruction at just 25% of the Nyquist rate, greatly speeding up performance compared to similar algorithms.

III. Cooperative Sensing

Through information exchange and the use of diversity, several nodes work together in cooperative sensing to enhance system performance as a whole. Improved coverage, precision, and dependability may result from this. There are two primary categories into which cooperative sensing falls:

- Data fusion: The nodes combine their individual measurements to obtain a more accurate estimate of the target signal.
- Distributed processing: The nodes collaborate to perform joint signal processing tasks, such as compressed sensing or detection.

The existence of channels with low SNR is one of the main obstacles in cooperative sensing. These channels have the potential to add noise and impair system performance as a whole. Consequently, identifying and reducing the impact of low-SNR channels is crucial.

By successfully addressing problems such as multipath fading, CSS increases spectrum sensing efficiency and detection accuracy. However, Cognitive Radios (CRs) need to

continuously identify Primary User (PU) activities and limit interference because spectrum availability is dynamic.

Cooperative spectrum sensing (CSS), in which users work together to improve detection reliability, has been the subject of research. To enhance fusion procedures and minimize shadowing effects, for example, Bouzegag et al. [15] presented Improved Square Law Combining (SLC) rules.

Furthermore, to enhance wideband spectrum sensing (WSS) at sub-Nyquist rates, the combination of collaboration and Compressive Sensing (CS) has been investigated. Single-node and multinode systems can benefit from two approaches that decrease complexity and increase noise resistance [16].

Secondary users (SUs) send autocorrelation data to a fusion center (FC) for decision making in a centralized CSS approach published in [17]. The Matching Pursuit algorithm is used to reconstruct signals, which reduces the impact of noise but increases complexity by requiring a high sampling rate for the computation of the autocorrelation matrix.

IV. Differential Measurement Technique for Noise-Reduced Signal Recovery

The two sets of compressed measurements, $\mathbf{y}_1$ and $\mathbf{y}_2$, are assumed to correspond to the same signal $\mathbf{x}$ but to have distinct noise realizations:

$$\mathbf{y}_1 = \mathbf{\Phi}\mathbf{x} + \mathbf{n}_1$$
$$\mathbf{y}_2 = \mathbf{\Phi}\mathbf{x} + \mathbf{n}_2$$

where $\mathbf{n}_1$ and $\mathbf{n}_2$ are noise vectors, $\mathbf{x}$ is the same signal and $\mathbf{\Phi}$ is the same measurement matrix. The result of subtracting the two measurements is the following:

$$\mathbf{y}_{\text{diff}} = \mathbf{y}_1 - \mathbf{y}_2 = (\mathbf{n}_1 - \mathbf{n}_2)$$

The noise difference $\mathbf{n}_1 - \mathbf{n}_2$ is all that remains after the signal $\mathbf{\Phi}\mathbf{x}$ is ideally canceled out. This makes it possible to estimate the noise power.

A. Noise Power Estimation

The noise difference $\mathbf{n}_1 - \mathbf{n}_2$ is a new noise vector, and its power can be calculated as:

$$P_{\text{new noise}} = \frac{1}{2}\|\mathbf{n}_1 - \mathbf{n}_2\|^2$$

Since $\mathbf{y}_{\text{diff}} = \mathbf{n}_1 - \mathbf{n}_2$, the noise power can be approximated as:

$$P_{\text{new noise}} \approx \frac{1}{2}\|\mathbf{y}_{\text{diff}}\|^2$$

B. Signal Power Estimation

Next, we estimate the signal power using one of the original measurements (assuming that both have the same signal):

$$P_{\text{signal}} = \|\mathbf{\Phi}\mathbf{x}\|^2 \approx \|\mathbf{y}_1\|^2 - \|\mathbf{n}_1\|^2$$

To estimate $\|\mathbf{n}_1\|^2$, we use the noise difference:

$$\|\mathbf{n}_1\|^2 \approx \frac{1}{2}\|\mathbf{y}_{\text{diff}}\|^2$$

Thus, the signal power is approximated as:

$$P_{\text{signal}} \approx \|\mathbf{y}_1\|^2 - \frac{1}{2}\|\mathbf{y}_{\text{diff}}\|^2$$

C. SNR Estimation

Finally, the SNR can be estimated as:

$$\text{SNR} = 10\log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{new noise}}}\right)$$

where P_{signal} and $P_{\text{new noise}}$ are the estimated signal and noise power, respectively.

D. Considerations

1) Signal Stability: This method assumes that the signal $\mathbf{x}$ remains constant over the two measurement periods. If the signal changes, the subtraction will not perfectly cancel it, affecting the noise estimate.

2) Independent Noise: The method also assumes that the noise vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ are independent and identically distributed (i.i.d.) with the same variance. Correlated noise can bias the SNR estimate.

E. Use of SNR in Sparse Recovery

- **Noise-aware algorithms**: Recovery performance can be enhanced by algorithms such as Lasso or Basis Pursuit Denoising, which use the noise level to better tune their parameters.
- **Adaptive thresholding**: SNR estimation aids in establishing suitable thresholds for detecting significant coefficients, enhancing recovery accuracy in techniques like Matching Pursuit.
- **Stopping criteria**: In order to prevent overfitting of the noise, iterative algorithms can utilize the noise level to decide when to stop.
- **Cooperation**: When making decisions, we can choose the optimal channels in cooperative networks.

V. Results

A sparse spectrum wide band signal was used to drive the simulation. When one Secondary User (SU) is removed, the network is large enough to retain the cooperation hypothesis. The AND rule serves as a hard fusion rule, and a greedy algorithm is used for recovery. The average SNR of the appropriate number of channels L on the y axis is represented by the SNR axis (x axis) in the 3D curves of Figure 1. The SNR fluctuation is the only difference between the first and second lines. $\Delta SNR = 1dB$, and $10dB$ are the expected values.

The recovered support portion of the cooperative compressed sensing technique is shown in Figures 1.a and 1.d. When removing the worst channel, both figures perform better in terms of cooperative detection probability Q_d. Cooperation in the original network (L-sized) is shown by the colored map, while the reduced network (L-1) is represented by the red grid.

Figures 1.b and 1.e demonstrate the improvement in Q_d. It is clearly seen that there is no negative value, indicating that the strategy is always effective. In medium SNR (around $6dB$) and smaller networks, greater improvements are discernible. The cooperative false alarm rate Q_f is shown in the final two figures, 1.c and 1.f, for both fluctuation scenarios according to

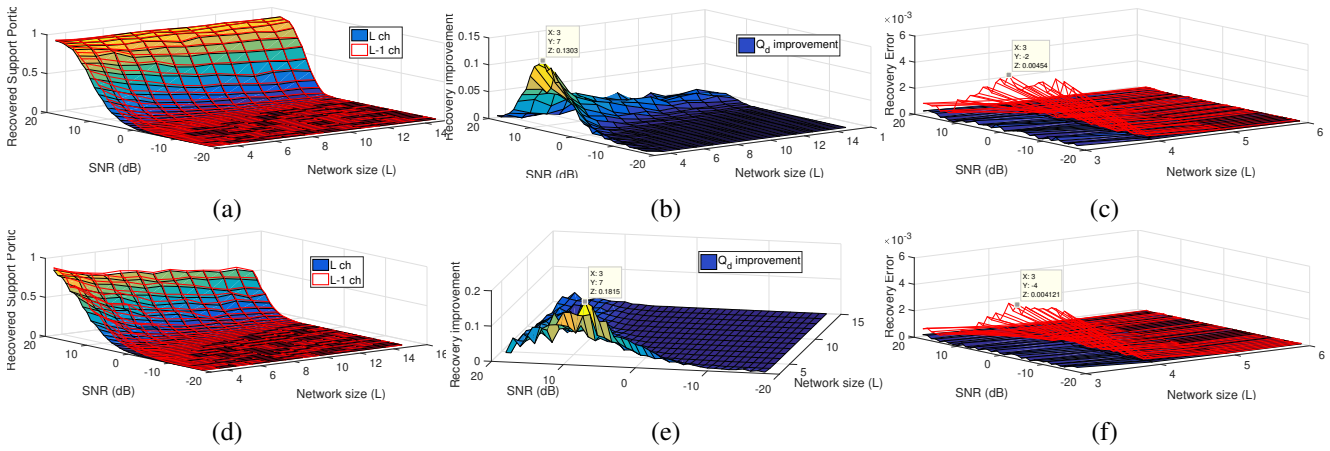


Fig. 1. Cooperative detection probability and false alarm rate: (a) Q_d of L Vs L-1 channels (Δ SNR=1dB). (b) Q_d improvement (Δ SNR=1dB). (c) Q_f deterioration (Δ SNR=1dB). (d) Q_d of L Vs L-1 channels (Δ SNR=10dB). (e) Q_d improvement (Δ SNR=10dB). (f) Q_f deterioration (Δ SNR=10dB).

SNR and L. Transformation degradation is comparable to the new Q_f of the smaller network because the AND rule initially shows a relatively low Q_f. The maximum value complies with the previously specified **IEEE 802.22** CR requirement since it is less than 5×10^{-3}.

VI. SDR Validation

As shown in Figure 2, we realized a framework in accordance with the methods described in [6] in order to validate the suggested strategy. Eight receivers (USRP B210), a centralized fusion center, and a single transmitter (HackRF One) make up the network included in this framework. The frequency, phase, and timing of the system are carefully adjusted to provide accurate cooperative detection and precise synchronization.

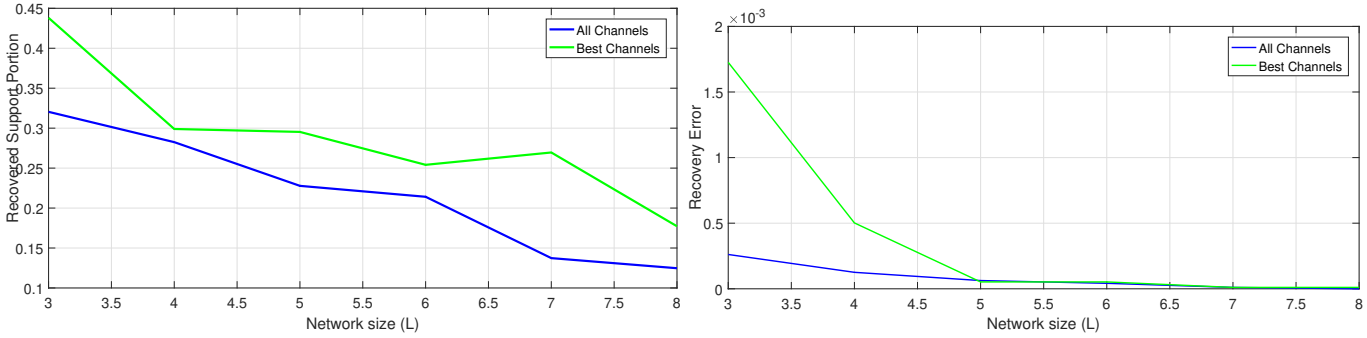


Fig. 2. Manipulations with Software Defined Radios (SDRs).

- Frequency Calibration: We used frequency calibration to ensure alignment at the receiver end and prevent frequency offset across receiving channels. This calibration process makes sure that frequency shifts will not interfere with signal detection, which is important for cooperative spectrum sensing.
- Phase Calibration: To align every channel with a specified reference channel, a phase correction mechanism is incorporated into the phase calibration procedure. Phase differences across receivers can reduce detection accuracy, particularly in cooperative setups where combined signals are examined; therefore, this alignment is crucial for coherent detection.
- Timing and Synchronization: A two-tiered clocking mechanism is used to produce synchronization. First, to guarantee alignment at a lower level, each pair of channels shares synchronization via the motherboard clock. Second, by coordinating the timing for every channel in the network, the central PC clock keeps the four motherboards in sync. Since any time misalignment could result in decreased cooperative detection efficiency, this multilayered synchronization strategy is crucial for efficient data fusion at the fusion center.

We used cooperative compressed sensing several times during successive periods of compressed measurements in order to apply Monte Carlo averaging to ensure accuracy. Because the SNR varies over time, the curves may not match those seen in simulations. As a result, the false alarm rate and detection probability are averages over several realizations.

In terms of performance, Figure 3(a) demonstrates that there is a significant performance gap across all network sizes, with the cooperative detection probability for L channels being lower than that of the $L-1$ best channels. This finding



(a) (b)

Fig. 3. Detection performance: (a) Q_d of L Vs L-1 channels. (b) Q_f of L Vs L-1 channels.

demonstrates how choosing the most dependable channels can increase detection accuracy. Meanwhile, the cooperative false alarm rate is shown in Figure 3(b), where a small decline is observed for smaller network sizes (3 or 4 users). However, this rise stays within the IEEE 802.22 standard's permitted bounds, showing that the framework retains a high degree of dependability even in the face of slight variations in false alarms. The resilience of the cooperative detection framework under varied situations is thus demonstrated by this SDR-based implementation, which confirms the suggested method in a real-world setting.

VII. Conclusion

In this study, we introduced a novel cooperative compressed sensing (CCS) method designed for cognitive radio networks (CRNs) that function in difficult environments like deep fading and high shadowing. Significant performance increases in compressed spectrum sensing were shown in early simulations driven by a sparse wideband signal, especially in challenging scenarios like the removal of a secondary user (SU) from the network. After these simulations, as detailed in the SDR Validation section, we used a Software-Defined Radio (SDR)-based implementation to evaluate our methodology in an actual environment. With exact frequency, phase, and timing calibrations to guarantee synchronized, accurate cooperative detection, this useful framework used a HackRF One transmitter, eight USRP B210 receivers, and a centralized fusion center. The SDR-based validation strengthened our approach's potential for practical implementation by confirming its effectiveness and robustness. Key findings from simulation and SDR-based assessments were as follows:

- Robust Cooperative Detection: Regardless of network size, the technique's capacity to maintain high detection accuracy in CRNs is demonstrated by the cooperative detection probability (Q_d), which continuously improved under all network conditions, including the removal of the least dependable channel. The SDR validation echoed the dependability observed in simulated environments by confirming that the cooperative CSS technique maintained strong detection performance across network sizes under real-world situations.
- Performance in Moderate SNR Regimes: The suggested approach is well-suited for networks with limited resources, as evidenced by the positive increases in Q_d that were especially apparent in moderate SNR regimes (around 6 dB) and with lower network sizes. These results were further supported by the SDR validation, which demonstrated constant efficacy in settings with low and moderate SNR.
- False Alarm Control: With an SDR-confirmed rate well below the typical threshold of 5×10^{-3}, the cooperative false alarm rate (Q_f) stayed below the strict IEEE 802.22 requirements. The AND rule continuously maintained Q_f at acceptable levels, even slight changes brought on by SNR variations, guaranteeing a trustworthy depiction of spectrum occupancy free of excessive false alarms.
- Flexibility and Adaptability: By keeping only the most dependable channels, the architecture was shown to be flexible enough to adjust to variations in network size in both simulations and SDR validation. This flexibility is crucial in dynamic wireless contexts where the amount of available SUs might be impacted by things like movement, interference, or device limitations.

Even though the current study uses both simulated and real-world SDR implementations to evaluate the suggested CSS strategy, there are still a number of areas that require more investigation and improvement:

- Extending the SDR implementation to handle real-time processing constraints will be an important step in assessing the performance of the approach in live wireless environments.
- Developing adaptive algorithms that can respond to real-time changes, such as the addition or removal of SUs, will be critical for the framework's practical viability in flexible network topologies.
- Enhanced Fusion Techniques: Additional research into other fusion rules or hybrid decision-making approaches could improve the resilience of detection, especially in

heterogeneous CRNs with variable SU dependability.

- Energy Efficiency and Computational Optimization: It will be crucial to look at simpler recovery techniques or ways to decrease computational complexity, particularly for dense or large-scale CRNs with constrained resources.

In conclusion, even in the face of difficult circumstances like fading and shadowing, the suggested cooperative CSS method is supported by the combined simulation and SDR validation as a viable way to improve spectrum sensing in cognitive radio networks. This method has a great deal of promise for real-world CRN applications with continued study and hardware optimization, which will ultimately help next-generation wireless networks use spectrum resources effectively and dependably.

REFERENCES

[1] Zaid Albataineh, Khaled Hayajneh, Haythem Bany Salameh, Chinh Dang, and Ahmad Dagmseh. Robust massive mimo channel estimation for 5g networks using compressive sensing technique. *AEU - International Journal of Electronics and Communications*, 120:153197, 2020.

[2] Ryo Hayakawa. Noise variance estimation using asymptotic residual in compressed sensing. *APSIPA Transactions on Signal and Information Processing*, 12(1), 2023.

[3] Alexandra Gallyas-Sanhueza and Christoph Studer. Blind snr estimation and nonparametric channel denoising in multi-antenna mmwave systems. In *ICC 2021 - IEEE International Conference on Communications*, pages 1–7, 2021.

[4] Venkata Vara Prasad and Polipalli Trinatha Rao. Adaptive cooperative sensing in cognitive radio networks with ensemble model for primary user detection. *International Journal of Communication Systems*, 35(2):e4247, 2022. e4247 dac.4247.

[5] Hadj Abdelkader Benzater, Djamal Teguig, and Nacerredine Lassami. Enhanced cooperative compressive spectrum sensing in cognitive radio networks. *Transactions on Emerging Telecommunications Technologies*, 35(11):e70000, October 2024.

[6] Hadj Abdelkader BENZATER, Nacerredine LASSAMI, and Djamal Teguig. New scheme of cooperative compressed spectrum sensing. *Physica Scripta*, October 2024.

[7] Haoran Qi. *Compressive Sensing of Multiband Spectrum towards Real-World Wideband Applications*. PhD thesis, Queen Mary University of London, 2021.

[8] Carlos Ramirez, Vladik Kreinovich, and Miguel Argaez. Why l1 is a good approximation to l0: A geometric explanation. March 2013.

[9] Salma Benazzouza, Mohammed Ridouani, Fatima Salahdine, and Aawatif Hayar. A survey on compressive spectrum sensing for cognitive radio networks. In *2019 IEEE International Smart Cities Conference (ISC2)*, Casablanca, Morocco, October 2019. IEEE.

[10] Fatima Salahdine, Elias Ghribi, and Naima Kaabouch. A cooperative spectrum sensing scheme based on compressive sensing for cognitive radio networks. *arXiv preprint arXiv:1912.04935*, 2019.

[11] Amirhossein Molazadeh, Mehrdad Ardebilipour, and Zahra Maroufi. Sensing matrix design for wideband channel estimation in millimeter-wave hybrid multiple-input multiple-output system. *Transactions on Emerging Telecommunications Technologies*, 33(12):e4639, September 2022.

[12] Thomas Peters. Data-driven science and engineering: machine learning, dynamical systems, and control: by s. l. brunton and j. n. kutz, 2019, cambridge, cambridge university press, 472 pp., £49.99 (hardback), isbn 9781108422093. level: postgraduate. scope: textbook. *Contemporary Physics*, 60(4):320–320, September 2019.

[13] Ameni Mejri, Moufida Hajjaj, and Salem Hasnaoui. Structured analysis/synthesis compressive sensing-based channel estimation in wideband mmwave large-scale multiple input multiple output systems. *Transactions on Emerging Telecommunications Technologies*, 31(7):e3903, March 2020.

[14] Michael M. Abdel-Sayed, Ahmed Khattab, and Mohamed F. Abu-Elyazeed. Fast matching pursuit for wideband spectrum sensing in cognitive radio networks. *Wireless Networks*, 25(1):131–143, June 2017.

[15] Younes Bouzegag, Djamal Teguig, and Abdelmadjid Maali. An enhanced slc scheme for cooperative spectrum sensing system in cognitive radio. *International Journal of Circuits, Systems and Signal Processing*, 17:153–162, March 2023.

[16] Zhijin Qin, Yue Gao, Mark D. Plumbley, and Clive G. Parini. Wideband spectrum sensing on real-time signals at sub-nyquist sampling rates in single and cooperative multiple nodes. *IEEE Transactions on Signal Processing*, 64(12):3106–3117, June 2016.

[17] Osama Elnahas and Maha Elsabrouty. Cyclostationary-based cooperative compressed wideband spectrum sensing in cognitive radio networks. In *2017 Wireless Days*, pages 77–82, Porto, Portugal, March 2017. IEEE.

Towards the new object-oriented metrics calculation system for software fault prediction

MEKAHLIA Fatma Zohra
Faculty of Computer Science, MOVEP Laboratory
University of Science and Technology Houari-Boumédiène
Bab Ezzouar, Algeria
mekahlia.fzohra@yahoo.fr 0000-0002-7369-3878

CHABANE CHAOUCHE Rabah
Faculty of Computer Science
University of Science and Technology Houari-Boumédiène
Bab Ezzouar, Algeria
chabanechaoucherabah4@gmail.com

***Abstract*—Testing is a crucial phase in the software engineering development lifecycle, represented by a well-defined chain of tasks to ensure that software quality objectives are met. With the advent of artificial intelligence algorithms, carrying out various types of software testing has become a costly and complex phase, requiring a great deal of time and effort. Object-oriented programming (OOP) is an influential software development methodology, emphasizing objects to promote modularity and code reusability. In this context, the evaluation of object-oriented metrics has become crucial for measuring software quality, complexity and performance. Some researchers have proposed the use of software metrics to describe the characteristics of software evolution. and improve the defect prediction rate. To achieve this, we have proposed new object-oriented software metrics that monitor import and exception conflicts. We propose a formal mathematical model based on Java lookaround assertions, which enables these aspects of the code to be automatically detected and analyzed. This structured approach is useful for software development environments, as it makes it easier for developers to spot specific types of problem. Our job is to: 1- exploring object-oriented metrics for software fault prediction. 2- propose new metrics to better understand the contemporary challenges of software development.**

***Index Terms*—software metrics, object-oriented, regular expression, Java, lookaround assertions, software engineering, formal model, software fault prediction.**

I. Introduction

A. Motivation

Several works in artificial intelligence have shown that there is a correlation between the variables of object-oriented metrics and the increase in software fault prediction rates: [1] [2] [3] [4] [5] [6] [7] [8]. Since the detection and determination of software defects is a very delicate, difficult and costly phase compared to developers who spend a lot of time to target and correct them in the different phases of the software life cycle such as unit tests, functional tests, integration tests, etc. The transition to software defect prediction becomes essential because it allows to reduce the effort, time and resources to identify any type of software defects before the delivery of the product to the end customer. Generally, software defect prediction is based on machine learning models [9] [10] and software metrics [11] [12]. Which makes the emergence and development of new software metrics play a very important role in the quality and accuracy of software defect prediction [13] and what prompted us to study the paradigm of object-oriented programming to propose new metrics that represent fundamental concepts of OOP and that positively influence the prediction of software defects. Existing solutions suffer from lack of certain metrics that represent fundamental concepts in object-oriented software such as:

- Import conflicts of classes in a program. The development tools of the object-oriented software that exist do not control the imports of a project which can generate conflict errorscompared to general import (*).
- Failure to measure the degree of use of exceptions in a program that can help detect potential security risks.

Our goal is to increase the set of object-oriented software metrics by addressing the gap we have targeted in order to increase the performance as well as the accuracy of predicting software defects.

Regular expressions are a formalization of John von Neumann's automata theory that was later formalized by the mathematician Stephen Kleene. Afterwards, regular expressions were first implemented by Ken Thompson in a software with a patent on the proper use of pattern matching and of course this formalism has been reimplemented in many ways up to the present day. Today regular expressions can be applied to extract information from text at a very high level and we can say that regular expressions have exceeded the traditional mathematical limits. However, we can use them to search for tokens or particular elements in a large text. For example, detecting email addresses in a text editor in order to identify spam emails. Also, they can be used to replace target tokens in a text by other particular elements in the case of data cleaning for example. Generally, in arithmetic, we use operations such as * and / to construct expressions. As example: 5*4/2. In the same way, we can use regular operations to construct regular expressions while describing a regular language, called regular expressions. For example: $(a \bigcup b)^*$.

B. Contributions

This work presents a new approach to improving software defect prediction by introducing two new object-oriented metrics. The proposed metrics are designed to meet the challenges of software development by improving defect prediction, re-

ducing development time and optimizing code quality. Our main points are:

1. proposal of two new object-oriented metrics on import conflicts and exception conflicts

2. proposing a formal model based on regular expressions to capture key elements of Java code, including strings, numbers, method calls... The aim is to provide a comprehensive approach that can be integrated into existing software analysis tools.

To the best of our knowledge, this work is the first to propose new object-oriented metrics that measure fundamental concepts of object-oriented programming such as: import and exception conflicts proven by a mathematical model. Generally, existing works have proposed metrics that manage the general aspect of software, such as the number of lines of code; moreover, the metrics have been presented by a simple formula.

The paper is organized as follows. Related work was presented in Section 2.In section 3, we present our two metrics. In Section 4, formal semantics have been presented and we will end with a conclusion and future work in the section 5.

II. Related works

Since the emergence of the object-oriented paradigm in 1990s, several software metrics have been proposed in the literature that allow to evaluate the characteristics of an O.O software. In [14], the authors invented a suite of metrics that allow you to measure the complexity of the object-oriented code. Commonly called CK metrics are still references today. The metrics were validated using commercial systems written in C++ and SmalItalk. Subsequently, in [15] the authors propose a data model and terminology with illustration of the importance of CK.

In [16] we find the MOOD metrics (Metrics for Object Oriented Design) which measure certain characteristics of object oriented such as Encapsulation, Coupling, Inheritance and Polymorphism. In [17]the authors present their tool, which is based on MOOD metrics proposed: 1) Method Hidden Factor, 2) Attribute Hidden Factor, 3) Method Inheritance Factor, 4) Attribute Inheritance Factor, 5) Polymorphism Factor, 6) Coupling factor.

In [18] the authors presented through their surveys a comparative study between proprietary, free and open source tools that evaluate static and dynamic object-oriented metrics. Moreover, the metrics supported in dynamic metrics calculation tools are very limited such as inheritance, dynamic binding and runtime polymorphism.

The authors in [19] investigated the relationship between centrality measures and O.O-metrics in order to predict the propensity for failure in three aspects: fault-prone classes, fault severity and number of faults. As a conclusion, he finds that the use of O.O-metrics and centrality measures improves the prediction of fault-prone classes and the number of failures in a program. To carry out this study, the authors focus on 9 metrics that belong to three families: complexity, coupling, and size.

Regular expressions, also called regex [20], are formal patterns that allow the extraction of tokens or target character strings from a text or a computer program. A regular expression is represented by a set of metacharacters where each metacharacter represents a given meaning. For example ([ ]) : Match any one of the enclosed characters. Example: [abc] matches 'a', 'b', or 'c'. In the literature, we find several works which focus on the use of regular expressions as a formal solution which makes it possible to target a certain text or program code in order to carry out certain processing on the targeted area. In [21] the authors proposed the FungiRegEx software which is based on regular expressions and which allows proteomic research. The software retrieves real-time data on several species from the JGI Mycocosm database.

III. Proposed software metrics

nous avons suivi plusieurs étapes clés pour évaluer et proposer deux nouvelles métriques orientées objet afin d'améliorer la qualité et l'efficacité des logiciels orientés objet, which are:

A. Import Conflict (IC)

Import conflicts often occur during software development. However, available Java development tools do not provide a comprehensive analysis on this topic. Therefore, we proposed the IC metric to fill this gap. This metric helps to manage the import of a Java file with all possible scenarios, it provides insights into the complexity and class dependencies. IC retrieves all the imports in a special file and then classifies them into four distinct categories which are:

- Used Imports: the import is used at least once. Example: import java.util.*. We propose to modify the import by using only the class concerned.
- Not Used Imports: the import is never used. We propose to delete the import.
- Duplicate Imports: if the import is repeated at least once. we propose to remove the painful copies.
- Conflict Imports: two or more imports have one class in common. Example: import java.sql.date and import java.util.date. We propose to delete one of the two.

Impact in software engineering:

- A large number of imports can increase code complexity and extend compilation time, thus affecting software maintainability and performance.
- Better visibility of imports in a class improves its understandability and maintainability, reducing unnecessary code complexity.
- This metric gives insight into a project's imports and helps in detecting conflict errors even in the general import case (*) that the IDE does not detect.

B. Java Exception Analyser (JEA)

Help a detect potential security risks and strengthen the overall software security by evaluating this aspect of the code. This metric analyzes the Java project to extract exceptions and classifies them into two groups: default Java exceptions and

non-default exceptions (exception made by the user), and itself divided into two subsections:

- Runtime exception: exception that extends RunTimeException.
- Compile time exception: exception that doesn't extends RunTimeException.

Impact in software engineering:

- A high number of exceptions handled indicates attention to error handling and exceptional cases.
- A large number of potential exceptions can indicate excessive complexity in the code, which can make the code more difficult to understand, maintain, and debug.
- The metric is designed to improve code quality assessment by providing information about exception handling practices.

IV. Formal semantics

In this section we will present our mathematical semantics which is based on regular expressions with lookaround to formalize the six metrics proposed in Java. Let r be a regular expression and $\sum$ be an alphabet. The language represented by r denoted L(r) is a regular language. r is a sequence of symbols, like: union, concatenation, alphabet. To present our proposed engine, we will start by defining the syntax of our Java's regex:

A. Java's regex syntax

1) Quantifiers: Quantifiers are symbols for specifying how many times a pattern should appear in a regular expression. Our used quantifiers are:

- *: Matches 0 or more occurrences of the preceding pattern.
- +: Matches 1 or more occurrences.
- ?: Matches 0 or 1 occurrences.
- {n,m}: Matches between n and m occurrences..
- {n,}: Matches n or more occurrences.
- {n}: Matches exactly n occurrences.

2) Alternation: Alternation is a strong concept in regular expressions which allows you to specify several alternatives for the same pattern:

- Vertical Bar (|): Acts as a logical OR between patterns. **Example:** a | b matches 'a' or 'b'.

3) Character Classes: Represent sets of characters that describe specific search patterns and which are:

- Square Brackets ([ ]) : Match any one of the enclosed characters.
  Example: [abc] matches 'a', 'b', or 'c'.
- Ranges: Specify a range of characters.
  Example: [a-z] matches any lowercase letter. i.e. match any character that belongs to the specified set
- Negation: Use ^ to negate the character class. i.e. match any character that does not belong to the specified set.
  Example: [^0-9] matches any character that is not a digit.

B. Escaped characters

In regular expressions, some characters cannot be used directly because they have a special meaning. To use them, we will need to escape them with an escape character (\):

- \d : Matches any digit, equivalent to [0-9].
- \D : Matches any non-digit, equivalent to [^0-9].
- \w : Matches any word character (alphanumeric plus underscore), equivalent to [a-zA-Z0-9_].
- \W : Matches any non-word character.
- \s : Matches any whitespace character (spaces, tabs, line breaks).
- \S : Matches any non-whitespace character.
- \n : Matches new line .

1) Special Characters:

- Dot (.) : Matches any single character except newline (\n) .

2) Lookahead and Lookbehind: Regular expression lookarounds are very practical for checking the presence or absence of a left or right subexpression in relation to the current position while using constraints on the context in the searched pattern [22]. Furthermore, if the regular expression engine tests a lookaround, it does not advance in the text but rather it stays in its place, it can advance if and only if the condition defined in the lookaround is tested correctly. You should also know that there are two types of lookaround which are lookaheads and lookbehinds.

Lookaheads to test the presence or absence of a pattern ahead of the searched pattern by specifying conditions to check after the current position. On the other hand, lookbehinds to test the presence or absence of a pattern after the searched pattern by specifying conditions to check before the current position.

We also have positive and negative lookaheads. For positive lookahead checks that the expression must be found after the current position and without including this expression in the global match ((?=expression)). On the other hand, negative lookaheads ((?!expression)) checks if the expression is not found after the current position and therefore the opposite.

Definition: Let $\sum$ be an alphabet. Lookaround tests the presence or absence of a pattern just before or just after the searched pattern as follows:

- ((?=pattern)): it is positive lookahead. Asserts that what follows the current position in the string matches the pattern inside the parentheses.
  Example: a(?=b) matches 'a' only if it is followed by 'b'.
- ((?!pattern)): it is negative lookahead asserts that what follows the current position in the string does not match the pattern inside the parentheses.
  Example: a(?!b) matches 'a' only if it is not followed by 'b'.
- ((?<=pattern)): it is positive lookbehind asserts that what precedes the current position in the string matches the pattern inside the parentheses.

Example: (?<=a)b matches 'b' only if it is preceded by 'a'.

- ((?<!pattern)): it is negative lookbehind asserts that what precedes the current position in the string does not match the pattern inside the parentheses.
  Example: (?<!a)b matches 'b' only if it is not preceded by 'a'.

Example:

Consider the regular expression \w+(?= email) searches any alphanumeric word that is followed by the word "email". Lookahead allows you to check the presence or not of the word "email" after the current position with non-inclusion in the global correspondence.

C. Language proposed

1) Type Parameter : Type parameter is used to build other regex like the genericity part of class and method detection regex.
Example: public class NumberBox <T extends Number> { public static <T extends Number> double sum(T[ ] array) {}}

L_1 = { Type Parameter Of Genericity }

- MultipleBoundPattern =(\s+extends\s+\w+(\s*<\w+>\s*)?(\s+&\s+\w+(\s*<\s*\w+\s*>)?)*\s*)?, matches one/multiple bounds in a Type Parameter extends NumberClass & <T >.
- TypeParameterGen = (\s*<\s*\w+(MultipleBoundPattern)\s*(,\s*\w+(MultipleBoundPattern)\s*)*\s*>\s*)?, matches Type Parameter <T extends NumberClass, k extends <V >>

2) Method Detection :

- IC(Import Conflict): it's needed to fetch the classes from the method prototype (return type , parameter type , exception thrown) because some of them need imports to be used.
  Example: public static ArrayList <ImportStatus>ImportFetch(File file){...}
- JEA(Java Exception Analysis): it needs to fetch the exception thrown in a method prototype.
  Example: public static int countOverrideMethods (Class<?>) throws FileNotFoundException{ ...}

L_2={Method Or Constructor Prototype}

- Bracket=(\[\s* \]){1,2} ,to match [ ] or [ ][ ] of arrays.
- ArrayDeclarationPattern=\w+\s+\w+\s*(Bracket) | \w+ \s*(Bracket)\s*\w+ \s*, to match 2d or 1d array declaration like : int[][] IntegrGrid, Student ArrayStudent [].
- ArrayTypePattern = \w+\s*(Bracket)\s*, to match 1d or 2d array type without name of the array example:int [].
- NormalPattern = \w+\s+\w+ ,to match simple type: int nb, float pi.
- WrapperClass = \s*\w+\s*, to match WrapperClass Inside the <>of a collection: Integer.
- WildCardGen = \s* \? (extends\s+ \w+ | super \s+ \w+)?\s*, matches wildcard genericity.
- SimpleInside = \s*(WrapperClass)\s* | \s*(ArrayTypePattern)\s* | (WildCardGen), to match either wrapper class or arrays inside <>of a collection.
- InsideCollection = (SimpleInside) | \s*\w+<\s*(SimpleInside)\s*> | \s*\w+\s*<\s* (SimpleInside)\s*,\s*(SimpleInside)\s*>\s* ,to match double nested or normal inside of a collection.
- SetListPattern = \w+\s*<\s*(InsideCollection)\s*>\s* ,to match set and list collection.
- MapPattern = \w+\s*<\s*(InsideCollection)\s*,\s* (InsideCollection)\s*>\s* ,to match map collection.
- CollectionPatten = (MapPattern) | (SetListPattern).
- Paramter = \s*(NormalPattern)\s* | \s*(ArrayDeclarationPattern)\s* | \s*(CollectionPattern) \w+\s* ,to match collection, array and simple type.
- Arg = \(\s*((Paramter)(,\s*(Paramter))*)?\s*\) ,to match the Arguments including the parenthesis of a method it also matches no arguments.
- AcessModifier = (private\s+ | public\s+ | protected\s+)?.
- NonAcessModifierSimple = (static\s+ | final \s+ | abstract\s+)?.
- ModifierSimple = (AcessModifier) (NonAcessModifier-Simple) | (NonAcessModifierSimple)(AcessModifier).
- ModifierComplex = (AcessModifier)final\s+static\s+ | (AcessModifier)static\s+final\s+ | final \s+(AcessModifier)static\s+ | static\s+(AcessModifier) final\s+ | static\s+final\s+(AcessModifier) | final\s+ static\s+(AcessModifier).
- ModifierPattern = ModifierSimple | ModifierComplex.
- ThrowsPattern = (\s*throws\s+\w+\s*(\s*,\s*\w+\s*)*)?, to match single or multiple exceptions throws.
- CurlyBraces = (\{\s* | {\s*\}\s*)?, to match { or {}.
- ReturnType = \s*\w+\s+ | (collectionPattern) | (Map-Pattern) | (ArrayTypePattern) ,to match return type could be array like int[ ] , simple type : Integer , collection : List <ArrayList<String []>>.
- ConstructorRegex=(AcessModfier)(TypeParameterGen)\w+ \s*(Arg)\s*(ThrowsPattern)(CurlyBraces) ,to match constructor prototype : ImportStatus(String ImportName,int ImportStatus,int LineNumber) { ...}.
- MethodPattern = (ModifierPattern)(TypeParameterGen) (RetunType)\w+\s*(Arg) (ThrowsPattern) ((CurlyBraces) | \s*;\s*) , to match normal method prototype: static Set<String>FetchSrcPackageFile (String AsterixImport){ ...}.
- MethodPrototypePattern= (MethodPattern) | (ConstructorRegex), to match method prototype.

3) Import Detection: IC(Import Conflict): it's needed to fetch the imports from a Java file.
L_3 = {Import Line}

- StaticAccesModifier = (\s*static\s+)? , to match static access modifier for static import.
- ImportPattern=\s*import\s+(StaticModifier)\w+(\s*\.\s* (* | \w+))+\s*;\s*, to match static and normal import line : import static java.util.math.*; , import

application.BackEnd.RegularExpression;

4) Package Detection: it's there to avoid package statement while reading the Java file line by line since it doest hold any interesting data.

L_4 = {Package Line}

- PackagePattern =\s*package\s+\w+(\s*\.\s*\w+)*\s*;\s*, to catch package line : pacakge application; , package application.Backend;

5) Catch Detection:

- IC(Import Conflict): some exceptions inside catch statement needs to be imported to be used.
  Example: catch (FileNotFoundException | MalformedURLException | ClassNotFoundException e).
- JEA(Java Exception Analysis): metric needs to fetch the exceptions inside a catch statement.
  Example: catch(Exception e)

6) String Literal : It's used to build the method call regex since a string literal can be passed as a parameter.

L_5={String Literal with all concatenation possible }

- Char= [^"\n]+, matches any characters beside" and newline.
- StringConcatElement= (ClassVariable) | (MethodCall) | \w+ | "(Char)" | (NumbersPattern), matches methodCall: token.getToken(), ClassVariable: Student.Name, VarName: Age, or a String: "Hello World!", Numbers: 21, -32.21f.
- StringConcat = (StringConcatElement)(\+(StringConcatElement))*, this matches one/multiple concatenation with StringConcatElement.
- LiteralStringPattern = ((StringConcat)\+)?"((Char) | " \+(StringConcat)\+ ")"(\+(StringConcat))? , matches string literal with optional concatenation in the beginning middle and end : Age + "Is My Age and My Name is "+ Student.Name + "My Grade Is"+ 13.21f.

7) Numbers: The numbers regex is used to build the string literal regex since a number can be concatenated with a string literal , it's also used as parameter in method call.

L_6={Int And Float And Double}

- SignPattern = (\+\s* | \-\s*)?, to match sign of numbers: none , + , -
- FloatPattern = \s*(SignPattern)\d+\.\d+(f)?\s*, to match double and float.
- IntPattern = \s*(SignPattern)\d+\s* , to match integers.
- NumbersPattern = (FloatPattern) | (IntPattern), to match int, float and double.

8) Method Call: It's used to build the static call regex

L_7={Method Call}

- ClassCall = (\w+\.)+\w+ matches static call of a method or object call method : list.size.
- ClassVariable = \w+\.\w+ , matches class variable like Student.age.
- SimpleArgMethodCall = (ClassVariable) | (NumbersPattern) | \w+ | (LiteralStringPattern), simple argument inside of a method call (): String literal: "Hello world!", Numbers: 32, ClassVariable: City.inhabitant.
- SimpleMethodCall = \s*(ClassCall)\(((SimpleArgMethodCall)(\s*,\s*SimpleArgMethodCall)*)?\s*\)\s*, matches simple method call with no method call as parameter inside the(): list.size(),Student.grade(grade1, grade2, 12.90).
- Inside = (SimpleArgMethodCall) | (Class) \(((SimpleMethodCall)(\s*,\s*SimpleMethodCall)*)? \s*\)\s* | (SimpleMethodCall), inside of a method call is either numbers, variable, class variable, String Literal, nested method call, simple method call.
- MethodCall = \s*(ClassCall)\(((Inside)(\s*,\s*(Inside))*)? \s*\)\s* , matches double nested method call.

9) Throw Detection:

- IC(Import Conflict): some exceptions inside a throw statement needs to be imported to be used.
  Example: throw new IllegalArgumentException("Number must be positive").
- JEA(Java Exception Analysis): metric needs to fetch the exception of a throw statement.
  Example: throw new Exception("ERROR EXCEPTION HAPPENING").

L_8={Throw Statement}

- ThrowPattern = \s*throw\s+new\s+\w+\s*\(\s*(Inside)\s*\) \s*;\s* , matches throw statment :
  throw new IllegalArgumentException("Age must be 18 or older.");

10) Instanciation:

- IC(Import Conflict): to fetch the constructor since some of them needs to be imported to be used.
  Example: try (BufferedReader reader = new BufferedReader(new FileReader(file))).
- SC(Swing Component): to fetch the constructor of swing element.
  Example: mainFrame.setLayout(new BorderLayout());

L_9={Line That Contains Instanciation}

- NewPattern = .+ (\(| =)\s*new\s+ .+ , matches line of code that contains instanciation: try(BufferedReader reader = new BufferedReader(new FileReader(file))).

11) Variable:

- IC(Import Conflict): it's needed to fetch the classes from the variables(reference type, type parameter of a collection) because some of them need imports to be used.
  Example: public static ArrayList<ImportStatus >ListImport.

L_{10}={Variables}

- PatternAcessModifiers = (private\s+ | protected\s+ | public\s+)?, to matches acces modfiers: private, public, protected or none.
- StaticModifier = (static\s+)?, macthes static modifier.
- FinalModifier = (final\s+)?, matches final modifier.
- VarModifer = (PatternAcessModifiers) (FinalModifier) (StaticModifier) | (PatternAcessModifiers) (StaticModi-

fier) (FinalModifier) | (StaticModifier) (PatternAcessModfiers) (FinalModifier) | (StaticModifier) (FinalModifier) (PatternAcessModifiers) | (FinalModifier) (PatternAcessModifiers) (StaticModifier) | (FinalModifier) (StaticModifier) (PatternAcessModifiers), matches all the possible combination of the modifiers: static final, final private static, public final, etc.

- VariablePattern = (VarModifer) ((?!return\s+)\w+\s+\w+ | (ArrayDeclarationPattern) | (CollectionPattern)\w+)\s*(=\s*.+)?;?, matches Variables: List<Integer>NbList; int a = 0; etc.

12) Annotation: Import Conflict: needed to fetch the annotation because some of them needs an import to be used. Example: @FXML.

L_{11}={Annotation Besides Overload and Override}

- AnnotationPatten = \s*@\s*(?!(Overload | Override))\w+ \s*, this matches all anotation beside Override and Overload: @FXML

13) Static Call: IC(Import Conflict): some static call method and variable class need to be imported to be used. Example: for (ImportStatus Import: ImportController.ListImport) Encapsulation encapsulation = Encapsulation.EncapsulationFetch(file).

L_{12}={Line That Contains Static Method Call}

- StaticCallPattern= .+ (MethodCall) .+, this matches line of code that contains Static Call: ListImport= ImportStatus.update(file,(ImportStatus.ImportFetch(file)));

V. Conclusion and futur work

The aim of this paper is to present a new approach to improving software defect prediction by introducing two new object-oriented metrics. The proposed metrics are designed to meet the challenges of software development by improving defect prediction by artificial intelligence algorithms, reducing development time and optimizing code quality. We have proposed a formal model based on regular expressions to capture key elements of Java code, including strings, numbers, method calls, etc. The aim is to provide a comprehensive approach that can be integrated into software development tools. The aim is to provide a comprehensive approach that can be integrated into existing software analysis tools. The use of regular expressions to define patterns in Java code for each metric is well executed, providing a clear mathematical model that is both theoretically sound and practical for implementation. This model is also adaptable, with potential for extension to other programming languages or paradigms in the future.

Finally, we conclude with the future perspectives: First, propose a tool based on the proposed engine. Second, incorporate existing metrics into the software. Third, creation of an artificial intelligence model to predict software defects in a way that improves

References

[1] Fenton, Norman E. Software metrics: a practical and rigorous approach. International Thomson Pub., 1996.

[2] BASILI, Victor R., BRIAND, Lionel C.. , et MELO, Walcélio L. A validation of object-oriented design metrics as quality indicators. IEEE Transactions on software engineering, 1996, vol. 22, no 10, p. 751-761.

[3] KAUR, Arvinder et MALHOTRA, Ruchika. Application of random forest in predicting fault-prone classes. In : 2008 international conference on advanced computer theory and engineering. IEEE, 2008. p. 37-43.

[4] Rathore, Santosh S., and Sandeep Kumar. "An empirical study of some software fault prediction techniques for the number of faults prediction." Soft Computing 21 (2017): 7417-7434.

[5] Rathore, Santosh S., and Sandeep Kumar. "A study on software fault prediction techniques." Artificial Intelligence Review 51 (2019): 255-327.

[6] Turhan, Burak, Ayşe Tosun Mısırlı, and Ayşe Bener. "Empirical evaluation of the effects of mixed project data on learning defect predictors." Information and Software Technology 55.6 (2013): 1101-1118.

[7] Singh, Yogesh, Arvinder Kaur, and Ruchika Malhotra. "Empirical validation of object-oriented metrics for predicting fault proneness models." Software quality journal 18 (2010): 3-35.

[8] Bhattacharya, Pamela, et al. "Graph-based analysis and prediction for software evolution." 2012 34th International conference on software engineering (ICSE). IEEE, 2012.

[9] Miholca, Diana-Lucia, Gabriela Czibula, and Istvan Gergely Czibula. "A novel approach for software defect prediction through hybridizing gradual relational association rules with artificial neural networks." Information Sciences 441 (2018): 152-170.

[10] Majd, Amirabbas, et al. "SLDeep: Statement-level software defect prediction using deep-learning model on static code features." Expert Systems with Applications 147 (2020): 113156.

[11] Khleel, Nasraldeen Alnor Adam, and Károly Nehéz. "A novel approach for software defect prediction using CNN and GRU based on SMOTE Tomek method." Journal of Intelligent Information Systems 60.3 (2023): 673-707.

[12] Yu, Qiao, et al. "Process metrics for software defect prediction in object-oriented programs." IET Software 14.3 (2020): 283-292.

[13] Tete, Akpedje, Fadel Toure, and Mourad Badri. "Using Deep Learning and Object-Oriented Metrics to Identify Critical Components in Object-Oriented Systems." Proceedings of the 2023 5th World Symposium on Software Engineering. 2023.

[14] Chidamber, Shyam R., and Chris F. Kemerer. "A metrics suite for object oriented design." IEEE Transactions on software engineering 20.6 (1994): 476-493.

[15] Churcher, Neville I., and Martin J. Shepperd. "Towards a conceptual framework for object oriented software metrics." ACM SIGSOFT Software Engineering Notes 20.2 (1995): 69-75.

[16] Abreu, Fernando Brito, and Rogério Carapuça. "Object-oriented software engineering: Measuring and controlling the development process." Proceedings of the 4th international conference on software quality. Vol. 186. 1994.

[17] Abreu, F. Brito, Miguel Goulão, and Rita Esteves. "Toward the design quality evaluation of object-oriented software systems." Proceedings of the 5th International Conference on Software Quality, Austin, Texas, USA. 1995.

[18] Manju, and Pradeep Kumar Bhatia. "A Survey of Static and Dynamic Metrics Tools for Object Oriented Environment." Emerging Research in Computing, Information, Communication and Applications: ERCICA 2020, Volume 2. Springer Singapore, 2022.

[19] Ouellet, Alexandre, and Mourad Badri. "Combining object-oriented metrics and centrality measures to predict faults in object-oriented software: An empirical validation." Journal of Software: Evolution and Process 36.4 (2024): e2548.

[20] Sipser, Michael. "Introduction to the theory of Computation." Thomson Course Technology. Book. 2020.

[21] Terrón-Macias, Victor, et al. "FungiRegEx: A Tool for Pattern Identification in Fungal Proteomic Sequences Using Regular Expressions." Applied Sciences 14.11 (2024): 4429.

[22] Mamouras, Konstantinos, and Agnishom Chattopadhyay. "Efficient matching of regular expressions with lookaround assertions." Proceedings of the ACM on Programming Languages 8.POPL (2024): 2761-2791.

Performance Evaluation of Uplink and Downlink Cooperative NOMA Communications

Faical Khennoufa[1], Abdellatif Khelil[1], Safia Beddiaf[1], Walid Beddiaf[2]
[1]*LGEERE Laboratory, Department of Electrical Engineering, Echahid Hamma Lakhdar University, El-Oued, Algeria, email: khennoufa-faical,abdellatif-khelil, beddiaf-safia@univ-eloued.dz.*
[2]*LEVE Laboratory, Department of Architecture, Mohamed Khider University, Biskra, Algeria, email: w.beddiaf@univ-biskra.dz.*

***Abstract*—This paper examines an uplink and downlink cooperative non-orthogonal multiple access (NOMA). The model consists of four users, each distributed in two regions. Both users communicate with the other two in the second region through a relay node. The primary objective of this system is to enable efficient and reliable communication between the users in the two regions via the relay node. We obtain the bit error rate (BER), capacity, and outage probability (OP) expressions for the proposed system. Then, we compare our system with traditional orthogonal multiple access (OMA). After completing all verifications, we conduct computer simulations to generate the results. The findings indicate that the proposed system achieves higher performance gain compared to traditional OMA.**

***Index Terms*—Cooperative, downlink, NOMA, OMA, and uplink.**

I. Introduction

Non-orthogonal multiple access (NOMA) is widely renowned for increasing the number of users and improving spectral efficiency in next-generation wireless networks [1]–[4]. Besides, NOMA can be easily integrated with other physical layer schemes. Hence, the cooperative NOMA has recently been received a great deal of attention due to its ability to increase spectral efficiency, user fairness and expand network coverage [5], [6].

In [7], three types of cooperative NOMA schemes have been analyzed to enhance outage probability performance. The authors of [8], [9] derive the outage probability performance of the cooperative NOMA with direct links by adopting an amplify and forward (AF) relay. A buffer-aided cooperative NOMA system in the Internet of Things (IoT) has been investigated in terms of outage probability in [10]. In [11], cooperative NOMA is evaluated in terms of outage probability over the Nakagami-m fading channel. In [12], a new scheme of AF cooperative NOMA with direct links has been investigated in terms of outage probability. The delay-tolerant capacity, outage probability, and delay-limited capacity of a deivce-to-device-aided cooperative NOMA using the overheard signal are analyzed [13]. In [14], the system throughput of cooperative NOMA with direct links has been evaluated over the Nakagami-m fading channel. The authors of [15] examine the coverage probability and average data rate of user pairing with full-duplex (FD) cooperative NOMA networks. The authors of [16] analyze the outage probability of decode and forward (DF) cooperative NOMA over the Nakagami-m fading channel. In [17], the incremental cooperative NOMA has been investigated in terms of outage probability to improve the performance of conventional cooperative NOMA. In [18], the outage probability of FD/half-duplex (HD) uplink cooperative NOMA has been considered. The authors of [19] analyzed the ergodic capacity and outage probability for threshold-based DF selective cooperative NOMA. All previous works considered the perfect case of SIC and CSI, which is not realistic. The investigation needs to consider and evaluate the systems with more practical assumptions such as imperfect-successive interference cancellation (SIC), imperfect channel state information (CSI), in-phase and quadrature imbalance (IQI), etc. In [20], the outage probability and system throughput of DF cooperative NOMA have been examined under imbalance CSI and IQI, while in [21], the authors evaluated the outage probability of DF cooperative NOMA under imperfect SIC, imperfect CSI, and hardware impairments (HWI). On the other hand, the bit error rate (BER) of FD/HD cooperative NOMA has been analyzed in [22], [23]. In [24], the authors investigated the BER of the threshold-based DF selective cooperative NOMA to increase the data reliability of conventional cooperative NOMA. For more practical assumptions, under imperfect SIC, imperfect CSI, and HWI the BER of HD cooperative NOMA has been performed and evaluated in [21], [25]–[33].

The interaction between NOMA and cooperative communication is one of the topics that drew the most attention to achieve higher spectral efficiency and compensation for path-loss [4]. As mentioned above, cooperative NOMA assisted one relay has been investigated in terms of BER, ergodic capacity, and outage probability [7]–[33]. In this regard, to the best of the authors' knowledge, uplink and downlink uplink cooperative NOMA has not been investigated in terms of BER, capacity, and outage probability over the Nakagami-m fading channel. Therefore, this paper investigates uplink and downlink uplink cooperative NOMA systems in terms of BER, capacity, and outage probability over the Nakagami-m fading channel. The original contributions of this paper are as follows:

- We present uplink and downlink uplink cooperative NOMA system models. The model consists of four users, each distributed in two regions. Both users communicate with the other two in the second region through a relay node.

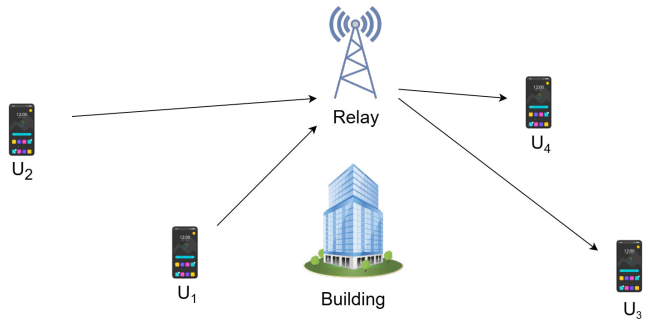


Fig. 1: Uplink and downlink cooperative NOMA system model.

- We obtain the BER, capacity, and outage probability over the Nakagami-m fading channel. We investigate the system using different channel parameters (i.e., shape parameters).
- Finally, the performance of the proposed system is compared with traditional OMA.

The paper is organized as follows. In section II, we present the proposed system model. In section III, we obtain the capacity and outage probability of the proposed system. Section IV presents the numerical results of the proposed system. Finally, section V concludes the paper with the conclusion.

II. System Model

We consider an uplink and downlink cooperative NOMA consisting of four users named as follows, U_1, U_2, U_3, U_4 and a relay as shown in Fig. 1. We assume that U_1 is far and U_2 is near from the BS. We assume that a relay node works in DF and HD modes.

In this system, We consider that the U_1 and U_2 communicate with U_3 and U_4 through the relay node. In the first phase, the U_1 and U_2 transmit their signal s_1 and s_2 simultaneously to the relay node. The received signals at relay node can be written as

$$y_R = \sqrt{P_1}\ h_1\ s_1 + \sqrt{P_2}\ h_2\ s_2 + \ n, \tag{1}$$

where h_1 and h_2 are the Nakagami-m fading channel coefficient[1], $n \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN).

The signals at the relay node are detected based on the channel gain i.e., $|h_1|^2 > |h_2|^2$, so the s_1 is detected first then using the SIC s_2 is detected. The MLD of s_1 symbols at the relay is designed by pretending s_2's symbols as noise, and it is given by

$$\tilde{s}_1 = \arg\min_m \{|y_R - \ \sqrt{P_1}\ h_1\ s_{1,m}|^2\}. \tag{2}$$

Thus, the signal-to-interference plus noise ratio (SINR) of s_1 at the relay node can be expressed as

$$\gamma_{R,s_1} = \frac{P_1|h_1|^2}{P_2|h_2|^2 + \sigma^2}, \tag{3}$$

However, in order to detect the s_2 at the relay node, a SIC should be implemented. The MLD of s_2 symbols at the relay node can be given as

$$\tilde{s}_2 = \arg\min_m \{|\acute{y}_R - \ \sqrt{P_2}\ h_2\ s_{2,m}|^2\}, \tag{4}$$

where

$$\acute{y}_R = y_R - \ \sqrt{P_1}\ h_1\ \tilde{s}_{1,m}, \tag{5}$$

Thus, the SINR of s_1 at the relay node can be expressed as

$$\gamma_{R,s_2} = \frac{P_2|h_2|^2}{\sigma^2}. \tag{6}$$

The relay re-encodes s_1 and s_2 in a superposition coding (SC) signal and transmits it to U_3 and U_4 according to the channel gain e.g., $|g_3|^2 > |g_4|^2$. The received signals at the relay node can be written as

$$y_i = \sqrt{P_t}\ g_i\ (\sqrt{\alpha_1}s_1 + \sqrt{\alpha_2}s_2) + \ n, \quad i = \{3, 4\}, \tag{7}$$

where α_1 and α_2 are the power allocation coefficient for U_i, $\alpha_1 > \alpha_2$, in which $\alpha_1 + \alpha_2 = 1$.

In the second phase, the MLD of s_1 symbols at U_3 is designed by pretending s_2's symbols as noise, and it is given by

$$\tilde{s}_1 = \arg\min_m \{|y_1 - \ \sqrt{P_t}\sqrt{\alpha_1}\ h_3\ s_{1,m}|^2\}. \tag{8}$$

The SINRs of the decoded signals at U_3 can be expressed by

$$\gamma_{U_3,s_1} = \frac{P_t\alpha_1|g_3|^2}{P_t\alpha_2|g_3|^2 + \sigma^2}. \tag{9}$$

Likewise, the MLD of s_1 symbols at U_4 node is created by portraying s_2's symbols as noise, and it is provided by

$$\tilde{s}_1 = \arg\min_m \{|y_2 - \ \sqrt{P_t}\sqrt{\alpha_1}\ h_4\ s_{1,m}|^2\}. \tag{10}$$

However, in order to detect the s_2 at U_4, a SIC should be implemented. The MLD of s_2 symbols at the U_4 can be given as

$$\tilde{s}_2 = \arg\min_m \{|\acute{y}_2 - \ \sqrt{P_t}\sqrt{\alpha_2}\ h_4\ s_{2,m}|^2\}, \tag{11}$$

where

$$\acute{y}_2 = y_2 - \ \sqrt{P_t}\ h_4\ \sqrt{\alpha_1}\ \tilde{s}_{1,m}, \tag{12}$$

where $s_{i,m}$ denotes the m^{th} constellation point of U_i.

The SINRs of the decoded signals at U_4 can be expressed by

$$\gamma_{U_4,s_1} = \frac{P_t\alpha_1|g_4|^2}{P_t\alpha_2|g_4|^2 + \sigma^2}, \tag{13}$$

$$\gamma_{U_4,s_2} = \frac{P_t\alpha_2|g_4|^2}{\sigma^2}. \tag{14}$$

[1]The envelope of h_r and h_1, h_2, h_3, and h_4 follows Nakagami-m fading with Ω_1, Ω_2, Ω_3, and Ω_4 spread and m_1 and m_2, m_3, and m_4 shape parameters with m_1 and m_2, m_3, and m_4 are greater than 1 [27].

III. Performance Analysis

In this section, we formulate the sum rate and outage probability of the proposed system.

A. Sum Rate

In this subsection, the capacity and sum rate of the received signals at U_3 and U_4 can be expressed by

$$C_{U_3} = \frac{1}{2}\log_2\left(1+\min\left(\frac{P_1|h_1|^2}{P_2|h_2|^2+\sigma^2}, \frac{P_t\alpha_1|g_3|^2}{P_t\alpha_2|g_3|^2+\sigma^2}\right)\right), \quad (15)$$

$$C_{U_4} = \frac{1}{2}\log_2\left(1+\min\left(\frac{P_2|h_2|^2}{\sigma^2}, \frac{P_t\alpha_2|g_4|^2}{\sigma^2}\right)\right). \quad (16)$$

Thus, the sum rate of the proposed system can be expressed as

$$\text{Sum}_{\text{Rate}} = C_{U_3} + C_{U_4}. \quad (17)$$

The outage probability of U_3 and U_4 are given by

$$P_{U_3}(out) = Pr\left(\frac{1}{2}\log_2\left(1+\min\left(\frac{P_1|h_1|^2}{P_2|h_2|^2+\sigma^2}, \frac{P_t\alpha_1|g_3|^2}{P_t\alpha_2|g_3|^2+\sigma^2}\right)\right) < R_1\right) \quad (18)$$

$$P_{U_4}(out) = Pr\left(\frac{1}{2}\log_2\left(1+\min\left(\frac{P_2|h_2|^2}{\sigma^2}, \frac{P_t\alpha_2|g_4|^2}{\sigma^2}\right)\right)\right), \quad (19)$$

where R_1 and R_2 are threshold of U_3 and U_4.

The outage probability of the system is given as in [**?**] by

$$P_{\text{system}}(out) = 1-\left(1-P_{U_3}(out)\right)\left(1-P_{U_4}(out)\right), \quad (20)$$

IV. Numerical Results

In this section, we obtain the simulation results of the proposed system in terms of the sum rate and outage probability, BER, and we compare the results with conventional OMA. The parameters used in the simulation are as follows, $\Omega_1 = \Omega_4 = 0.5$, $\Omega_2 = \Omega_3 = 1$, $m_1 = m_2 = m_3 = m_4 = m_r$, and $P_t = P_1 = P_2$.

In Fig. 2, we present the BER performance of the received signals at U_3 and U_4 with different shape parameter, $m_r = \{1, 1.5, 2\}$. We observe that the BER of U_4 has better performance than U_3. The BER of both users' signals has an error floor at higher transmit power. It can be argued that increasing the transmit powers of users does not guarantee an enhancement in the BER performance of the uplink cooperative NOMA due to the increase in inter-user interference. We can also see that increasing m_r improves the system performance.

For further comparison, in Fig. 3 we compare the BER of system performance of uplink cooperative NOMA with uplink cooperative OMA. We observe that OMA systems have better BER performance compared to uplink cooperative NOMA. This is because there is no inter-user interference in OMA, while there is in NOMA.

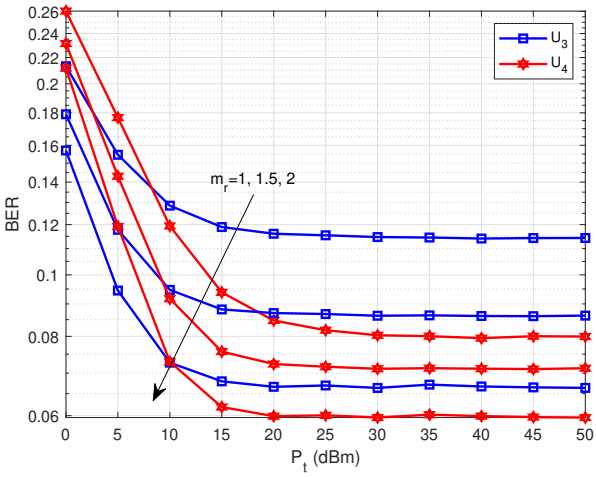


Fig. 2: BER of the proposed system with two users i.e., U_3 and U_4.

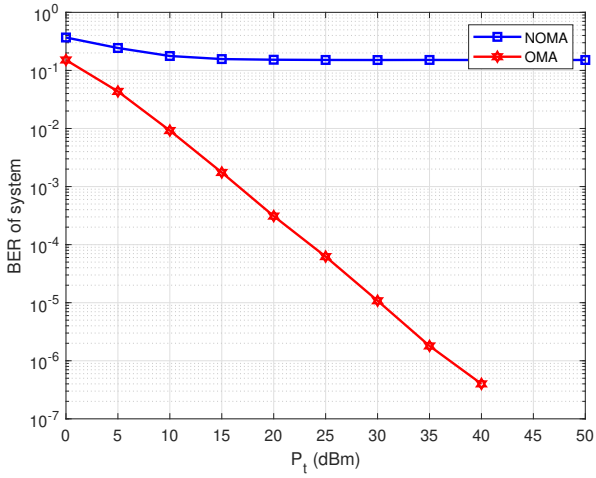


Fig. 3: BER of system comparison between NOMA and OMA..

Moreover, the outage probability performance of the received signals at U_3 and U_4 is presented in Fig. 4 with different shape parameter, $m_r = \{1, 1.5, 2\}$. We observe that U_4 achieves higher performance gain compared to U_3. This latter has an error floor at high transmit power values. This is due to the inter-user interference from the first phase. We can also see that increasing m_r improves the system performance.

For further comparison, we compared the outage probability of system performance of uplink cooperative NOMA and uplink cooperative OMA is presented in Fig. 5. We observe that uplink cooperative NOMA systems have better outage probability performance compared to uplink cooperative OMA at lower transmit power value. At high transmit power values,

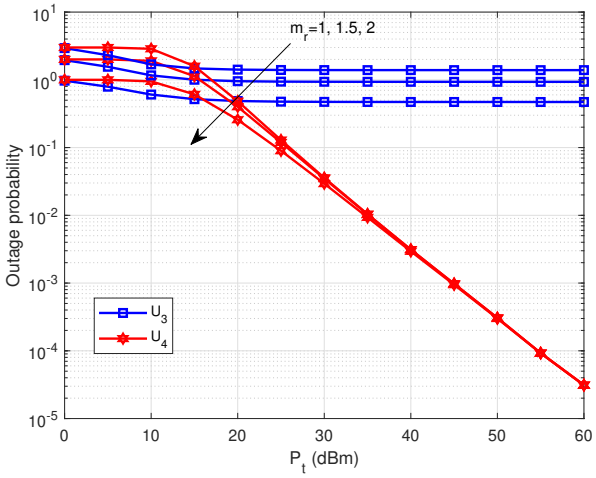


Fig. 4: Outage probability of the proposed system with two users i.e., U_3 and U_4.

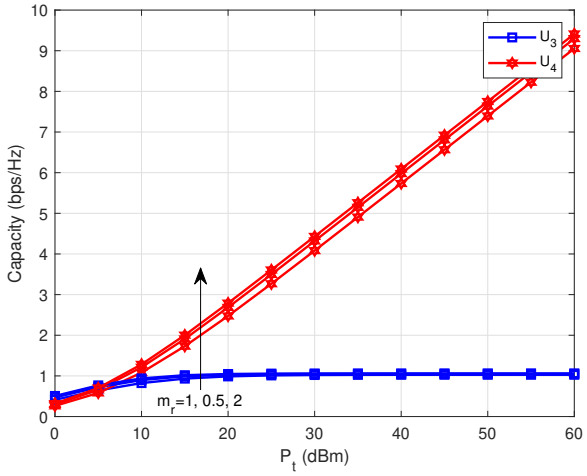


Fig. 6: Capacity of the proposed system with two users i.e., U_3 and U_4

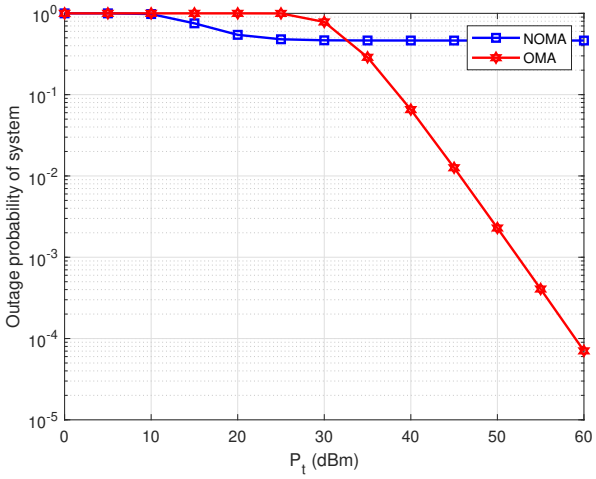


Fig. 5: Outage probability of system comparison between NOMA and OMA.

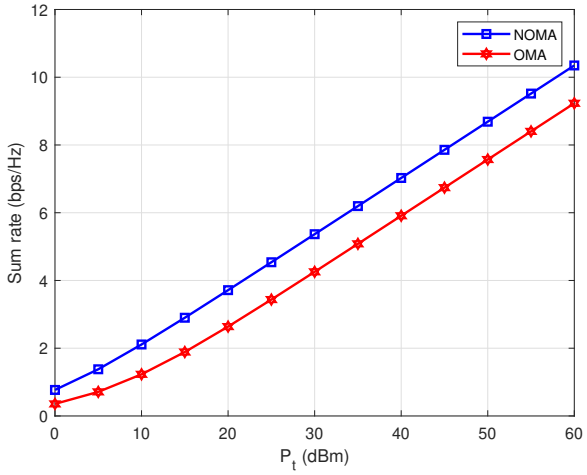


Fig. 7: Sum rate comparison between NOMA and OMA.

while OMA outperforms, uplink cooperative NOMA performance has an error floor. It means that uplink cooperative NOMA suffers from inter-user interference at high transmit power values.

Furthermore, the capacity performance of the received signals at U_3 and U_4 is presented in Fig. 6 with different shape parameters, $m_r = \{1, 1.5, 2\}$. We observe that the capacity of U_4 achieves higher performance gain compared to U_3. This latter is saturated at high transmit power values due to the inter-user interference from the first phase. We can also see that increasing m_r improves the capacity performance of both users.

For further comparison, in Fig. 7 we compare the sum rate performance of uplink cooperative NOMA with uplink cooperative OMA. We observe that the uplink cooperative NOMA system has better potential compared to the uplink cooperative NOMA. Although there is inter-user interference in OMA, NOMA achieves better performance potential.

V. Conclusion

This paper investigated the uplink and downlink cooperative NOMA. We have obtained the BER, capacity, and outage probability expressions for the considered systems. We compare the performance of the proposed system with traditional OMA. The Monte Carlo simulations are used to obtain the presented results. The results demonstrate that the proposed system achieves high-performance gains compared to its OMA counterpart. NOMA systems have inter-user interference, which effects the SIC process, especially at the uplink phase.

References

[1] Y. Qi, X. Zhang, and M. Vaezi, "Over-the-air implementation of NOMA: New experiments and future directions," *IEEE Access*, vol. 9, pp. 135 828–135 844, 2021.

[2] Z. Q. Al-Abbasi, K. M. Rabie, and D. K. So, "EE optimization for downlink NOMA-based multi-tier CRANs," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 5880–5891, 2021.

[3] D.-T. Do, T.-T. T. Nguyen, C.-B. Le, M. Voznak, Z. Kaleem, and K. M. Rabie, "UAV relaying enabled NOMA network with hybrid duplexing and multiple antennas," *IEEE Access*, vol. 8, pp. 186 993–187 007, 2020.

[4] M. Aldababsa, M. Toka, S. Gökçeli, G. K. Kurt, and O. Kucur, "A tutorial on nonorthogonal multiple access for 5G and beyond," *Wireless communications and mobile computing*, vol. 2018, no. 1, p. 9713450, 2018.

[5] I. Budhiraja, N. Kumar, S. Tyagi, S. Tanwar, Z. Han, M. J. Piran, and D. Y. Suh, "A systematic review on NOMA variants for 5G and beyond," *IEEE Access*, vol. 9, pp. 85 573–85 644, 2021.

[6] N. Ashraf, S. A. Sheikh, S. A. Khan, I. Shayea, and M. Jalal, "Simultaneous wireless information and power transfer with cooperative relaying for next-generation wireless networks: A review," *IEEE Access*, vol. 9, pp. 71 482–71 504, 2021.

[7] H. Liu, Z. Ding, K. J. Kim, K. S. Kwak, and H. V. Poor, "Decode-and-forward relaying for cooperative NOMA systems with direct links," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8077–8093, 2018.

[8] X. Liang, Y. Wu, D. W. K. Ng, Y. Zuo, S. Jin, and H. Zhu, "Outage performance for cooperative NOMA transmission with an AF relay," *IEEE Communications Letters*, vol. 21, no. 11, pp. 2428–2431, 2017.

[9] M. Yang, J. Chen, L. Yang, L. Lv, B. He, and B. Liu, "Design and performance analysis of cooperative NOMA with coordinated direct and relay transmission," *IEEE Access*, vol. 7, pp. 73 306–73 323, 2019.

[10] P. Xu, Y. Wang, G. Chen, G. Pan, and Z. Ding, "Design and evaluation of buffer-aided cooperative NOMA with direct transmission in IoT," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8145–8158, 2020.

[11] H. Liu, N. I. Miridakis, T. A. Tsiftsis, K. J. Kim, and K. S. Kwak, "Coordinated uplink transmission for cooperative NOMA systems," in *IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[12] G. Li and D. Mishra, "Cooperative NOMA networks: User cooperation or relay cooperation?" in *IEEE international conference on communications (ICC)*, 2020, pp. 1–6.

[13] M. F. Kader, S. R. Islam, and O. A. Dobre, "Device-to-device aided cooperative NOMA transmission exploiting overheard signal," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1304–1318, 2021.

[14] S. Joshi and R. K. Mallik, "Cooperative NOMA with AF Relaying over Nakagami-m Fading in a D2D Network," in *IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, 2019, pp. 1–6.

[15] J. Zhang, X. Tao, H. Wu, and X. Zhang, "Performance analysis of user pairing in cooperative NOMA networks," *IEEE Access*, vol. 6, pp. 74 288–74 302, 2018.

[16] X. Gong, X. Yue, and F. Liu, "Performance analysis of cooperative NOMA networks with imperfect CSI over Nakagami-m fading channels," *Sensors*, vol. 20, no. 2, p. 424, 2020.

[17] G. Li, D. Mishra, and H. Jiang, "Cooperative NOMA with incremental relaying: Performance analysis and optimization," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 11 291–11 295, 2018.

[18] Y. Zhang, Z. Yang, Y. Feng, and S. Yan, "Performance analysis of a novel uplink cooperative NOMA system with full-duplex relaying," *IET Communications*, vol. 12, no. 19, pp. 2408–2417, 2018.

[19] F. Kara and H. Kaya, "Threshold-based selective cooperative NOMA: Capacity/outage analysis and a joint power allocation-threshold selection optimization," *IEEE Communications Letters*, vol. 24, no. 9, pp. 1929–1933, 2020.

[20] S. Beddiaf, A. Khelil, F. Khennoufa, and K. Rabie, "On the impact of IQI on cooperative NOMA with direct links in the presence of imperfect CSI," *Physical Communication*, vol. 56, p. 101952, 2023.

[21] S. Beddiaf, A. Khelil, F. Khennoufa, F. Kara, H. Kaya, X. Li, K. Rabie, and H. Yanikomeroglu, "A Unified Performance Analysis of Cooperative NOMA With Practical Constraints: Hardware Impairment, Imperfect SIC and CSI," *IEEE Access*, vol. 10, pp. 132 931–132 948, 2022.

[22] A. A. Hamza, I. Dayoub, I. Alouani, and A. Amrouche, "On the error rate performance of full-duplex cooperative NOMA in wireless networks," *IEEE Transactions on Communications*, vol. 70, no. 3, pp. 1742–1758, 2021.

[23] F. Kara and H. Kaya, "On the error performance of cooperative-NOMA with statistical CSIT," *IEEE Communications Letters*, vol. 23, no. 1, pp. 128–131, 2018.

[24] ——, "Threshold-based selective cooperative-NOMA," *IEEE Communications Letters*, vol. 23, no. 7, pp. 1263–1266, 2019.

[25] F. Khennoufa and A. Khelil, "Exact bit error rate of cooperative NOMA without diversity combining," in *IEEE International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP)*, 2021, pp. 1–5.

[26] F. Khennoufa, A. Khelil, and E. Kenane, "BER and outage performance for multi-hop user-relays NOMA with imperfect SIC and CSI," in *IEEE International Conference of Advanced Technology in Electronic and Electrical Engineering (ICATEEE)*, 2022, pp. 1–6.

[27] F. Khennoufa, A. Khelil, K. Rabie, H. Kaya, and X. Li, "An efficient hybrid energy harvesting protocol for cooperative NOMA systems: Error and outage performance," *Physical Communication*, vol. 58, p. 102061, 2023.

[28] F. Khennoufa, K. Abdellatif, and F. Kara, "Bit error rate and outage probability analysis for multi-hop decode-and-forward relay-aided NOMA with imperfect SIC and imperfect CSI," *AEU-International Journal of Electronics and Communications*, vol. 147, p. 154124, 2022.

[29] ——, "Bit error rate evaluation of relay-aided cooperative NOMA with energy harvesting under imperfect SIC and CSI," *Physical Communication*, vol. 52, p. 101630, 2022.

[30] F. Khennoufa, A. Khelil, S. Beddiaf, F. Kara, K. Rabie, H. Kaya, A. Emir, S. Ikki, and H. Yanikomeroglu, "Wireless powered cooperative communication network for dual-hop uplink NOMA with IQI and SIC imperfections," *IEEE Access*, 2023.

[31] S. Beddiaf, A. Khelil, F. Khennoufa, F. Kara, K. Rabie, X. Li, H. Kaya, A. Emir, and H. Yanikomeroglu, "Impact of hardware impairment on the uplink SIMO Cooperative NOMA with selection relay under imperfect CSI," *IEEE Access*, 2023.

[32] F. Khennoufa and K. Abdellatif, "BER performance of co-operative relay NOMA-assisted PS protocol with imperfect SIC and CSI," *International Journal of Sensors Wireless Communications and Control*, vol. 13, no. 2, pp. 99–107, 2023.

[33] F. Khennoufa, K. Abdellatif, F. Kara, H. Yanikomeroglu, K. Rabie, T. Y. Elganimi, and S. Beddiaf, "Error performance analysis of UAV-mounted RIS for NOMA systems with practical constraints," *IEEE Communications Letters*, 2024.

Cooperative NOMA assisted Multiple Relays with Direct Links

Faical Khennoufa[1], Abdellatif Khelil[1], Safia Beddiaf[1], Walid Beddiaf[2]
[1]*LGEERE Laboratory, Department of Electrical Engineering, Echahid Hamma Lakhdar University, El-Oued, Algeria, email: khennoufa-faical,abdellatif-khelil, beddiaf-safia@univ-eloued.dz.*
[2]*LEVE Laboratory, Department of Architecture, Mohamed Khider University, Biskra, Algeria, email: w.beddiaf@univ-biskra.dz.*

***Abstract*—This paper investigates a multi-hop decode-and-forward (DF) relay-aided cooperative non-orthogonal multiple access (MH-DF-NOMA) scheme. The model solves the issues of poor channel quality and users losing connection to the broadcast station by delivering the signal to distant users via the path from intermediate relay nodes. If direct links are available, the system can also reduce path loss and improve the signal quality. We derive the closed-form bit error rate (BER) and outage probability expressions. Additionally, imperfect successive interference cancellation (ISIC) and imperfect channel state information (ICSI) are considered. We present computer simulations to validate the theoretical analyses. The results indicate that the proposed system achieves a higher performance gain compared to the traditional single relay scheme.**

***Index Terms*—BER, cooperative, multi-hop, NOMA, and outage probability**

I. Introduction

The ambitious high data rate applications in the beyond 5G wireless networks will require new solutions and technologies [1]. NOMA is one of the promising radio access technologies for improved performance in cellular communications of the next generation [2]. Compared to orthogonal multiple access (OMA), non-OMA (NOMA) can serve multiple users in the same resource block. Thus, it reduces latency and promotes massive connectivity. On the other hand, cooperative communications provide a useful method of combating channel disorders, such as fading, path-loss, and shadowing [3]. Thus, it extends the communication coverage area, improves the quality of service, and increases spectral efficiency [3].

The advantages of NOMA over OMA are validated in [4]. The cooperative (CNOMA) system outperforms OMA in terms of ergodic sum rate and OP [4]. CNOMA under the effect of hardware impairments, imperfect channel state information (ICSI), and imperfect successive interference cancelation (ISIC) has been investigated regarding outage probability and bit error rate (BER) [5]. The BER of NOMA and half duplex (HD) CNOMA was analyzed in [6]–[10]. The investigation of CNOMA has been extended to full-duplex, where the outage probability and ergodic sum capacity performance of the full-duplex (FD) CNOMA under the realistic assumption of imperfect self-interference cancellation have been performed [11]. The BER of FD CNOMA was analyzed in [12]. There are two types of CNOMA: CNOMA with direct links and CNOMA without direct links. In the first type, both direct and indirect links are available between users and the source, while in the second type, indirect links are not available. In this paper, we focus on the second type. CNOMA with the direct links has been examined in terms of BER, outage probability, and ergodic capacity [13]–[15]. CNOMA under hardware impairments and in-phase/quadrature-phase imbalance was examined in terms of BER and outage probability [5], [16].

On the other hand, the multi-hop relay has gotten great attention in academia and industry because it can efficiently extend coverage to severely shadowed areas in the cell or places outside of the cell range. Meanwhile, it improves throughput, especially at the cell edge [17]–[20]. The multi-hop relay schemes' have been also investigated with the NOMA implementation. In the presence of an active eavesdropper for the multi-hop relay, the OP and throughput have been derived [17], [21]. The OP and throughput have been examined for multi-point cooperative relay NOMA with EH [22].

The interaction between NOMA and cooperative communication is one of the topics that drew the most attention to achieve higher spectral efficiency and compensation for path-loss [4]. As mentioned above, CNOMA assisted one relay has been investigated in terms of BER, EC, and OP [6]–[16]. The idea is to help the very far user or when the direct link between the source and users is not available due to the large distance or obstacles, etc. Although there is a relay to help users, the channel quality may still be poor between each node (i.e., source-relay and relay-users), and the performance will not be good. Therefore, channel degradation and long distances between nodes may cause a significant deterioration in performance for users. Hence, it is very important to resolve this problem. As mentioned, studies in the open literature consider single relay (or relay selection) in CNOMA. In the end, they are all two-hop scenarios. However, we consider a multi-hop scenario which is quite important to extend cell coverage to heavily shadowed areas in the cell or places outside the cell range [17]–[20]. Therefore, a NOMA with multi-hop relaying can be employed through intermediate relays that form a multi-hop path from the source to users to compensate the path-loss, expand network coverage, and deliver messages to users everywhere. Furthermore, in the cell, the multi-hop system can be used to improve the quality of the received signal when assistance relays establish a path to the remote user. At the same time, the remote user can receive a direct signal

from the source. In this regard, based on the discussions that we have mentioned above, and to the best of the authors' knowledge, the BER and outage probability expressions of multi-hop decode-and-forward (DF)-aided CNOMA (MH-DF-CNOMA) when the direct and indirect links are available have not been reported. Therefore, this paper investigates downlink MH-DF-CNOMA schemes with direct links to achieve ubiquitous coverage and connectivity in the future NOMA network. The e2e BER and outage probability expressions of MH-DF-CNOMA are analyzed at each hop under ICSI and ISIC, which are practical assumptions. We compare our model with traditional CNOMA with direct links of [13], [14]. The original contributions of this paper are as follows:

- We present downlink MH-DF-CNOMA direct links system models. The model consists of a base station (BS) and a group of intermediate assistance relay that form a multi-hop path. The far user can receive the BS signal directly, and the multi-hop path.
- The BER analysis for even single relay schemes is limited. In this paper, we perform a comprehensive analysis of the considered schemes. The considerations of the ISIC and ICSI are more accurate/realistic for the practical scenarios. We investigate ISIC and ICSI in the MH-DF-CNOMA. We obtain the e2e BER for each user. We discuss the effect of the ICSI and ISIC on the BER performance of the users.
- Moreover, we obtain the e2e outage probability for each user. The outage probability performance of users at each hop with the effect of ICSI and ISIC is evaluated.
- In BER and outage probability performance, we examine the impact of the increasing number of relays.
- Finally, the performance of the proposed system is compared with [13], [14] of a single relay scheme.

The paper is organized as follows. In section II, we present the proposed system model. In section III, we obtain the BER and outage probability of the proposed system. Section IV presents the numerical results of the proposed system. Finally, section V concludes the paper with the conclusion.

II. System Model

We consider a downlink MH-DF-CNOMA with direct link consisting of one BS, U_1, U_2, and a group of relay named R_L, where $L = 1, 2, 3, ..., N$ which creates a way to U_1 as shown in Fig. 1. We assume that U_1 is far and U_2 is near from the BS. Our assumption is that only R_1 from the group of relays is capable of directly receiving the signal from the BS signal, while R_{L+1} facilitates forwarding the signal to U_1. We assume all node works in DF and HD modes. We suppose also the CSI between all nodes is imperfect. The Rayleigh fading channel coefficient of R_L links is denoted by $g_{RL} \sim \mathcal{CN}(0, \sigma_{RL}^2 = d_{RL}^{\varsigma})$ and the Rayleigh fading channel coefficient of the U_1 and U_2 links are denoted by $h_i \sim \mathcal{CN}(0, \sigma_i^2 = d_i^{\varsigma})$, where $i = \{1, 2\}$. The channel estimation coefficients are presented as $\hat{g}_{RL} = g_{RL} - e \sim \mathcal{CN}(0, \hat{\sigma}_{RL}^2 = \sigma_{RL}^2 - \sigma_e^2)$ and $\hat{h}_i = h_i - e \sim \mathcal{CN}(0, \hat{\sigma}_i^2 = \sigma_i^2 - \sigma_e^2)$, where e is the error factor of the channel estimation, d_i is the distance between BS-U_1, d_{R1} is the distance between BS-R_1, d_{RL} is the distance between the relay nodes, and ς is the path loss factor.

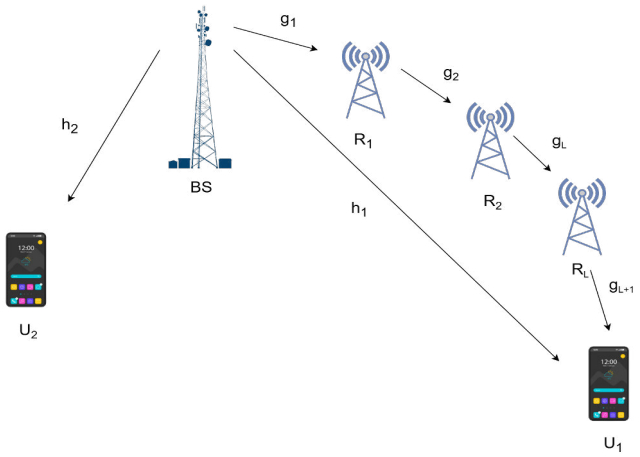


Fig. 1: MH-DF-CNOMA with direct links scheme.

In this system, We consider that the BS transmits as a superposition coding (SC) signal consisting of s_1 and s_2 to U_1, U_2, and R_1 directly. U_1 detect its signal directly while R_1 and U_2 uses the SIC to detect s_1 and s_2 signals. We suppose that the R_1 re-encode the U_1 signal (s_1) and forward it to R_{L+1} until it reaches U_1 (see Fig. 1). The power of the BS and relay nodes R_L are assumed to be equal and named P_t. The received signals at U_1, U_2, and R_1 can be written, respectively, as

$$y_i = \sqrt{P_t}\ (\hat{h}_i + e)\ (\sqrt{\alpha_1} s_1 + \sqrt{\alpha_2} s_2) + \ n, \quad i = \{1, 2\}, \quad (1)$$

$$y_{R1} = \sqrt{P_t}\ (\hat{g}_{R1} + e)\ (\sqrt{\alpha_1} s_1 + \sqrt{\alpha_2} s_2) + \ n, \quad (2)$$

where α_1 and α_2 are the power allocation coefficient for U_i, $\alpha_1 > \alpha_2$, in which $\alpha_1 + \alpha_2 = 1$, and n is the additive white Gaussian noise (AWGN) which follows $n \sim \mathcal{CN}(0,\ \sigma^2)$.

The U_1 decodes its signal using maximum likelihood detection (MLD) while the U_2 and R_1 nodes decode the s_1 signal first then using the SIC decodes the s_2 signal. The signal-to-interference-plus-noise ratios (SINRs) of the decoded signals at U_1 and U_2, and R_1 under ICSI and ISIC can be expressed, respectively, by

$$\gamma_{U_1, s_1} = \frac{P_t \alpha_1 |\hat{h}_1|^2}{P_t \alpha_2 |\hat{h}_1|^2 + P_t \sigma_e^2 + \sigma^2}, \quad (3)$$

$$\gamma_{U_2, s_1} = \frac{P_t \alpha_1 |\hat{h}_2|^2}{P_t \alpha_2 |\hat{h}_2|^2 + P_t \sigma_e^2 + \sigma^2}, \quad (4)$$

$$\gamma_{U_2, s_2} = \frac{P_t \alpha_2 |\hat{h}_2|^2}{\varrho P_t \alpha_1 |\hat{h}_2|^2 + P_t \sigma_e^2 + \sigma^2}, \quad (5)$$

$$\gamma_{R_1, s_1} = \frac{P_t \alpha_1 |\hat{g}_{R1}|^2}{P_t \alpha_2 |\hat{g}_{R1}|^2 + P_t \sigma_e^2 + \sigma^2}, \quad (6)$$

$$\gamma_{R_1, s_2} = \frac{P_t \alpha_2 |\hat{g}_{R1}|^2}{\varrho P_t \alpha_1 |\hat{g}_{R1}|^2 + P_t \sigma_e^2 + \sigma^2}, \quad (7)$$

where ϱ is the factor of ISIC.

The R_1 re-encode s_1 and forward it over the path of R_{L+1} to U_1. Thus, s_1 is decoded and forwarded over R_{L+1} until reaches U_1. The received signal at R_{L+1} and U_1 over L hop under ICSI can be expressed by

$$y_{L+1} = \sqrt{P_t}\ (\hat{g}_{R(L+1)} + e)\ s_1 + \ n, \tag{8}$$

where $\hat{g}_{R(L+1)} = g_{R(L+1)} - e \sim \mathcal{CN}(0, \hat{\sigma}^2_{R(L+1)} = \sigma^2_{R(L+1)} - \sigma^2_e)$ is channel estimation coefficient of between R_1-R_2 and R_{L-1}-R_L. $d_{R(L+1)}$ is the distance between R_1-R_2 and R_{L-1}-R_L.

The R_{L+1} and U_1 decodes s_1 signal using the MLD. In practical scenarios, the CSI is not imperfect. Therefore, the decoding error is expected in the detection and SIC process. Thus, the SINR of the decoded signals at R_{L+1} and U_1 under ICSI and ISIC can be expressed by

$$\gamma_{L+1,s_1} = \frac{P_t|\hat{g}_{R(L+1)}|^2}{P_t\sigma_e^2 + \sigma^2}. \tag{9}$$

III. Performance Analysis

A. Outage Probability

The user U_1 uses maximum ratio combining (MRC) to combine the two received signals (directly from BS and indirectly from $R(L+1)$) to improve the received signal quality. The OP expression of U_1 using the MRC is given as [5] by

$$P_{MRC,U1}(out) = (1 - P_{BS-L}(out)) \\ P(\gamma_{U_1,s_1} + \gamma_{L+1,s_1} < \theta_1^{(L)}) + P_{BS-L}(out)P(\gamma_{U_1,s_1} < \theta_1^{(L)}), \tag{10}$$

where $P_{BS-L}(out)$ is the OP of s_1 at U_1 over the path of R_L, $P(\gamma_{BS1,s_1} < \theta_1)$ is the OP of s_1 at U_1 in the direct like, and $P(\gamma_{U_1,s_1} + \gamma_{L+1,s_1} < \theta_1^{(L)})$ is the OP of s_1 at U_1 when MRC is implemented. We calculate each term of (10) as

$$P(\gamma_{U_1,s_1} < \theta_1^{(L)}) = 1 - exp\left(-\frac{(P_t\sigma_e^2 + \sigma^2)\theta_1^{(L)}}{(\alpha_1 - \alpha_2\theta_1^{(L)})P_tE[|\hat{h}_1|^2]}\right), \tag{11}$$

where $\theta_1^{(L)} = 2^{Tr_1(L)} - 1$. $P_{BS-L}(out)$ is the e2e OP of s_1 at R_L over L hop, so the end-to-end (e2e) outage probability of 2 hops can be calculated as [17], [23]

$$P_{BS-2}(out) = (1 - P_{BS-R_1}(out))\, P_{R_1-R_2}(out) \\ + P_{BS-R_1}(out)\,(1 - P_{R_1-R_2}(out))\,. \tag{12}$$

For more than two hops i.e., $L > 2$, the e2e outage probability of s_1 at R_L can be expressed as

$$P_{BS-L}(out) = \big(1 - P_{BS,R_{L-1}}(out)\big)\, P_{R_{L-1}-R_L}(out) \\ + P_{BS,R_{L-1}}(out)\,\big(1 - P_{R_{L-1}-R_L}(out)\big)\,. \tag{13}$$

If $L = 1$ (which means that R_1), the outage probability of BS-R_1 can be calculated as

$$P_{BS-R_1}(out) = 1 - exp\left(-\frac{(P_t\sigma_e^2 + \sigma^2)\theta_1^{(L)}}{(\alpha_1 - \alpha_2\theta_1^{(L)})P_tE[|\hat{g}_{RL}|^2]}\right), \tag{14}$$

If $L > 1$, which means that the outage probability of R_{L-1}-R_L can be calculated as

$$P_{R_{L-1}-R_L}(out) = 1 - exp\left(-\frac{(P_t\sigma_e^2 + \sigma^2)\theta_1^{(L)}}{P_tE[|\hat{g}_{R(L+1)}|^2]}\right). \tag{15}$$

By substituting (14) and (15) into (12) and (13), we find the e2e outage probability of s_1 at R_L over L hop.

The last R_L forward s_1 to U_1. Thus, the U_1 implements the MRC to combine the received signal. Since γ_{U_1,s_1} and γ_{L+1,s_1} are two independent random variables, where $\gamma_{U_1,s_1} \neq \gamma_{L+1,s_1}$, so $P(\gamma_{U_1,s_1} + \gamma_{L+1,s_1} < \theta_1^{(L)})$ can be computed as [5], [13] by

$$P(\gamma_{U_1,s_1} + \gamma_{L+1,s_1} < \theta_1^{(L)}) = 1 - [A_1 + A_2]\,, \tau_a \neq \tau_b, \tag{16}$$

where $A_1 = \frac{\tau_a}{\tau_a - \tau_b}\exp\left(-\tau_a\theta_1^{(L)}\right)$, $A_2 = \frac{\tau_b}{\tau_b - \tau_a}\exp\left(-\tau_b\theta_1^{(L)}\right)$, $\tau_a = \frac{P_t\sigma_e^2 + \sigma^2}{(\alpha_1 - \alpha_2\theta_1^{(L)})P_tE[|\hat{h}_1|^2]}$, and $\tau_b = \frac{P_t\sigma_e^2 + \sigma^2}{P_tE[|\hat{g}_{R(L+1)}|^2]}$.

The e2e outage probability of s_1 using MRC is obtained by substituting (12), (13), (14), (15), and (16) into (10).

Moreover, the outage probability of U_2 can be expressed as

$$P_{U_2}(out) = 1 - exp\left(-\frac{(P_t\sigma_e^2 + \sigma^2)\theta_2^{(L)}}{(\alpha_1 - \alpha_2\theta_2^{(L)})P_tE[|\hat{h}_2|^2]}\right) \\ exp\left(-\frac{(P_t\sigma_e^2 + \sigma^2)\theta_2^{(L)}}{(\alpha_2 - \varrho\alpha_1\theta_2^{(L)})P_tE[|\hat{h}_2|^2]}\right). \tag{17}$$

B. Bit Error Rate

In the MH-DF-CNOMA with direct links, after the R_1 detects the transmitted symbols, it re-encodes the s_1 symbol and forwards it to R_{L+1}. In this case, the U_1 receives its signal s_1 from two different sources (i.e., BS and R_L). Hence, we consider that U_1 implements the diversity combining technique MRC to improve its received signal quality. We obtain the e2e BER expression of MH-DF-CNOMA with direct links for U_1 as in [5] by

$$P_{MRC,U_1}(e) = \\ \frac{1}{2}\sum_{j=1}^{2}[P_{prop,MH}(e|j) \times P_{e2e,s_1,(L)hop}(e|j) \\ + (1 - P_{e2e,s_1,(L)hop}(e|j))P_{coop,MH}(e|j)], \tag{18}$$

where $P_{prop,MH}(e|j)$ is the BER in the presence of error propagation from R_L to U_1 achieved by the MRC to users achieved by the MRC, $P_{e2e,s_1,(L)hop}(e|j)$ is the BER of the s_1 at R_L, and $P_{coop,MH}(e|j)$ is the BER when the symbols of U_1 are detected correctly at the R_L and forwarded to the U_1 and then combined with MRC. Each term of (18) is calculated below.

The e2e BER of $P_{e2e,s_1,(L)hop}(e|j)$ for two hops i.e., BS-R_1-U_1 is given as in [17] by

$$P_{e2e,s_1,(2)hop}(e|j) = \\ P_{R_1,s_1}(e|j)\ (1 - P_{R_2,s_1}(e|j)) + P_{R_2,s_1}(e|j)\ (1 - P_{R_1,s_1}(e|j)). \tag{19}$$

Further, for more than two hops i.e., $L > 2$ (BS-R_{L-1}-R_L), we express e2e BER of L hops as in [17], [23] by

$$P_{e2e,s_1,(L)hop}(e|j) = \underbrace{P_{e2e,s_1,(L-1)hop}(e|j)}_{\iota_1} (1 - P_{R_L,s_1}(e|j)) + P_{R_L,s_1}(e|j) (1 - \underbrace{P_{e2e,s_1,(L-1)hop}(e|j)}_{\iota_1}), \tag{20}$$

where ι_1 is the e2e BER from BS to R_{L-1} which can be expressed as

$$P_{e2e,s_1,(L-1)hop}(e|j) = P_{R_L,s_1}(e|j) (1 - P_{e2e,s_1,(L-1)hop}(e|j)) + P_{e2e,s_1,(L-1)hop}(e|j) (1 - P_{R_L,s_1,s_1}(e|j)). \tag{21}$$

Each term of (19), (20), and 21) can be calculated by

$$P_{R_L,s_1}(e|j) = \frac{1}{4}\left(\sqrt{\frac{(P_t\gamma\chi_j E[|\hat{g}_{R_L}|^2]/(P_t\gamma\chi_j\sigma_e^2+1))}{2+(P_t\gamma\chi_j E[|\hat{g}_{R_L}|^2]/(P_t\gamma\chi_j\sigma_e^2+1))}}\right). \tag{22}$$

$$P_{R_{L+1},s_1}(e|j) = \frac{1}{2}\left(\sqrt{\frac{(P_t\gamma E[|\hat{g}_{R_{L+1}}|^2]/(P_t\gamma\sigma_e^2+1))}{2+(P_t\gamma E[|\hat{g}_{R_{L+1}}|^2]/(P_t\gamma\sigma_e^2+1))}}\right). \tag{23}$$

where $\chi_j = [(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2, (\sqrt{\alpha_1}-\sqrt{\alpha_2})^2]$,

The e2e BER of s_1 over the path of R_L is obtained by substituting (22) and (23) into (19-21).

Since the U_1 receives its signal from BS and, the U_1 employs MRC, the total SNR is the sum of the SNRs after combining the two received signals.

The average BER of the cooperative of two branches using MRC over Rayleigh fading in (24) when the mean of the SNRs of branches are different $\overline{\gamma}_{Lhop,s1} \neq \overline{\gamma}_{1,s1,j}$ as given in [5], [13] by

$$P_{coop,MH}(e|j) = \frac{1}{2}\left(1 - B_1\left(B_2 - B_3\right)\right), \tag{24}$$

where $\gamma_{Lhop,s1} = P_t\gamma|\hat{g_{R_L}}|^2/(P_t\gamma\sigma_e^2+1)$, $\gamma_{1,s1,j} = P_t\gamma\chi_j|\hat{h}_1|^2/(P_t\gamma\chi_j\sigma_e^2+1)$, $B_1 = \frac{1}{\overline{\gamma}_{1,s1,j}-\overline{\gamma}_{Lhop,s1}}$, $B_2 = \overline{\gamma}_{1,s1,j}\sqrt{\frac{\overline{\gamma}_{1,s1,j}}{1+\overline{\gamma}_{1,s1,j}}}$, $B_3 = \overline{\gamma}_{Lhop,s1}\sqrt{\frac{\overline{\gamma}_{Lhop,s1}}{1+\overline{\gamma}_{Lhop,s1}}}$, $\overline{\gamma}_{Lhop,s1} = P_t\gamma E[|\hat{g}_{R_L}|^2]/(P_t\gamma\sigma_e^2+\sigma^2)$, and $\overline{\gamma}_{1,s1,j} = P_t\gamma\chi_j E[|\hat{h}_1|^2]/(P_t\gamma\chi_j\sigma_e^2+\sigma^2)$.

After utilizing MRC at U_1, the incoming signals from BS and over the path of R_L are summed. If R_L (i.e., the last one) detects the symbol of U_1 erroneously, it will be forwarded to U_1 as an error propagation. Hence, the received signals will be combined at U_1 using MRC in the case of error propagation. Thus, the BER using MRC with error propagation at the U_1 is given as in [13] by

$$P_{Prop,MH}(e|j) = \frac{\overline{\gamma}_{Lhop,s1}}{\overline{\gamma}_{Lhop,s1}+\overline{\gamma}_{1,s1,j}}. \tag{25}$$

The e2e BER of MH-DF-CNOMA with direct links using MRC is obtained by substituting (19-21), (25), and (26) into (18).

TABLE I: Distances used for scenarios.

	Distances
Scenario I	d_1=5, d_{R1}=1, d_{R2}=4
Scenario II	d_1=5, d_{R1}=1, d_{R2}=2, D_{R3}=2
Scenario III	d_1=5, d_{R1}=1, d_{R2}=1.5, d_{R3}=1.5, d_{R4}=1
Scenario IV	d_1=5, d_{R1}=1, d_{R2}=1, D_{R3}=1, d_{R4}=1, d_{R5}=1
Scenario V ([13], [14])	d_1=5, d_{R1}=2.5, d_{R2}=2.5

Moreover, the average BER of U_2 can be expressed as in [5] by

$$P_{U_2}(e) = \frac{1}{4}\sum_{v=1}^{6}\kappa_v\left(\sqrt{\frac{(P_t\gamma\aleph_v E[|\hat{g}_2|^2]/(P_t\gamma\beth_v\sigma_e^2+1))}{2+(P_t\gamma\aleph_v E[|\hat{g}_2|^2]/(P_t\gamma\beth_v\sigma_e^2+1))}}\right). \tag{26}$$

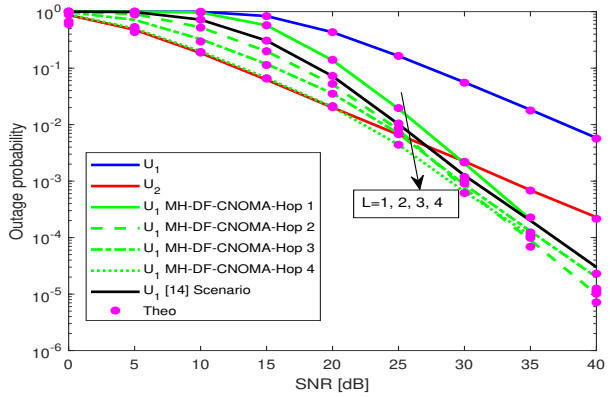
where $\kappa_v = [1,1,-1,1,1,-1]$, $\aleph_v = [\alpha_2, \alpha_2, (\sqrt{\alpha_1}+\sqrt{\alpha_2})^2, (\sqrt{2\alpha_1}+\sqrt{\alpha_2})^2, (\sqrt{\alpha_1}-\sqrt{\alpha_2})^2, (2\sqrt{\alpha_1}-\sqrt{\alpha_2})^2]$, and $\beth_v = [(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2, (\sqrt{\alpha_1}+\sqrt{\alpha_2})^2, (\sqrt{\alpha_1}+\sqrt{\alpha_2})^2, (\sqrt{\alpha_1}+\sqrt{\alpha_2})^2, (\sqrt{\alpha_1}-\sqrt{\alpha_2})^2, (\sqrt{\alpha_1}-\sqrt{\alpha_2})^2]$.

IV. Numerical Results

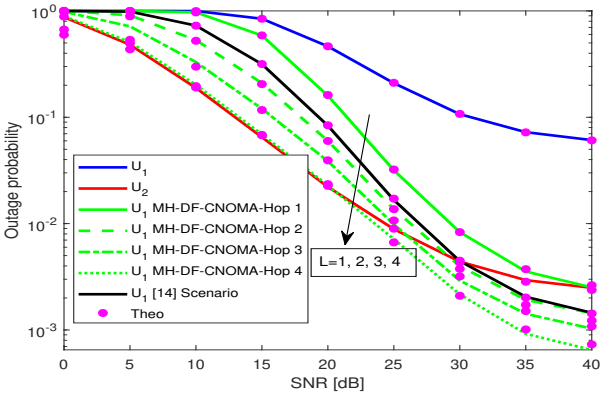
The parameters used in the simulation are as follows, d_2=1 m, ς=4, α_1=0.85, α_2=0.15, ϱ=0.001 and σ_e^2=0.01. The curves denote simulations, and markers represent the analytical derivation in our results. It can be easily observed that our derivations are matched perfectly with the simulations that validate our derivations. The simulation parameters of [13], [14] are used to evaluate the performance gain of our proposed system. Thus, the distance parameters of the simulation of [13], [14] and our system are presented in table I. As we can see through table I, the relays are placed differently, where users are distributed almost with equal distances between them until reaching U_1. The fixed simulation parameters are presented as follows, ς=3, α_1=0.8, α_2=0.2, ϱ=0.001, and σ_e^2=0.001. The curves denote simulations, and markers represent the analytical derivation in our results. It can be easily observed that our derivations are matched perfectly with the simulations that validate our derivations.

In Fig. 2, we present the OP performance of MH-DF-CNOMA with DL under perfect CSI, ICSI, SIC, and ISIC. We observe that the proposed scheme improved the performance of U_1. It is observed also the OP of our proposed scheme improved as the number of relays increased. It can be easily seen that the proposed scheme achieves higher performance gain compared to [14]. Fig. 2 (b) presents the OP performance of our system under ICSI and ISIC. It is observed that in the presence of ICSI and ISIC, the OP performance of our proposed system deteriorates compared to the perfect case in Fig. 2 (a) and causes an error floor at the high SNR.

To evaluate the effect of power allocation on the performance of our proposed system, Fig. 3, presents the OP of MH-DF-CNOMA with direct links when SNR= 20 dB. It is observed that increasing α_2 decreases the OP of U_2 and increases the performance of U_1. Also, increasing the number of relays improves the performance of our system. It can be seen that the OP performance of our proposed system



(a)



(b)

Fig. 2: OP of MH-DF-CNOMA with direct links w.r.t. SNR: (a) With perfect CSI, (b) With ICSI.

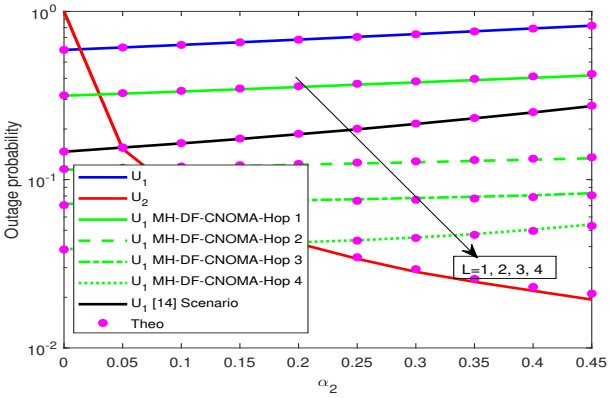
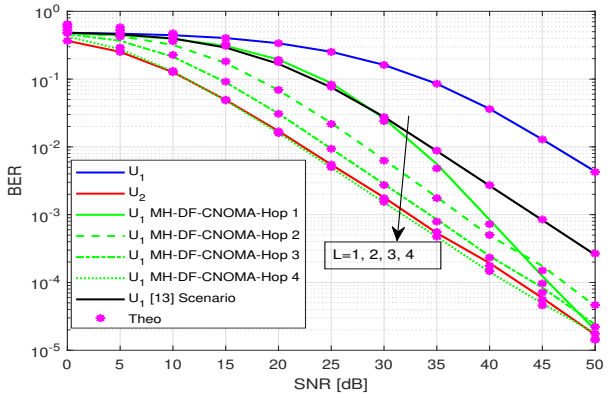


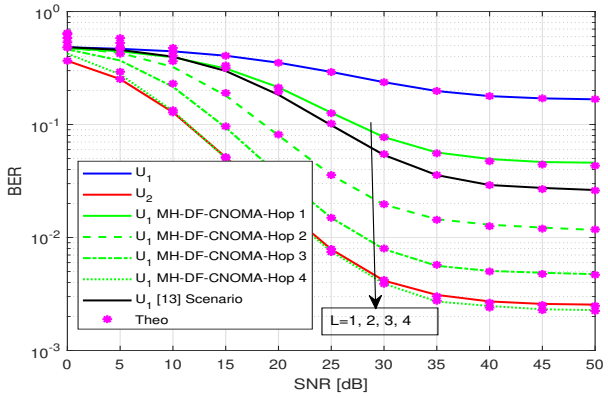
Fig. 3: OP of MH-DF-CNOMA with direct links w.r.t. α_2 with different relay.

achieves higher performance gains compared to [14] in all power allocation values.

On the other hand, Fig. 4 and Fig. 5 present the BER performance of MH-DF-CNOMA with direct links compared to a single relay scheme. In Fig. 4, we present the BER



(a)



(b)

Fig. 4: BER of MH-DF-CNOMA with direct links w.r.t. SNR: (a) With perfect CSI, (b) With ICSI.

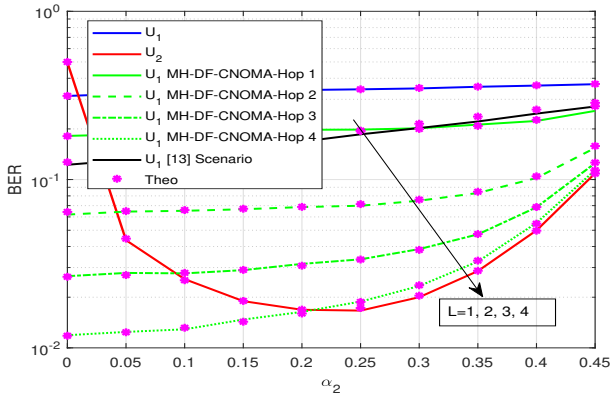


Fig. 5: OP of MH-DF-CNOMA with direct links w.r.t. α_2 with different relay.

performance of our proposed system compared to a single relay scheme of [13]. We observe that the BER performance of MH-DF-CNOMA with direct links improves as the number of relays increases. Moreover, the BER performance of our proposed system achieves higher performance gain compared to [13] of a single relay scheme. In Fig. 4 (a) and Fig. 4 (b), we present the BER of our system with perfect CSI and ICSI respectively. It can be seen that the BER performance of the system degrades in the presence of ICSI and achieves an error floor at the high SNR.

To examine the impacts of the power allocation of our system, in Fig. 5, we present the BER of our system w.r.t α_2 when SNR= 20 dB. We observe that the change of α_2 leads to a change performance of the users which affects the SIC process of each node. Again, increasing the number of relays improves the performance of our scheme. Furthermore, it can be easily seen that our proposed scheme outperforms the single relay scheme of [13].

V. Conclusion

This paper investigated the MH-DF-CNOMA with direct links. For practical scenarios, the ICSI and ISIC are taken into account. We have derived the e2e OP and BER expressions for the considered systems under ICSI and ISIC. We have discussed the effect of power allocation, the number of relays, ICSI, and SIC on the performance of systems. We have validated our mathematical derivations by using Monte Carlo simulations. The results demonstrate that the proposed system achieves high-performance gain compared to a single relay scheme and [13], [14], and the performance improves as the number of relays increases. Moreover, the changes in the power allocation affect the performance of the systems, which affects the SIC process. The ICSI decreases the performance of the systems and leads to an error floor at the high SNR. The presence of the ICSI affects signal detection and the SIC process at nodes.

References

[1] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020.

[2] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, and S. M. Sait, "A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2192–2235, 2020.

[3] F. Khennoufa and A. Khelil, "Capacity performance comparison of cooperative NOMA and OMA systems in DF and AF," *Electrical Engineering*, vol. 2, p. 1, 2021.

[4] M. Aldababsa, M. Toka, S. Gökçeli, G. K. Kurt, and O. Kucur, "A tutorial on nonorthogonal multiple access for 5G and beyond," *Wireless communications and mobile computing*, vol. 2018, no. 1, p. 9713450, 2018.

[5] S. Beddiaf, A. Khelil, F. Khennoufa, F. Kara, H. Kaya, X. Li, K. Rabie, and H. Yanikomeroglu, "A unified performance analysis of cooperative NOMA with practical constraints: Hardware impairment, imperfect SIC and CSI," *IEEE Access*, vol. 10, pp. 132 931–132 948, 2022.

[6] F. Kara and H. Kaya, "BER performances of downlink and uplink NOMA in the presence of SIC errors over fading channels," *Iet Communications*, vol. 12, no. 15, pp. 1834–1844, 2018.

[7] F. Khennoufa and A. Khelil, "Exact bit error rate of cooperative NOMA without diversity combining," in *2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP)*. IEEE, 2021, pp. 1–5.

[8] F. Khennoufa, K. Abdellatif, F. Kara, H. Yanikomeroglu, K. Rabie, T. Y. Elganimi, and S. Beddiaf, "Error performance analysis of UAV-mounted RIS for NOMA systems with practical constraints," *IEEE Communications Letters*, 2024.

[9] F. Khennoufa and K. Abdellatif, "BER performance of co-operative relay NOMA-assisted PS protocol with imperfect SIC and CSI," *International Journal of Sensors Wireless Communications and Control*, vol. 13, no. 2, pp. 99–107, 2023.

[10] F. Khennoufa, A. Khelil, K. Rabie, H. Kaya, and X. Li, "An efficient hybrid energy harvesting protocol for cooperative NOMA systems: Error and outage performance," *Physical Communication*, vol. 58, p. 102061, 2023.

[11] C. Zhong and Z. Zhang, "Non-orthogonal multiple access with cooperative full-duplex relaying," *IEEE Communications Letters*, vol. 20, no. 12, pp. 2478–2481, 2016.

[12] M. Sashiganth, S. Thiruvengadam, and D. S. Kumar, "BER analysis of full duplex NOMA downlink and uplink co-operative user relaying systems over Nakagami-m fading environment," *Physical Communication*, vol. 38, p. 100963, 2020.

[13] F. Kara and H. Kaya, "On the error performance of cooperative-NOMA with statistical CSIT," *IEEE Communications Letters*, vol. 23, no. 1, pp. 128–131, 2018.

[14] ——, "Threshold-based selective cooperative NOMA: Capacity/outage analysis and a joint power allocation-threshold selection optimization," *IEEE Communications Letters*, vol. 24, no. 9, pp. 1929–1933, 2020.

[15] ——, "Threshold-based selective cooperative-NOMA," *IEEE Communications Letters*, vol. 23, no. 7, pp. 1263–1266, 2019.

[16] S. Beddiaf, A. Khelil, F. Khennoufa, and K. Rabie, "On the impact of IQI on cooperative NOMA with direct links in the presence of imperfect CSI," *Physical Communication*, vol. 56, p. 101952, 2023.

[17] F. Khennoufa, K. Abdellatif, and F. Kara, "Bit error rate and outage probability analysis for multi-hop decode-and-forward relay-aided NOMA with imperfect SIC and imperfect CSI," *AEU-International Journal of Electronics and Communications*, vol. 147, p. 154124, 2022.

[18] G. Shen, J. Liu, D. Wang, J. Wang, and S. Jin, "Multi-hop relay for next-generation wireless access networks," *Bell Labs Technical Journal*, vol. 13, no. 4, pp. 175–193, 2009.

[19] M. O. Hasna and M.-S. Alouini, "Outage probability of multihop transmission over Nakagami fading channels," *IEEE Communications Letters*, vol. 7, no. 5, pp. 216–218, 2003.

[20] N. C. Beaulieu and S. S. Soliman, "Exact analytical solution for end-to-end SNR of multihop AF relaying systems," in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*. IEEE, 2011, pp. 580–585.

[21] T. T. Phu, T. D. Tran, M. Voznak *et al.*, "Security-reliability analysis of noma-based multi-hop relay networks in presence of an active eavesdropper with imperfect eavesdropping CSI," *Advances in Electrical and Electronic Engineering*, vol. 15, no. 4, pp. 591–597, 2017.

[22] T.-N. Tran and M. Voznak, "Multi-points cooperative relay in NOMA system with N-1 DF relaying nodes in HD/FD mode for N user equipments with energy harvesting," *Electronics*, vol. 8, no. 2, p. 167, 2019.

[23] F. Khennoufa, A. Khelil, and E. Kenane, "BER and outage performance for multi-hop user-relays NOMA with imperfect SIC and CSI," in *2022 International Conference of Advanced Technology in Electronic and Electrical Engineering (ICATEEE)*. IEEE, 2022, pp. 1–6.

Dual Band antenna for 5G millimeter-wave applications

HAMMAD Ilhem
Tlemcen Telecommunications Laboratory (LTT)
Abou Bekr Belkaid university, Tlemcen, Algeria
Hammad-ttl@outlook.fr

MERIAH Sidi Mohammed
Tlemcen Telecommunications Laboratory (LTT)
Abou Bekr Belkaid university, Tlemcen, Algeria
meriah_m@yahoo.com

MERAD Lotfi
Tlemcen Telecommunications Laboratory (TTL)
Abou Bekr Belkaid university, Tlemcen, Algeria
merad77@yahoo.fr

Abstract—The objective of this paper is to design a patch antenna that operates in two frequency bands for 5G applications. It consists of a multi-slotted shape rectangular radiating element with a 50 Ω microstrip line feed. The overall size of the antenna is 12.50mm × 12mm, it is printed on a Rogers RT Duroid 5880 substrate with a dielectric constant ε_r = 2.2 and a dielectric loss tangent δ= 0.0009, and a thickness of 0.508 mm. We used the commercially available software CST Micro Wave Studio to simulate and analysis the patch antenna characteristics. The results show that the antenna operates at two resonant frequencies with a return loss of -19.36 dB at 29.06GHz and -29.50 dB at 37.68GHz Also, the proposed antenna provides the VSWR<1.5 in each frequency band.

Index Terms— Microstrip patch antenna, Return loss, millimeter-wave.

I. INTRODUCTION

One of the major elements affecting modern wireless communication is a scarcity of viable frequency resources. To address this issue, research has begun in 5G wireless communication at the millimeter frequency band, which extends from 20 GHz to 300 GHz. The frequency band that is typically utilized for 5G research is between 24 GHz and 60 GHz [1]. As a result, 5G employs a broader range of frequencies, such as millimeter waves (mmWave), which allow for high speeds but necessitate a dense base station infrastructure because of their short range.

In the context of this infrastructure, Low profile antennas are used in modern, high quality, high-gain, and simple wireless communication systems to provide reliability, flexibility, and high performance [2]. Microstrip antennas are therefore highly recommended due to their advantage having important attributes such as low cost, light weight, low profile and compatibility with monolithic microwave integrated circuits are widely used in mobile communications. A microstrip patch antenna is a printed antenna that typically has a ground plane on one side of the substrate and a radiating patch on the other. The patch comes in a variety of shapes and is typically composed of copper, silver, or gold [3].

In this paper, we present a novel unique multi-slotted rectangular microstrip antenna operating at two frequencies of 29.06 GHz and 37.68 GHz for 5G applications with high performance.

II. Antenna design

The geometrical configuration of the proposed antenna is shown in Fig. 1 presents a new design of a multi-slotted rectangular microstrip antenna. The patch is printed on a Rogers RT/Duroid5880 substrate, which has a relative permittivity ε_r= 2.2 and a dielectric loss tangent, tan δ = 0.0009. The ground plane is made of a PEC (Perfect Electric Conductor) material with a thickness of 0.017 mm, a length of 12.50 mm and a width of 12 mm.

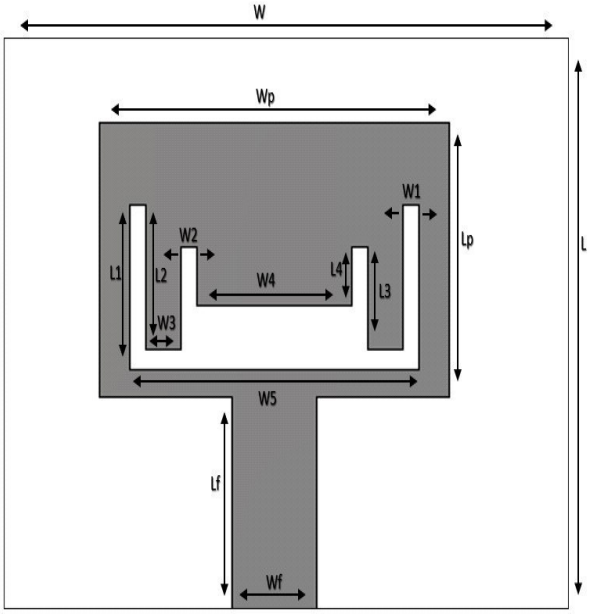


Fig. 1. Proposed multi-slotted patch antenna.

The proposed printed type antenna is fed by a 50 Ω microstrip line of 1.56mm wide and 4mm long. In order to determine the width of the transmission line, the following formula is used [4].

$$z_c = \frac{120\pi}{\sqrt{\varepsilon_{eff}}\left[\frac{w}{h}+1.393+0.667\ln\left(\frac{w}{h}+1.444\right)\right]} \quad (1)$$

The effective dielectric constant is given by:

$$\varepsilon_{eff} = \frac{\varepsilon_{r+1}}{2} + \frac{\varepsilon_{r-1}}{2}\left[1+12\frac{w}{h}\right]^{-\frac{1}{2}} \quad (2)$$

where:

ε_r Is: the dielectric permittivity of the substrate.
w: is the width of the microstrip line.
h: is the height of the dielectric substrate.

The analytical calculations for the dimensions of the single patch element as shown in Fig. 1 are shown below by using equations (3) and (4) [5,6].

$$W_p = \frac{c}{2f_r\sqrt{\frac{\varepsilon_r+1}{2}}} \quad (3)$$

The width of the patch W_P can be determined from equation (3), where c is the velocity of light, fr is the frequency of operation and ε_r is the dielectric permitivity of the substrate.

$$L_p = L_{eff} - 2\Delta L \quad (4)$$

From equation (4), the length of the patch L_p can be determined after obtaining the valued of the effective length of the patch (L_{eff}) and the length of extension (ΔL).

Table 1 shows the optimized dimensions of the proposed antenna demonstrated in Fig. 1.

Table I
DIMENSION OF PROPOSED ANTENNA STRUCTUR

Parameters	Value(mm)
W	12
L	12.50
Wp	5.8
Lp	4.50
Wf	1.56
Lf	4
L1	2.70
L2	2.37
L3	1.67
L4	0.95
W1	0.30
W2	0.30
W3	0.65
W4	2.90
W5	5.40

In this paper, the effect of different widths and lengths of the slots and different patch widths has been analysed to have the Optimized antenna parameters in order to provide lower return loss. The slots on the patch are shown in Fig. 1, where, L and W are the length and width of the slots, W_P is the width of the patch.

The multi-slots on the patch decrease the size of the proposed patch and demonstrate better reflection coefficient at resonant frequencies. Inaddition, Multi-slots limit patch currents at resonance frequencies, resulting in a lower reflection coefficient [7].

III. RESULTS AND DISCUSSION

A. Parametric antenna analysis:

Fig. 2 Shows the variation on the reflection coefficient with different widths of the slots (w1) on the patch. It can be observed that the first resonant frequency increase by decreasing the patch widths, then we get higher reflection coeffitient at resonant frequencie, but the second resonant frequency remains constant and we get lower reflection coefficient by decreasing the patch widths. Thus, the slot width, w1 equals 0.30mm is used as the optimized value.

Fig. 3 Shows the variation on the reflection coefficient with varying lengths of slots (L4) on the patch. The first resonant frequency falls as the slot length increases, but the second resonant frequency remains constant as the slot lengths increase, L4 equals 0.95mm is used as optimized value.

Fig. 4 shows the variation on the reflection coefficient with varying widths of the patch (W_P). The first resonant frequency remains constant, but the second resonant frequency increase with the decreasing of the patch widths, it can also be observed that by decreasing the patch widths, we get lower reflection coeffitient at resonant frequencies, then, changing the widths of the patch effect directly on the antenna reflection coefficient and correspondingly on the antenna matching. Wp equals 5.8mm is used as optimized value.

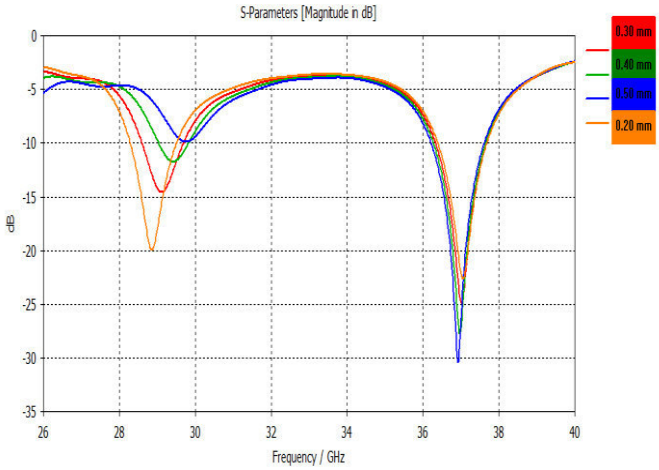


Fig. 2. Reflection coefficient of different slot widths (W1).

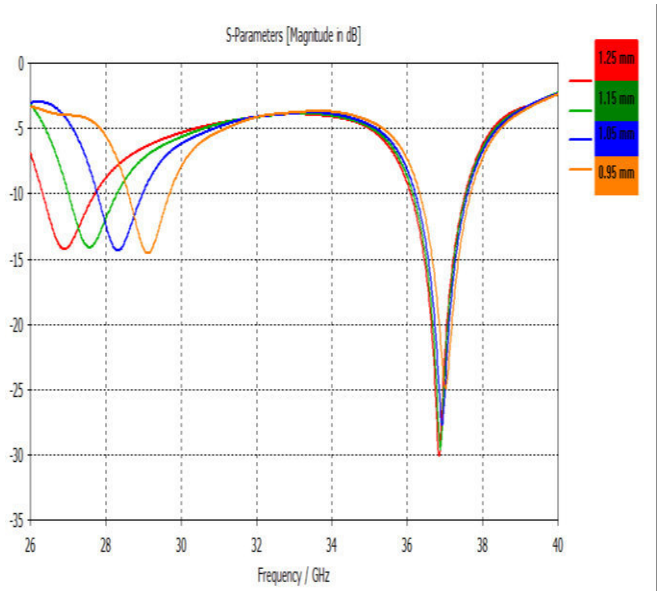


Fig. 3. Reflection coefficient of different slot length (L4).

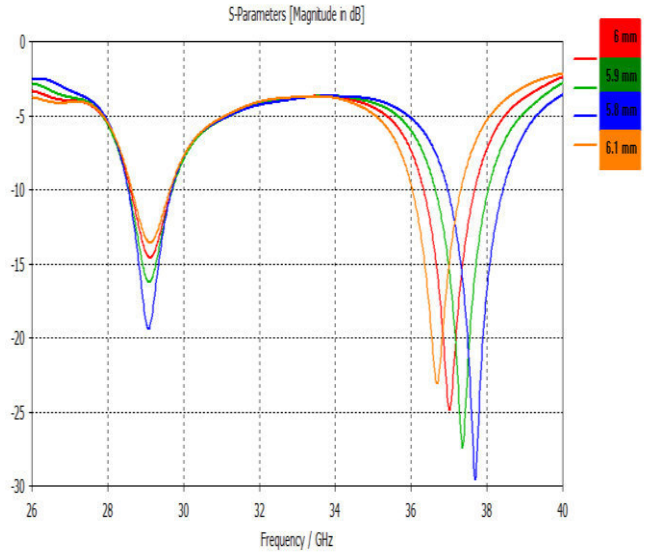


Fig. 4. Reflection coefficient of different patch widths (Wp).

B. S11 parameter & VSWR

The ratio of the incident wave to the reflected wave at the antenna input is known as the reflection coefficient. It is dependent on the transmission line's characteristic impedance z_0 and the antenna's input impedance z_{in} such that [8].

$$\Gamma = \frac{z_{in}-z_0}{z_{in}+z_0} \qquad (5)$$

Fig. 5 illustrate the reflection coefficient of the proposed patch antenna, we can see that the antenna operates in two frequency bands 29.06GHz and 37.68GHz, with peak of -19.36dB and -29.50dB, respectively.

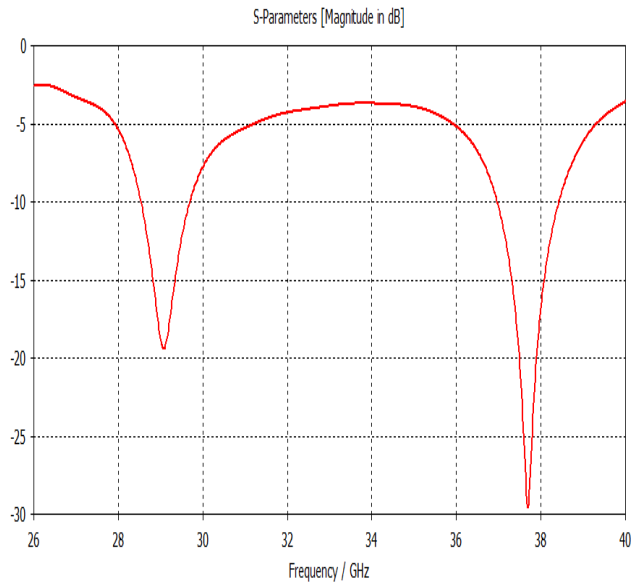


Fig. 5. Reflection coefficient of the proposed antenna.

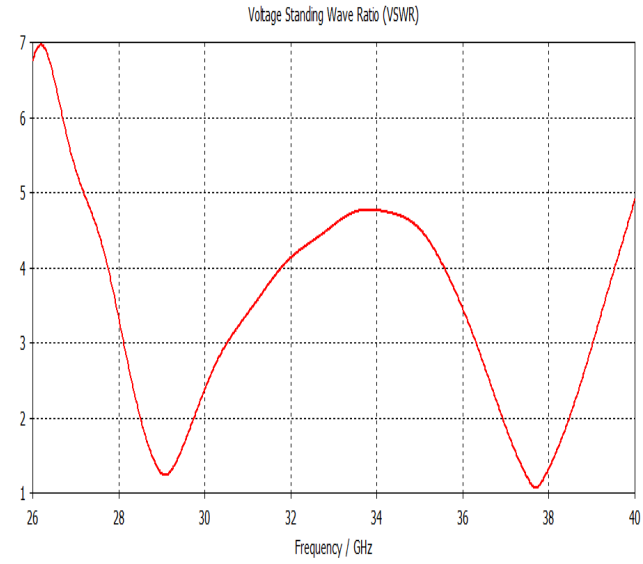


Fig. 6. VSWR of the proposed antenna.

Another method of describing the adaptation is the voltage standing wave ratio, which is determined by the reflection coefficient module [8].

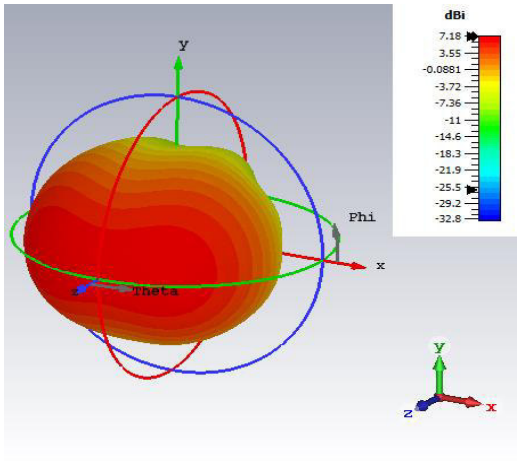
$$VSWR = \frac{1+|\Gamma|}{1-|\Gamma|} \qquad (6)$$

Fig. 6 shows the voltage standing wave ratio of the proposed patch antenna. (VSWR <2), indicates a good match for the antenna.

C. Radiation pattern

The 3D radiation pattern of the proposed patch antenna is shown in the following figure, this antenna has a gain of 7.18dB at 29.06GHz and 7.68dB at 37.68GHz.

a)



b)

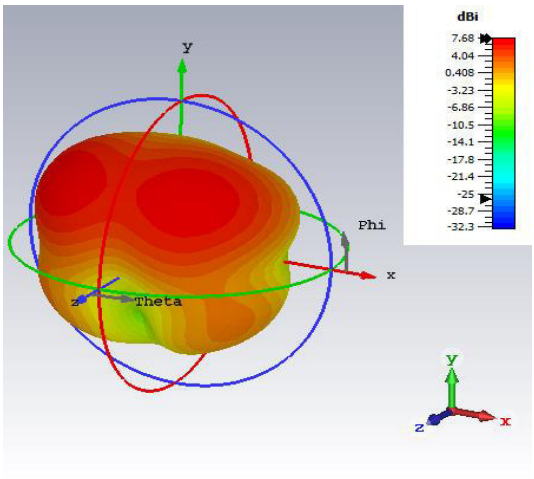


Fig. 7. 3D Radiation pattern of the proposed antenna at (a) 29.06GHz and (b) 37.68GHz.

Table II
COMPARISION OF DESIGNED ANTENNA WITH OTHER REFERENCE ANTENNA

parameter	Our work	Reference antenna[9]
Bandwidth (1).	[28.5-29.7] GHz	[25.2-34.5] GHz
Bandwidth (2).	[36.9-38.4] GHz	[37.5-41.5] GHz
Gain (1)	7.18dB	2.3dBi
Gain (2)	7.68dB	4.8dBi

IV. CONCLUSION

A novel rectangular multi-slotted patch antenna is presented in this work. The antenna is printed on a Rogers RT Duroid 5880 substrate with a dielectric constant ε_r = 2.2 and a thickness of 0.508 mm. Its dimensions are 12.5mmx12mm. It is shown that the effect of different widths and lengths of the slots and different patch widths can be used to improve lower return loss at two resonant frequencies 29.06GHz and 37.68GHz, with peak of -19.36dB and -29.50dB, respectively. The antenna analysis was carried out using CST Microwave Studio Ver. 2019.

REFERENCES

[1] T.Kiran, "Design of Microstrip Patch Antenna for 5g Applications," *IOSR J. Electron. Commun. Eng.*, vol. 13, no. 1, pp. 14–17, 2018, doi: 10.9790/2834-1301011417.

[2] M. M. A. Saman Khabbat Ezzulddin, Sattar Othman Hasan, "Design and Simulation of Microstrip Patch Antenna for 5G Application using CST Studio," *Int. J. Adv. Sci. Technol. Vol.*, vol. 29, no. 04, pp. 7193 – 7205, 2020.

[3] D. Mungur and S. Duraikannan, "Design and Analysis of 28 GHz Millimeter Wave Antenna Array for 5G Communication Systems," *J. Sci. Technol. Eng. Manag. Res. Innov. ISSN*, vol. 1, no. 3, pp. 1–10, 2018.

[4] a J. Wiley, *Antenna Theory Analysis and Design Third Edition (BOOK)*. 2005.

[5] S. Mungur, Dheeraj; Duraikannan, "Microstrip Patch Antenna at 28 GHz for 5G Applications," *J. Sci. Technol. Eng. Manag. Res. Innov.*, vol. 1, no. 1, pp. 20–22, 2018.

[6] A. Z. M. Imran, M. L. Hakim, M. R. Ahmed, M. T. Islam, and E. Hossain, "Design of microstrip patch antenna to deploy unmanned aerial vehicle as UE in 5G wireless network," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 5, pp. 4202–4213, 2021, doi: 10.11591/ijece.v11i5.pp4202-4213.

[7] P. In, "Multi-Slotted Microstrip Patch Antenna for Wireless Communication", *Prog. Electromagn. Res. Lett.*, vol. 10, pp. 11–18, 2009.

[8] D. El Hadri, A. Zakriti, and A. Zugari, "Reconfigurable antenna for Wi-Fi and 5G applicationsR," *Procedia Manuf.*, vol. 46, no. 2019, pp. 793–799, 2020, doi: 10.1016/j.promfg.2020.04.007.

[9] C. Z. Han, G. L. Huang, T. Yuan, and C. Y. D. Sim, "A dual-band millimeter-wave antenna for 5G mobile applications," *2019 IEEE Int. Symp. Antennas Propag. Usn. Radio Sci. Meet. APSURSI 2019 - Proc.*, pp. 1083–1084, 2019, doi: 10.1109/APUSNCURSINRSM.2019.8888328.

Stochastic Geometry-based Analysis of Cellular-Connected UAVs

Dina Alkama[1], Lynda Alkama [2]

[1] *Université de Bejaia, Faculté de Technologie, Laboratoire d'Informatique Médicale (LIMED), Bejaia 06000, Algérie*

[2]*Laboratoire LITAN, École supérieure en Sciences et Technologies de l'Informatique et du Numérique, RN 75, Amizour 06300, Bejaia, Algérie*

dina.alkama@univ-bejaia.dz, alkama@estin.dz

***Abstract*—Cellular-connected unmanned aerial vehicles (UAVs) have become attractive for Internet of Things (IoT) applications due to their flexibility and lower cost. Based on the stochastic geometry theory, this paper investigates the 3D downlink coverage probability in millimeter-wave (mmWave) cellular networks. The terrestrial base stations (TBSs) are distributed according to the homogeneous Poisson point process (HPPP) and are equipped with up-tilt antennas serving UAV users (AUEs) hovering at a fixed altitude h_u. Monte Carlo simulations are used to validate our analytical results.**

***Index Terms*—Cellular networks, Coverage probability, Stochastic geometry, UAV, Up-tilted antennas.**

I. Introduction

With the recent significant technological advancements in the fifth generation (5G) networks, Unmanned Aerial Vehicles (UAVs) have been widely used in various scenarios and fields, including telecommunications, Internet of Things (IoT) and real-time data transmission. The advantage of UAVs is that they can efficiently perform previously difficult or impossible tasks with conventional methods, thanks to their great ability to quickly navigate and rationalize their trajectory [1]. Cellular-connected UAVs can be defined as UAVs equipped with small sensors that are used as aerial users in a terrestrial cellular network to provide reliable and secure connectivity for a range of applications, including monitoring and package delivery.

On the other hand, the deployment of cellular-connected UAVs presents some challenges in terms of coverage due to line-of-sight (LoS) interference from neighbouring cells, UAV altitude and base station (BSs) antenna pattern. In fact, conventional cellular networks were originally designed for ground-based users, where BSs antennas are typically tilted downwards [2]. However, UAVs typically operate at much higher altitudes, so the vertical antenna pattern must be modified to serve UAV UEs (AUEs) achieving 3D coverage. In addition, the altitude of the UAVs must be carefully chosen, taking into account the nature of the environment.

Stochastic geometry-based methods are an important approach widely used to study the coverage problem in cellular networks and provide more accurate estimates of the coverage probability compared to heuristic methods [3]. It is used for complete mathematical modeling by averaging over all network topologies observed from a generic node, weighted by their probability of occurrence due to the irregularity of node locations (BSs, UEs, etc.).

Research on cellular networked UAVs using stochastic geometry has received increasing attention in the literature. The authors in [4] adopted a location-BS cooperation strategy to analyze the coverage probability and local delay of a cellular-connected UAV network. The authors in [5] address LoS interference in cellular-connected UAVs based on BS cooperation, assuming that the BSs are equipped with directional antennas and the gain of the serving BS is the side-lobe gain. In [6], an analysis of the coverage and handover probabilities was performed where the BSs are distributed according to the Poisson Point Process (PPP) and equipped with downtilt antennas. The results obtained showed that the optimal beamwidth of the AUE antenna decreases with the density of the BSs. The coexistence of conventional down-tilt and up-tilt antennas has been investigated in [7], [8], but the ITU channel model was used, making the analysis infeasible for cellular-connected UAVs.

Based on the above motivations, in this paper, using the tools of stochastic geometry, we analyze the coverage probability of the cellular-connected UAVs, where the UAVs are equipped with up-tilted antennas, using a 3D channel model to capture LoS and non-line-of-sight (NLoS) communications. We study the effect of AUE altitude, BS density and up-tilt angle on the coverage probability.

The rest of the paper is organized as follows: in Section 2, we present the network and channel model. In Section 3, we derive the coverage probability expression taking into account the LoS and NLoS propagation. In Section 4, the simulation results are given and the paper is concluded in Section 5.

II. System Model

As illustrated in Figure 1, we consider a 3D downlink millimeter-wave (mmWave) cellular-connected UAV network in a dense urban environment, where the AUEs are randomly distributed in an infinite area and located at a fixed altitude

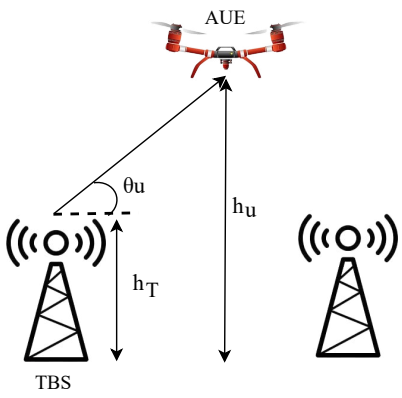


Fig. 1. Illustration of the system model.

h_u. The terrestrial BSs (TBSs) transmit at the same power denoted by P_T, and are assumed to be spatially distributed according to the homogeneous Poisson point process (HPPP) $\Phi = \{M_j, j \in \mathbb{N}\}$, with density and height denoted by λ_T and h_T, respectively. The TBSs are equipped with up-tilt antennas with an up-tilt angle denoted by ω to provide 3D coverage for AUEs. Specifically, each TBS is assumed to be equipped with horizontal omnidirectional antennas and vertical directional antennas with fixed radiation patterns. To perform our downlink analysis, we randomly select a typical AUE and assume that it is located at the origin $(0, 0, h_u)$, i.e., the projection of the typical AUE in the ground is $(0, 0, 0)$.

A. LoS and NLoS Propagation Model

In light of the distinctive attributes of mmWave communications in densely populated urban settings, it is reasonable to posit that the link between the TBS and the AUE may manifest as either a LoS or NLoS connection. The cellular-to-air (C2A) channel differs from the air-to-ground (A2G) channel developed in [9], which is intended for ground users. The system model incorporates TBSs situated at considerable elevations. For macro BSs, the typical height is 25 meters or more [10]. The ITU recommendation in [11], has formulated the probabilities of link states between TBS and user located at a certain height. However, this proposed channel model makes the analysis unmanageable for cellular-connected UAVs. According to [12], the probability of LoS and NLoS communication can be simplified as follows

$$P_{LoS}(W_j) = -b_1 \exp\left(-b_2 \tan^{-1}\left(\frac{h_u - h_T}{W_j}\right)\right) + b_3, \tag{1}$$

$$P_{NLoS}(W_j) = 1 - P_{LoS}(W_j), \tag{2}$$

where b_1, b_2 and b_3 are coefficients related to the environment and the height of TBS.

1) Large Scale Path Loss: The path loss in the link between AUE and TBS is modeled as follows

$$PL_{LoS}(W_j) = \zeta_{LoS}\left(W_j^2 + (h_u - h_T)^2\right)^{\frac{\alpha_{LoS}}{2}}, \tag{3}$$

$$PL_{NLoS}(W_j) = \zeta_{NLoS}\left(W_j^2 + (h_u - h_T)^2\right)^{\frac{\alpha_{NLoS}}{2}}, \tag{4}$$

where ζ_{LoS} and ζ_{NLoS} are the path losses at a reference distance $\left(W_j^2 + (h_u - h_T)^2\right)^{\frac{1}{2}} = 1$Km and α_{LoS} and α_{NLoS} are the path loss exponents for LoS/NLoS cases.

2) Small-Scale Fading: The Nakagami-*m* fading is assumed to all the links. The fading gain of the TBS in $_j$ is denoted by $h^l_{M_j}$ and follows a gamma distribution $h^l_{M_j} \sim \Gamma\left(m_l, \frac{1}{m_l}\right)$, with probability density function (PDF) given by [13]

$$f_H(g) = \frac{m^m g^{m-1}}{\Gamma(m)} \exp(-mg), \tag{5}$$

where m_l is the fading parameter.

B. Antenna Gain

The up-tilted model is accomplished by outfitting the TBS with a uniform linear array of M_t vertically positioned elements. Similar to [14] the antenna gain can be formulated as

$$G_{M_j}(\omega, W_j) = \frac{1}{M_t} \frac{\sin^2 \frac{M_t \pi}{2}\left(\sin(\theta(W_j)) - \sin(\omega)\right)}{\sin^2 \frac{\pi}{2}\left(\sin(\theta(W_j)) - \sin(\omega)\right)} \times 10^{\min\left(-1.2\left(\frac{\theta(W_j)}{\theta_u}\right)^2, \frac{G_{th}}{10}\right)}, \tag{6}$$

where $\theta(W_j) = \tan^{-1}\left(\frac{h_T - h_u}{W_j}\right)$ is the elevation angle between the TBS and the AUE, ω and θ_u are the up-tilt angle and the half power beam width respectively, G_{th} is the threshold for antenna nulls.

C. Cell association

In this study we employ the nearest BS association approach, i.e., the nearest TBS to the typical AUE is assumed to be the serving TBS. The received signal power at a typical AUE from GBS M_j can be calculated as

$$\xi_T = P_T h^l_{M_j} G_{M_j}(W_j) PL_l(W_j)^{-1}, \tag{7}$$

The serving TBS M_0, can be summarized as

$$M_0 = \arg\min_{i \in \mathbb{N}} \left\{\left(W_j^2 + (h_u - h_T)^2\right)^{\frac{1}{2}}\right\}, \tag{8}$$

III. Coverage probability performance

In this section, we derive the coverage probability expression, which is typically defined as the probability that the AUE received a SINR greater than a designated SINR threshold, denoted as γ. The mathematical definition of the coverage probability is as follows:

$$P^c(\gamma) = \mathbb{P}(SINR > \gamma), \tag{9}$$

In this analysis, we exclusively consider the impact of up-tilted BS antennas, while ignoring any interference from down-tilted BS antennas. It is assumed that all TBSs other than the one closest to the AUE deliver interference signals. Denote by W_0 the horizontal distance between the typical AUE and the serving TBS. The received SINR by the typical AUE can be expressed as follows:

$$SINR = \frac{P_T G_{M_0}(\omega, \theta(W_0)) h^l_{M_j} PL_l(W_0)}{I + N_0}, \quad (10)$$

where $l \in \{LoS, NLoS\}$ and N_0 is the noise power.
The aggreagate interference I from the TBSs in v link can be formulated as

$$I = \sum_{M_j \in \Phi / M_0} P_T h^v_{M_j} G_{M_j}(\omega, W_j) PL_v(W_j), \quad (11)$$

The coverage probability expression is given by

$$P^c(\gamma) = \sum_l \sum_{s=1}^{m^l} (-1)^{s+1} \binom{m^l}{s} \int_0^\infty e^{-\beta N_0} \mathcal{L}_I(\beta) P_l(w) \times f_{W_0}(w) dw, \quad (12)$$

where $\beta = \frac{s\,(m_l!)^{\frac{1}{m_l}}\, m_l \gamma PL_l(w)}{P_T G_{M_0}(\omega, \theta(w)}$,
$f_{W_0}(w)$ is the PDF of the serving distance W_0 given by

$$f_{W_0}(w) = 2\pi\lambda w \exp(-\lambda\pi w^2), \quad (13)$$

The Laplace transform of the interference is formulated by

$$\mathcal{L}_I(\beta) = \exp\Big(- 2\lambda_G \pi \int_w^\infty \sum_v \Big[1 - \Big(\frac{m_v}{m_v + \beta P_T G_{M_j}(\omega, \theta(w)) PL_v(w)} \Big)^{m_v} w P_v(w) dw \Big] \Big), \quad (14)$$

IV. Numerical Results

In this section, we validate the analytical expression of the coverage probability using Mathematica and discuss the impact of the UAV altitude and the up-tilt angle on the performance. We use Matlab for the Monte Carlo simulation across a wide simulation area. Table I summarizes the simulation parameters.

In Figure 2, we have plotted the coverage probability as a function of the SINR threshold γ for different TBS densities. First of all, we can see that the simulation results, represented by the circles, are in perfect agreement with the analytical results, validating our model. In addition, we observe that increasing the density of the BSs reduces the coverage probability as the interferences increase, which worsens the received SINR.

Figure 3 shows the effect of altitude and TBS density on the coverage probability. It's obvious that the coverage probability decreases with increasing TBS density and increasing altitude h_u, this is because the distance between the typical AUE and the serving TBS becomes larger, which increases the path loss.

TABLE I
Simulation parameters.

Parameter	Value
λ	10^{-5}
h_T	25m
h_u	$>$ 25m
P_T	20 W
α_{LoS}, α_{NLoS}	2.2, 3.75
$\zeta_{LoS}, \zeta_{NLoS}$	41.1dB, 32.9dB
m_{LoS}, m_{NLoS}	2, 1
b_1, $b2$, b_3	1, 0.106, 1
θ_u	60°
G_{th}	60 dBm
N_0	-90 dBm
γ	-35 dB
Monte Carlo Runs	10^5

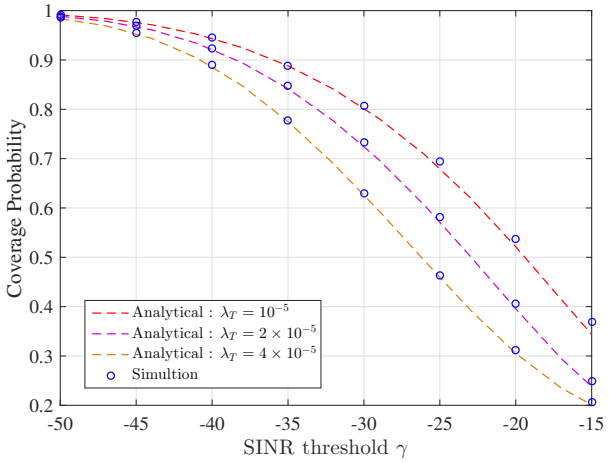


Fig. 2. Coverage probability vs. SINR threshold.

Additionally, the level of interference in the LoS from the adjacent TBSs becomes important and consequently, the SINR reduces.

In Figure 4 we have investigated the impact of the up-tilt angle on the coverage probability for different AUE altitudes. It's evident that the coverage probability initially increases with the tilt angle w and then, after an optimal angle value, it starts to deteriorate and then becomes stable. This can be explained by the fact that the important tilt angle reduces the antenna range. Furthermore, we observe that varying the AUE altitude does not really affect the optimal tilt angle, which is around 50°.

V. Conclusion

In this paper, we have analyzed the coverage probability of cellular-connected UAVs using the stochastic geometry approach. The positions of the TBSs are modeled as HPPP at a given height lower than the AUEs altitude. For 3D coverage, the TBSs are assumed to be equipped with up-tilt antennas to cover the AUEs, where each TBS is assumed to be equipped with horizontal omnidirectional antennas and vertical directional antennas with fixed radiation patterns. The

FT/Boumerdes/NCASEE-24-Conference

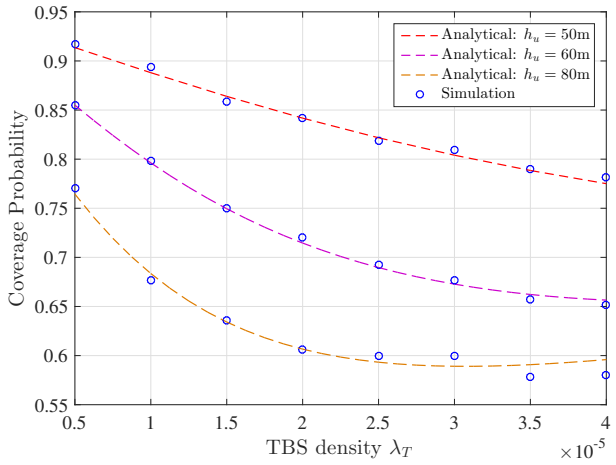


Fig. 3. Coverage probability vs. TBS density.

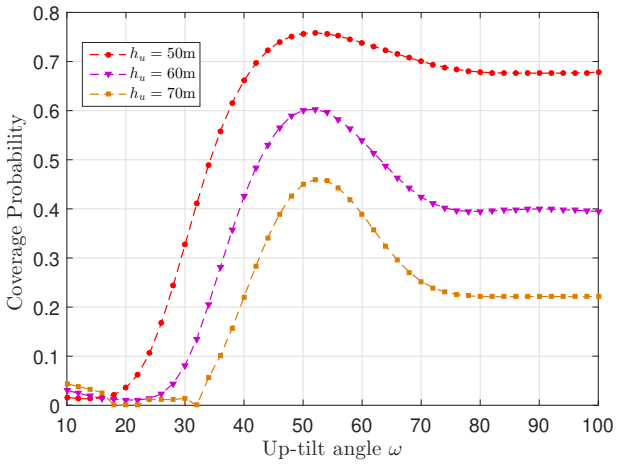


Fig. 4. Coverage probability vs. up-tilt angle.

numerical results have shown that the coverage probability can be improved by carefully adjusting the up-tilt angle and the AUE altitude.

REFERENCES

[1] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE communications surveys & tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.

[2] X. Li, X. Zhou, and S. Durrani, "SWIPT-enabled cellular-connected UAV: Energy harvesting and data transmission," in *2022 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1107–1112, IEEE, 2022.

[3] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

[4] Z. Wang and J. Zheng, "Performance analysis of location-based base station cooperation for cellular-connected UAV networks," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 11, pp. 14787–14800, 2023.

[5] Z. Wang and J. Zheng, "Performance modeling and analysis of base station cooperation for cellular-connected UAV networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 1807–1819, 2021.

[6] H. Sun, C. Ma, L. Zhang, J. Li, X. Wang, S. Li, and T. Q. Quek, "Coverage analysis for cellular-connected random 3D mobile UAVs with directional antennas," *IEEE Wireless Communications Letters*, vol. 12, no. 3, pp. 550–554, 2023.

[7] M. M. U. Chowdhury, I. Guvenc, W. Saad, and A. Bhuyan, "Ensuring reliable connectivity to cellular-connected UAVs with up-tilted antennas and interference coordination," *arXiv preprint arXiv:2108.05090*, 2021.

[8] "Base station antenna uptilt optimization for cellular-connected drone corridors, author=Maeng, Sung Joon and Chowdhury, Md Moin Uddin and Güvenç, İsmail and Bhuyan, Arupjyoti and Dai, Huaiyu," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 4, pp. 4729–4737, 2023.

[9] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, 2014.

[10] D. S. Baum, J. Hansen, J. Salo, G. Del Galdo, M. Milojevic, and P. Kyösti, "An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM)," in *2005 IEEE 61st Vehicular Technology Conference*, vol. 5, pp. 3132–3136, IEEE, 2005.

[11] P. Series, "Propagation data and prediction methods required for the design of terrestrial broadband radio access systems operating in a frequency range from 3 to 60 GHz," *Recommendation ITU-R*, pp. 1410–1415, 2013.

[12] N. Cherif, M. Alzenad, H. Yanikomeroglu, and A. Yongacoglu, "Downlink coverage and rate analysis of an aerial user in vertical heterogeneous networks (VHetNets)," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1501–1516, 2020.

[13] R. Amer, W. Saad, and N. Marchetti, "Mobility in the sky: Performance and mobility analysis for cellular-connected UAVs," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 3229–3246, 2020.

[14] Y. Du, H. Zhang, and J. Peng, "Modeling and coverage analysis for cellular-connected UAVs with up-tilted antenna," *IEEE Communications Letters*, vol. 26, no. 11, pp. 2572–2575, 2022.

Comparative Analysis of Machine Learning Algorithms for IoT Network Traffic Classification

DJEGHLOUF Asmaa[1,2]
Electronics department
[1] Faculty of technology, Hassiba Ben Bouali University
Chlef, Algeria
[2] National School of Cybersecurity, Algiers, Algeria
a.djeghlouf@univ-chlef.dz
asmaa.djeghlouf@enscs.edu.dz

***Abstract*— The Internet of Things (IoT) is transforming industries by enabling the seamless communication between devices, creating new opportunities and efficiencies. However, the exponential growth of connected devices poses significant security challenges, particularly in detecting malicious network traffic. Machine learning (ML) has emerged as a promising solution for addressing these challenges by enabling accurate classification of IoT network traffic and identifying potential threats. This study conducts a comparative analysis of several ML algorithms, including Naive Bayes, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, and Random Forest, for IoT network traffic classification. Using the ACI-IoT 2023 dataset, we evaluate the performance of these models based on accuracy, precision, recall, and F1-score. The results demonstrate that SVM and Random Forest outperform other models, particularly in terms of classification accuracy and robustness in handling imbalanced data. Furthermore, the impact of preprocessing techniques and hyperparameter tuning on model performance is examined. This study highlights the strengths and limitations of each algorithm within the context of IoT network security and provides actionable insights for selecting optimal ML models for real-time IoT traffic monitoring. These findings contribute to advancing IoT security through the application of sophisticated ML techniques.**

Keywords*— *Decision Tree; Internet of Things; K-Nearest Neighbors; Machine learning; Naive Bayes; Support Vector Machine; Random Forest;

I. INTRODUCTION

The rapid growth of Internet of Things (IoT) devices has revolutionized modern connectivity, facilitating applications in areas like healthcare [1], industrial automation [2], smart homes [3], and smart cities [4]. However, alongside these advancements, IoT networks have become a prime target for cyber threats due to the limited computational resources and heterogeneous operating environments of IoT devices [5]. These networks are especially susceptible to attacks such as Distributed Denial of Service (DDoS), data breaches, and unauthorized access.

Traditional methods for enhancing IoT security, such as rule-based and statistical approaches [6], have formed the foundation of network traffic classification techniques. These methods, often employed in intrusion detection systems, rely on predefined rules or basic statistical analyses to detect anomalies. While effective in certain scenarios, these approaches struggle with scalability and adapting to the rapidly evolving nature of IoT data, resulting in challenges like high false-positive rates and reduced accuracy in dynamic environments.

Machine learning (ML) techniques have emerged as a promising solution for IoT network security [7]. By analyzing large volumes of network traffic, ML models can detect complex patterns, adapt to new threats, and significantly improve classification accuracy compared to rule-based methods. Supervised learning algorithms such as Logistic Regression [8], Support Vector Machines (SVM) [9], Naive Bayes (NB) [10], Decision Trees (DT) [11], and K-Nearest Neighbors (KNN) [12] have been extensively studied for IoT traffic classification. These models, while effective, often fall short in capturing the intricate patterns of high-dimensional IoT traffic datasets.

Recent advancements in ensemble methods have further improved the capabilities of machine learning for IoT network traffic classification. Techniques such as Random Forest (RF) [13] and LightGBM [14] combine the strengths of multiple base estimators, achieving higher precision, recall, and robustness compared to individual algorithms. Ensemble methods have demonstrated notable success in identifying uncommon threat classes, but their computational complexity poses challenges for deployment in resource-constrained IoT environments.

IoT network security has garnered significant research attention in recent years. Researchers have explored various methods for network traffic classification, focusing on both traditional machine learning models and advanced ensemble techniques. Hussain et al. [7] used an ensemble of machine and deep learning algorithms and discovered that they can be used effectively for IoT Security. Ioannou et al. [9] recently published a paper in which they used SVM techniques in the computer network security field to detect unauthorized intervention. They shown that the C-SVM achieves up to 100% classification accuracy when evaluated with unknown data taken from the same network topology it was trained with and 81% accuracy when operating in an unknown topology. The OC-SVM that is created using benign activity achieves at most 58% accuracy.

In this work, we present a comprehensive comparative analysis of machine learning algorithms applied to IoT network

traffic classification. Using the ACI-IoT network traffic dataset, we evaluate the performance of algorithms such as Naive Bayes, Decision Trees, KNN, and various ensemble methods across metrics like accuracy, precision, recall, F1-score, and computational complexity. This study aims to identify the most effective algorithm(s) for real-time IoT network security and provide insights into their practical applications.

By combining the advantages of machine learning with the unique characteristics of IoT data, this research seeks to enhance IoT network security. In addition to offering practical recommendations for algorithm deployment, the findings aim to advance the state of the art in IoT network traffic classification and support the development of more secure and adaptive IoT ecosystems.

II. METHODOLOGY

This section provides a detailed description of the methodology employed in this research, which comprises five distinct steps.

A. Data Collection

For this project, The ACI-IoT Network Traffic dataset, which was created especially to capture the distinctive features of IoT network traffic has been collected from Kaggle [15]. This dataset includes a wide range of records from several classes, each of which represents a distinct kind of network activity or security incident, including benign traffic and different kinds of attacks. These attacks are summarized in Table 1.

TABLE I. TRAFFIC BREAKDOWN BY ATTACK TYPE.

Attack type	Number of records	Percentage
Distributed denial of service (ddos)	120,000	12%
Data exfiltration	40,000	4%
Unauthorized access	30,000	3%
Malware activity	10,000	1%
Total malicious traffic	200,000	20%

The ACI-IoT Network Traffic dataset comprises 1,000,000 examples of network traffic, with 800,000 (80%) considered as benign traffic and 200,000 (20%) considered as malicious traffic, these values are summarized in Table 2.

TABLE II. ACI-IOT NETWORK TRAFFIC DATASET OVERVIEW.

Attribute	50 features
Number of Instances	1,000,000
Number of benign traffic instances	800,000 (80%)
Number of malicious traffic instances	200,000 (20%)

Essential characteristics of the dataset, which act as markers for differentiating between typical and unusual traffic, include about 50 features such as packet size, time intervals, protocol type, and source-destination information. Fig.1 shows a review of the original dataset downloaded.

	Flow ID	Src IP	Src Port	Dst IP	Dst Port	Protocol	Timestamp	Flow Duration	Total Fwd Packet	Total Bwd packets	Total Length of Fwd Packet	Total Leng of Bv Pack
count	1231411	1231411	1.231411e+06	1231411	1.231411e+06	1.231411e+06	1231411	1.231411e+06	1.231411e+06	1.231411e+06	1.231411e+06	1.231411e+
unique	865406	159381	NaN	89089	NaN	NaN	75403	NaN	NaN	NaN	NaN	Na
top	8.6.0.1-8.0.6.4-0-0-0	192.168.1.45	NaN	192.168.1.1	NaN	NaN	2023-10-30 12:21:01	NaN	NaN	NaN	NaN	Na
freq	3685	621729	NaN	255771	NaN	NaN	16208	NaN	NaN	NaN	NaN	Na
mean	NaN	NaN	3.680030e+04	NaN	1.540297e+04	6.433792e+00	NaN	5.236489e+06	5.456022e+00	5.439780e+00	4.998453e+03	2.078230e+
std	NaN	NaN	2.331748e+04	NaN	2.044498e+04	4.919235e+00	NaN	1.188271e+07	1.819997e+02	2.185122e+02	4.582882e+05	1.237793e+
min	NaN	NaN	0.000000e+00	NaN	0.000000e+00	0.000000e+00	NaN	0.000000e+00	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+
25%	NaN	NaN	5.353000e+03	NaN	5.300000e+01	6.000000e+00	NaN	3.357000e+03	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+
50%	NaN	NaN	4.814600e+04	NaN	3.017000e+03	6.000000e+00	NaN	6.631000e+03	1.000000e+00	1.000000e+00	0.000000e+00	0.000000e+
75%	NaN	NaN	5.529500e+04	NaN	3.096650e+04	6.000000e+00	NaN	3.117183e+06	2.000000e+00	1.000000e+00	2.400000e+01	0.000000e+
max	NaN	NaN	6.553500e+04	NaN	6.553500e+04	1.700000e+01	NaN	6.094632e+07	7.418000e+04	7.247000e+04	2.382375e+08	2.646032e+

Fig. 1. General view of the dataset used.

Due to its scale and relevance, the dataset provides an ideal benchmark for evaluating the effectiveness of machine learning algorithms in classifying IoT network traffic.

B. Data Cleaning

Prior to analysis, the dataset was examined for any missing or inconsistent values, which could potentially impact model performance. Missing values were either imputed using suitable statistical techniques or removed if they constituted a small percentage of the dataset. Additionally, duplicated entries were identified and eliminated to ensure data integrity. Certain categorical features, such as protocol types, were standardized to enable uniform processing across all records.

Data cleaning involves preparing the data in a format that allows for effective processing by a machine, facilitating the construction of an efficient machine learning model.

C. Feature Engineering

To optimize the dataset for machine learning classification, several feature engineering steps were performed:

1) Encoding Categorical Variables

Features such as protocol type and source-destination ports were transformed into numerical representations using one-hot encoding. This step was necessary to allow machine learning algorithms to process these categorical attributes effectively.

2) Scaling Numerical Features

Continuous features, such as packet size and time intervals, were scaled to a standardized range using either Min-Max scaling or Standardization. This step ensured that features on different scales would not disproportionately influence model training.

3) Feature Selection

Given the high dimensionality of the dataset, feature selection techniques, including Recursive Feature Elimination (RFE) and correlation analysis, were employed to reduce feature redundancy and retain only the most relevant features for classification. Reducing the feature space helped improve computational efficiency and mitigate the risk of overfitting.

D. Data Splitting

To evaluate model performance, the dataset was split into training and testing subsets, with an 80-20 split ratio.

The training subset was used to fit the machine learning models, while the testing subset was held out for final evaluation. This split was stratified to maintain a balanced representation of each class in both subsets, thereby preventing class imbalance issues that could skew the results.

E. Handling Class Imbalance

IoT network traffic datasets often exhibit class imbalance, as certain types of attack traffic may be much rarer than benign traffic.

In this dataset, significant class imbalance was observed, with certain attack classes represented by far fewer records. To address this, several strategies were implemented:

1) Oversampling Minority Classes

The Synthetic Minority Over-Sampling Technique (SMOTE) was applied to the training dataset to artificially increase the number of samples in underrepresented classes.

2) Class Weight Adjustment

For models that support class weights, such as Decision Trees and SVMs, class weights were adjusted to penalize misclassifications in minority classes more heavily, thereby encouraging the model to focus on accurate classification across all classes.

F. Classification & Prediction

The methodology for this study involves selecting, training, and evaluating multiple machine learning algorithms to classify IoT network traffic into distinct categories, including benign and various malicious activities. This section outlines the machine learning models, hyperparameter tuning approach, and evaluation metrics used to assess model performance.

To achieve a comprehensive comparison, we applied various machine learning models known for their effectiveness in classification tasks, particularly in network security applications:

1) Naive Bayes (NB)

A probabilistic classifier based on Bayes' theorem. This method assumes feature independence and is computationally efficient, making it suitable for large datasets with minimal training time.

2) Support Vector Machine (SVM)

An effective classification model that seeks to maximize the margin between different classes. SVM is known for its robustness, particularly in high-dimensional spaces, making it suitable for complex datasets like network traffic data.

3) K-Nearest Neighbors (KNN)

A non-parametric, instance-based learning algorithm. The KNN classifier labels each data point based on the majority class of its K nearest neighbors, which makes it simple but computationally intensive for large datasets.

4) Decision Tree (DT)

A tree-based model that splits data based on feature importance to maximize classification purity. Decision Trees are interpretable models that perform well with imbalanced datasets by handling categorical and continuous variables.

5) Random Forest (RF)

An ensemble method that builds multiple decision trees and averages their outputs to improve accuracy and control overfitting. Random Forest is particularly effective for high-dimensional datasets and can handle noisy data robustly.

G. Training and Validation

The dataset was split into training, validation, and testing sets. After splitting the dataset into an 80-20 train-test ratio, an additional validation set was created from the training data (70% for training and 30% for validation). This split was stratified to maintain the class distribution across all subsets. Each model was trained on the training data and validated using the validation set. The validation step helped to refine model selection and avoid overfitting on the training data.

H. Evaluation Metrics

To assess the performance of each model in classifying IoT network traffic accurately, we utilized multiple metrics.

Since class imbalance was a significant factor in the dataset, metrics were selected to capture different aspects of model performance. These functions, namely Accuracy, Precision, Recall and F1-score, are derived from the analysis of the confusion matrix detailed in Table 2.

TABLE III. CONFUSION MATRIX RELATED TO NETWORK TRAFFIC CLASSIFICATION.

		Predicted	
		malicious	***benign***
True	***malicious***	True Positive (TP)	False positive (FP)
	benign	False Negative (FN)	True negative (TN)

In a classification issue, the confusion matrix is seen as a summary of the predicted outcomes. In this matrix, the number of properly predicted malicious emails is indicated by TP, the number of correctly forecast malicious mails is indicated by FP, the number of correctly predicted malicious emails is indicated by FN, and the number of correctly anticipated ***benign*** is indicated by TN.

a) Accuracy: represents the proportion of correctly classified instances made by the model. It is defined as shown in Eq. (1).

$$Accuracy = \frac{TP + TN}{\mathrm{TP + FP + FN + TN}} \qquad (1)$$

b) Precision: The precision measures the ratio of correctly predicted malicious traffic to all predicted malicious traffic. Precision is given in Eq. (2).

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

c) Recall: The recall quantifies the ability of the model to identify all actual malicious traffic. mathematically, recall is defined by Eq (3).

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

d) F1_score: F1 score is the harmonic mean of precision and recall, providing a balanced measure of model performance, presented as follows:

$$f_1_score = \frac{2(precision \times recall)}{(precision + recall)} \quad (4)$$

III. Experimental Results

In this section, we present the evaluation results for each machine learning model on the IoT network traffic dataset, analyzing their performance based on accuracy, precision, recall, F1-score, and other relevant metrics. The performance of each model is compared, followed by a discussion of the implications of these results for IoT network security.

Through training with the Naïve Bayes algorithm, the model showed poor overall performance with an accuracy of only 75.7% was attained. The precision for identifying malicious traffic stood at 0.71, with a corresponding recall rate of 0.69. Detailed results are depicted in Fig. 2.

```
---------Naive Bayes model---------
Accuracy: 0.7575456359852156
Precision: 0.7159665233548625
Recall: 0.6933256685912563
F1 Score: 0.7022549638791256
```

Fig. 2. Training result of the Naïve Bayes model.

Naïve Bayes algorithm simplicity and assumption of feature independence contributed to its limitations, as it struggled with the complexities and high dimensionality of the network traffic data.

Upon training the model with the SVM algorithm, the algorithm demonstrated an accuracy of 99.6%. It exhibited a precision of 0.88, with a recall rate of 0.85 for malicious traffic patterns. The evaluation results for the SVM model are presented in Fig 3.

```
-----------SVM model-----------
Accuracy: 0.9962548796352145
Precision: 0.8824569587547863
Recall: 0.8565544122563585
F1 Score: 0.8624789645789562
```

Fig. 3. Training result of the SVM model.

SVM performed exceptionally well, achieving 99.6% accuracy. Its ability to find an optimal margin between classes contributed to its high precision and recall, making it one of the top performers in this study.

When employing the KNN algorithm for model training, an accuracy of 94.2% was achieved. The precision reached 0.70, while the recall rate was 0.69. Detailed results can be observed in Fig. 4.

```
---------KNN model---------
Accuracy: 0.9425888563325623
Precision: 0.7002251536665287
Recall: 0.6998545520365245
F1 Score: 0.6933254452332019
```

Fig. 4. Training result of the KNN model.

KNN achieved good accuracy, it was computationally intensive and struggled with class imbalance. However, its performance highlights its potential usefulness for smaller IoT datasets with more balanced classes.

The Decision tree Classifier exhibited an accuracy of 96.8%, with a precision for malicious traffic detection of 0.81 and a recall of 0.82. The findings from the model training with the Random Forest Classifier are illustrated in Fig.5.

```
-----Decision tree model-----
Accuracy: 0.9687544800261056
Precision: 0.8114205895210251
Recall: 0.8222451108745019
F1 Score: 0.8225447859012587
```

Fig. 5. Training result of the decision tree model.

Decision Tree achieved moderate performance. While interpretable, it tended to overfit on certain classes, impacting precision and recall on less frequent classes.

By training with the Random Forest classifier, an accuracy of 99.3% was attained. Consequently, it demonstrated superior performance compared to the preceding models after the SVM model. The precision for identifying malicious traffic stood at 0.87, with a corresponding recall rate of 0.87. Detailed results are depicted in Fig. 6.

```
-----Random Forest  model-----
Accuracy: 0.9932420015698527
Precision: 0.8745214852014589
Recall: 0.8795425487596015
F1 Score: 0.8745126980159875
```

Fig. 6. Training result of the Random Forest model.

Random Forest performed well, with 99.3% accuracy, leveraging multiple decision trees to improve robustness. However, its complexity increased computational time, especially during training.

Variations in performance were observed across Accuracy, Precision, Recall, and f1_score parameters among the five machine learning algorithms. A summarized representation of these observations is provided in Fig.7.

Fig.7 has the figure has clearly shown that the Naïve Bayes algorithm displayed the lowest accuracy at 75.7%, along with the minimum f1_score in in distinguishing between benign and malicious traffic patterns. On the other hand, Support Vector Machine showed a reasonable accuracy rate of 99.6%, boasting a precision of 0.88, recall and f1_score of 0.85 and 0.86,

respectively. As a result, it outperformed the Decision Tree Classifier and KNN model, which exhibited lower accuracy, precision, recall, and f1_score, albeit achieving an accuracy of 94.2% and 96.8%, respectively.

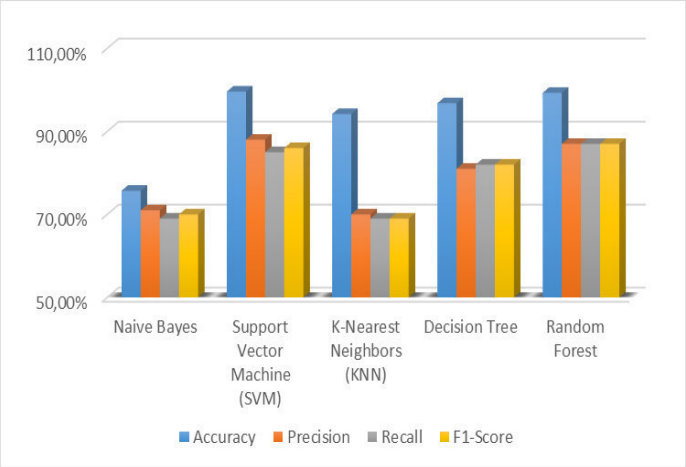


Fig. 7. Results of the fifth machine learning models.

While Random Forest model demonstrated results comparable to the SVM Classifier, it exhibited lower values in terms of accuracy and precision of 99.3% and 0.87, respectively. Overall, the SVM algorithm emerged with the highest accuracy, coupled with the lowest error rate for malicious traffic. Consequently, in the realm of distinguishing between benign and malicious traffic patterns, the SVM algorithm emerges as the most effective technique among those investigated for training the model.

IV. Conclusions

This paper presented a comparative analysis of various machine learning algorithms for IoT network traffic classification, evaluating their effectiveness in identifying and distinguishing between benign and malicious traffic patterns.

Using the ACI-IoT network traffic dataset, we implemented and assessed several models, including Naive Bayes, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree and Random Forest classifier.

Our findings indicate that Support Vector Machine demonstrated strong classification capabilities, though it required careful parameter tuning. Simpler models, such as Naive Bayes and KNN, underperformed due to their limited capacity to handle complex decision boundaries and imbalanced data distributions. The results underline the potential of tree-based ensemble models for high-stakes IoT applications where accurate detection of malicious activity is essential.

Several practical insights emerge from this study. First, model choice should balance accuracy with computational efficiency, especially for real-time IoT systems where processing resources may be limited.

SVM stands out as a particularly promising model for such environments due to its computational efficiency and high accuracy. Second, addressing class imbalance through techniques such as sampling or advanced ensemble methods is crucial for enhancing the reliability of IoT network security models.

References

[1] M. Haghi Kashani, M. Madanipour, M. Nikravan, P. Asghari, and E. Mahdipour, "A systematic review of IoT in healthcare: Applications, techniques, and trends," J. Netw. Comput. Appl., vol. 192, p. 103164, 2021, doi: https://doi.org/10.1016/j.jnca.2021.103164.

[2] A. Deshpande, P. Pitale, and S. Sanap, "Industrial automation using Internet of Things (IOT)," Int. J. Adv. Res. Comput. Eng. Technol., vol. 5, no. 2, pp. 266–269, 2016.

[3] D. Mocrii, Y. Chen, and P. Musilek, "IoT-based smart homes: A review of system architecture, software, communications, privacy and security," Internet of Things, vol. 1, pp. 81–98, 2018.

[4] H. Arasteh et al., "Iot-based smart cities: A survey," in 2016 IEEE 16th international conference on environment and electrical engineering (EEEIC), IEEE, 2016, pp. 1–6.

[5] E. Altulaihan, M. A. Almaiah, and A. Aljughaiman, "Cybersecurity threats, countermeasures and mitigation techniques on the IoT: Future research directions," Electronics, vol. 11, no. 20, p. 3330, 2022.

[6] M. W. A. Ashraf, A. R. Singh, A. Pandian, R. S. Rathore, M. Bajaj, and I. Zaitsev, "A hybrid approach using support vector machine rule-based system: detecting cyber threats in internet of things," Sci. Rep., vol. 14, no. 1, p. 27058, 2024.

[7] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, "Machine learning in IoT security: Current solutions and future challenges," IEEE Commun. Surv. Tutorials, vol. 22, no. 3, pp. 1686–1721, 2020.

[8] A. Churcher et al., "An experimental analysis of attack classification using machine learning in IoT networks," Sensors, vol. 21, no. 2, p. 446, 2021.

[9] C. Ioannou and V. Vassiliou, "Network attack classification in IoT using support vector machines," J. Sens. actuator networks, vol. 10, no. 3, p. 58, 2021.

[10] D. Jeevaraj, "Feature selection model using naive bayes ML algorithm for WSN intrusion detection system," Int. J. Electr. Comput. Eng. Syst., vol. 14, no. 2, pp. 179–185, 2023.

[11] Z. Azam, M. M. Islam, and M. N. Huda, "Comparative analysis of intrusion detection systems and machine learning based model analysis through decision tree," IEEE Access, 2023.

[12] M. Mohy-Eddine, A. Guezzaz, S. Benkirane, and M. Azrour, "An efficient network intrusion detection model for IoT security using K-NN classifier and feature selection," Multimed. Tools Appl., vol. 82, no. 15, pp. 23615–23633, 2023.

[13] N. Kandhoul, S. K. Dhurandher, and I. Woungang, "Random forest classifier-based safe and reliable routing for opportunistic IoT networks," Int. J. Commun. Syst., vol. 34, no. 1, p. e4646, 2021.

[14] S. Dalal et al., "Optimized LightGBM model for security and privacy issues in cyber-physical systems," Trans. Emerg. Telecommun. Technol., vol. 34, no. 6, p. e4771, 2023.

[15] "IoT Network Security Dataset." [Online]. Available: https://www.kaggle.com/code/aaradhyabadal/iot-network-security-anomaly-sml-project/input

Performance Analysis of Uplink NOMA under In-phase and Quadrature Imbalance

Beddiaf Safia[1], Abdellatif Khelil[1], Faical Khennoufa[1] and Beddiaf Walid[2]
[1]*LGEERE Laboratory, Department of Electrical Engineering, Echahid Hamma Lakhdar University, El-Oued, Algeria*
[2]*LEVE Laboratory, Department of Architecture, University Mouhamed Khidar, Biskra, Algeria*
Email: beddiaf-safia@univ-eloued.dz, abdellatif-khelil@univ-eloued.dz, khennoufa-faical@univ-eloued.dz, w.beddiaf@univ-biskra.dz.

***Abstract*—Non-Orthogonal Multiple Access (NOMA) has emerged as a promising solution to meet the demands of 5G networks, offering improved connectivity, low latency, and high spectral efficiency. However, practical wireless systems face challenges due to imperfections in RF frontend components, particularly in-phase (I) and quadrature-phase (Q) imbalance (IQI). These imperfections lead to amplitude and phase discrepancies between I and Q branches, causing poor image rejection and degrading system performance. While ideal systems achieve perfect phase shifts for optimal performance, real-world direct-conversion transceivers suffer from analog front-end impairments, resulting in signal distortion and increased error rates. This paper provides a comprehensive analysis of the outage probability (OP) and ergodic capacity (EC) performance in uplink NOMA systems under the impact of IQI, considering a Rayleigh fading channel and direct transmission between the source and users.**

***Index Terms*—outage , NOMA cooperative relay, power splitting, SIC, and CSI.**

I. Introduction

The emergence of Non-Orthogonal Multiple Access (NOMA) presents a significant breakthrough in wireless communication, offering enhanced connectivity, low latency, and superior spectral efficiency to meet the demands of 5G networks. However, practical communication systems face challenges due to imperfections in RF frontends, specifically in-phase (I) and quadrature-phase (Q) imbalance (IQI). IQI refers to discrepancies in amplitude and phase between the I and Q branches of the transceiver, which lead to imperfect image rejection and degraded system performance. In ideal systems, mixers maintain balanced I and Q branches with a 90-degree phase shift for optimal image band attenuation. In contrast, real-world systems with direct-conversion transceivers are prone to analog front-end impairments, resulting in phase and amplitude mismatches that distort the signal and increase error rates.

Driven by practical considerations, recent studies have focused on developing models to mitigate, reduce, or even exploit the effects of IQI. In [1], the authors derive the Signal-to-Interference-plus-Noise Ratio (SINR) by accounting for the correlation between subcarriers in Orthogonal Frequency Division Multiplexing (OFDM) systems. A more precise analytical expression for the average subcarrier SINR in beamforming OFDM systems, considering joint transmit-receive IQI, is presented in [2], offering insights into the influence of various system parameters on performance. The impact of IQI is further examined in [3], where the authors highlight both its detrimental and, in some cases, beneficial effects, depending on the specific scenario. In [4], the combined effects of IQI and phase noise (PHN) on beamforming OFDM direct-conversion transceivers are analyzed, with an exact normalized mean squared error expression and several special cases explored. The impact of IQI imbalance in amplify-and-forward (AF) relaying within the OFDM system is discussed in [5], where an analytical approximation for Outage Probability (OP) and Error Vector Magnitude (EVM) is derived. The effect of RF impairments on NOMA systems is investigated in [6], with a closed-form expression for conditional Pairwise Error Probability (PEP) for two users in the downlink NOMA system presented in [7]. The outage probability of IQI impairment in NOMA systems under Nakagami-m fading is explored in [8]–[10]. Security performance in multi-antenna systems is addressed in [11], While [12] examines the impact of IQI on the security of ambient backscatter (AmBC) NOMA systems, deriving the outage probability in the presence of an eavesdropper for two NOMA users. Additionally, the works of [13]–[15] investigate the NOMA performance in terms of OP for cooperative, cognitive radio, and vehicle-to-vehicle systems.

Inspired by this motivation, the current study focuses on quantifying and examining the impact of IQI on downlink and uplink direct NOMA systems. The primary objective of this article is to establish a comprehensive framework for analyzing both downlink and uplink NOMA systems under various IQI scenarios. This framework encompasses cases involving the combination of the Transmit (TX) IQI and Receive (RX) IQI. Specifically, we derive the Signal-to-Interference-plus-Noise Ratio (SINR) for both uplink and downlink NOMA systems operating in the presence of TX and RX IQI.

The rest of the paper is presented as follows: System model of the direct link of the NOMA system model is presented in Section II. In Section III, we analyze the OP and EC performance of the users under the imperfect CSI. Analytical results are described in Section IV. Finally, we conclude the paper in Section V.

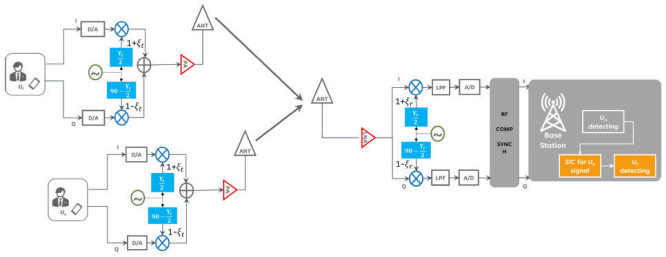


Fig. 1: System model of the uplink NOMA in the presence of IQI.

II. System Model

The Uplink NOMA enables multiple transmitters to send their signals to an access point using the same time and frequency resources. The primary characteristic of uplink NOMA is its power allocation strategy, where the signal from the near user transmitter is assigned a higher power level than that from the far user transmitter. In this system model, we consider an access point (AP), a far user (U1), and a near user (U2), as illustrated in Fig. 1. We assume that all nodes are equipped with a single antenna and are subject to IQI impairments.

1) NOMA TX impaired IQI

The received signal, affected by transmitter impairments, can be expressed as

$$\begin{aligned} y_{IQI}^{TX} =& u_t \left(h_1 \sqrt{\alpha_1 P} x1 + h_2 \sqrt{\alpha_2 P} x2 \right) + \\ & v_t \left(h_1 \sqrt{\alpha_1 P} x1 + h_2 \sqrt{\alpha_2 P} x2 \right)^* + n. \end{aligned} \tag{1}$$

The SINR of the U_2 can be expressed as

$$\gamma_2^{TX} = \frac{\mu_t^2 |h_2|^2 P \alpha_2}{\mu_t^2 |h_1|^2 P \alpha_1 + \nu_t^2 |h_1|^2 P \alpha_1 + \nu_t^2 |h_2|^2 P \alpha_2 + N_0}. \tag{2}$$

According to the uplink NOMA protocol, U_1 performs successive interference cancellation (SIC) to eliminate interference from U2. Assuming ideal SIC and perfect channel state information (CSI), the SINR for U_1 is expressed as

$$\gamma_1^{TX} = \frac{\mu_t^2 |h_1|^2 P \alpha_1}{\nu_t^2 |h_1|^2 P \alpha_1 + \nu_t^2 |h_2|^2 P \alpha_2 + N_0}. \tag{3}$$

2) NOMA RX impaired IQI

The received signal for uplink NOMA, considering an ideal transmitter (TX) and an impaired receiver (RX), can be expressed as

$$\begin{aligned} y_{IQI}^{RX} =& u_r \left(h_1 \sqrt{\alpha_1 P} x1 + h_2 \sqrt{\alpha_2 P} x2 \right) + \\ & v_r \left(h_1 \sqrt{\alpha_1 P} x1 + h_2 \sqrt{\alpha_2 P} x2 \right)^* + u_r n + v_r n^*. \end{aligned} \tag{4}$$

The SINR of the U_2 and U_1 can be defined, respectively as

$$\gamma_2^{RX} = \frac{\mu_r^2 |h_2|^2 P \alpha_2}{\mu_r^2 |h_1|^2 P \alpha_1 + \nu_r^2 |h_1|^2 P \alpha_1 + \nu_r^2 |h_2|^2 P \alpha_2 + (\mu_r^2 + \nu_r^2) N_0}, \tag{5}$$

$$\gamma_1^{RX} = \frac{\mu_r^2 |h_1|^2 P \alpha_1}{\nu_r^2 |h_1|^2 P \alpha_1 + \nu_r^2 |h_2|^2 P \alpha_2 + (\mu_r^2 + \nu_r^2) N_0}. \tag{6}$$

3) NOMA TX and RX impaired IQI

The received signal for the uplink NOMA when both TX and RX are impaired can be defined as

$$\begin{aligned} y_{IQI}^{TX/RX} =& u_r \left(u_t \left(h_1 \sqrt{\alpha_1 P} x1 + h_2 \sqrt{\alpha_2 P} x2 \right) + \right. \\ & \left. v_t \left(h_1 \sqrt{\alpha_1 P} x1 + h_2 \sqrt{\alpha_2 P} x2 \right)^* + n \right. \\ & \left. + v_r \left(u_t \left(h_1 \sqrt{\alpha_1 P} x1 + h_2 \sqrt{\alpha_2 P} x2 \right) + \right. \right. \\ & \left. \left. v_t \left(h_1 \sqrt{\alpha_1 P} x1 + h_2 \sqrt{\alpha_2 P} x2 \right)^* + n^* \right. \right. \end{aligned} \tag{7}$$

The SINR of the U_2 and U_1 while both the TX and RX are impaired can be given as

$$\gamma_2^{TX/RX} = \frac{\epsilon_1 |h_2|^2 P \alpha_2}{\epsilon_1 |h_1|^2 P \alpha_1 + \epsilon_2 |h_1|^2 P \alpha_1 + \epsilon_2 |h_2|^2 P \alpha_2 + (\mu_r^2 + \nu_r^2) N_0}, \tag{8}$$

$$\gamma_1^{TX/RX} = \frac{\epsilon_1 |h_1|^2 P \alpha_1}{\epsilon_2 |h_1|^2 P \alpha_1 + (\mu_r^2 + \nu_r^2) N_0}. \tag{9}$$

III. Performance Analysis of the uplink NOMA with IQI

A. Outage probability

In this subsection, the OP of the uplink NOMA is analyzed for three scenarios: transmitter (TX) impaired by IQI, receiver (RX) impaired by IQI, and both TX and RX jointly impaired by IQI. over the Rayleigh fading channel for the two users .

1) NOMA TX impaired IQI

The OP of the U_2 and U_1 when the TX is impaired can be given, respectively as

$$\begin{aligned} P_{out,2}^{TX} &= \left(|h_2|^2 \leq \frac{\gamma_{th,2} \left((\mu_t^2 + \nu_t^2) \alpha_1 |h_1|^2 P + N_o \right)}{(\mu_t^2 + \nu_t^2) \alpha_2 |h_1|^2 P} \right) \\ &= 1 - \exp \left(- \frac{\gamma_{th,2} N_o}{P \left(\mu_t^2 \alpha_2 - \gamma_{th,2} \nu_t^2 \alpha_2 \right) \phi_2} \right) \times \\ & \quad \frac{P \left(\mu_t^2 \alpha_2 - \gamma_{th,2} \nu_t^2 \alpha_2 \right) \phi_2}{\gamma_{th,2} (\mu_t^2 + \nu_t^2) \alpha_1 P \phi_1 + P \left(\mu_t^2 \alpha_2 - \gamma_{th,2} \nu_t^2 \alpha_2 \right) \phi_2}, \end{aligned} \tag{10}$$

$$\begin{aligned} P_{out,1}^{TX} &= \left(|h_{1,S}|^2 \leq \frac{\gamma_{th,1} N_o}{(\mu_t^2 + \nu_t^2) \alpha_1 |h_{1,S}|^2 P} \right) \\ &= 1 - \exp \left(- \frac{\gamma_{th,1} N_o}{P \left(\mu_t^2 \alpha_1 - \gamma_{th,1} \nu_t^2 \alpha_1 \right) \phi_{1,S}} \right). \end{aligned} \tag{11}$$

2) NOMA RX impaired IQI

Assuming an ideal TX, the OP of the U_2 and U_1 while the RX is impaired can be expressed, respectively as

$$P_{out,2}^{RX} = 1 - \exp\left(-\frac{\gamma_{th,2}(\mu_r^2+\nu_r^2)N_o}{P\left(\mu_r^2\alpha_2 - \gamma_{th,2}\nu_r^2\alpha_2\right)\phi_2}\right) \times \frac{P\left(\mu_r^2\alpha_2 - \gamma_{th,2}\nu_r^2\alpha_2\right)\phi_{2,S}}{\gamma_{th,2}(\mu_r^2+\nu_r^2)\alpha_1 P\phi_1 + P\left(\mu_r^2\alpha_2 - \gamma_{th,2}\nu_r^2\alpha_2\right)\phi_2}, \quad (12)$$

3) NOMA TX and RX impaired IQI

The OP of U_2 and U_1 is obtained, respectively as

$$P_{out,2}^{TX/RX} = 1 - \exp\left(-\frac{\gamma_{th,2}(\mu_r^2+\nu_r^2)N_o}{P\Big(\epsilon_1\alpha_2 - \gamma_{th,2}\epsilon_2\alpha_2\Big)\phi_2}\right) \times \frac{P\Big(\epsilon_1\alpha_2 - \gamma_{th,2}\epsilon_2\alpha_2\Big)\phi_2}{\gamma_{th,2}\Big(\epsilon_1+\epsilon_2\Big)\alpha_1 P\phi_1 + P\Big(\epsilon_1\alpha_2 - \gamma_{th,2}\epsilon_2\alpha_2\Big)\phi_2}, \quad (13)$$

$$P_{out,1}^{TX/RX} = 1 - \exp\left(-\frac{\gamma_{th,2}(\mu_r^2+\nu_r^2)N_o}{P\Big(\epsilon_1\alpha_2 - \gamma_{th,2}\epsilon_2\alpha_2\Big)\phi_2}\right) \times \frac{P\Big(\epsilon_1\alpha_2 - \gamma_{th,2}\epsilon_2\alpha_2\Big)\phi_2}{\gamma_{th,2}\Big(\epsilon_1+\epsilon_2\Big)\alpha_1 P\phi_1 + P\Big(\epsilon_1\alpha_2 - \gamma_{th,2}\epsilon_2\alpha_2\Big)\phi_2} \times \exp\left(-\frac{\gamma_{th,1}(\mu_r^2+\nu_r^2)N_o}{P\Big(\epsilon_1\alpha_1 - \gamma_{th,1}\epsilon_2\alpha_1\Big)\phi_1}\right). \quad (14)$$

B. Ergodic Capacity of the uplink NOMA with IQI

In this subsection, we evaluate the performance of uplink NOMA with the term of ergodic capacity (EC), considering different cases of in-phase and quadrature imbalance (IQI) imperfections.

1) NOMA TX impaired IQI

The EC of the U_2 and U_1 in the case of the impaired TX can be given, respectively as

$$C_2^{TX} = \frac{1}{ln2}\int_0^{\frac{\mu_t^2}{\nu_t^2}} \frac{\exp\left(-\frac{xN_o}{P(\mu_t^2\alpha_2 - x\nu_t^2\alpha_2)\phi_2}\right)}{1+x} \frac{P\Big(\mu_t^2\alpha_2 - x\nu_t^2\alpha_2\Big)\phi_2}{x(\mu_t^2+\nu_t^2)\alpha_1 P\phi_1 + P(\mu_t^2\alpha_2 - x\nu_t^2\alpha_2)\phi_2} dx, \quad (15)$$

$$C_1^{TX} = \frac{1}{ln2}\int_0^{\frac{\mu_t^2}{\nu_t^2}} \frac{\exp\left(-\frac{xN_o}{P(\mu_t^2\alpha_1 - x\nu_t^2\alpha_1)\phi_1}\right)}{1+x} dx. \quad (16)$$

2) NOMA RX impaired IQI

Assuming an ideal TX, the EC of the U_2 and U_1 while the RX is impaired can be expressed, respectively as

$$C_2^{RX} = \frac{1}{ln2}\int_0^{\frac{\mu_r^2}{\nu_r^2}} \frac{\exp\left(-\frac{x(\mu_r^2+\nu_r^2)N_o}{P(\mu_r^2\alpha_2 - x\nu_r^2\alpha_2)\phi_2}\right)}{1+x} \frac{P\Big(\mu_r^2\alpha_2 - x\nu_r^2\alpha_2\Big)\phi_{2,S}}{x(\mu_r^2+\nu_r^2)\alpha_1 P\phi_1 + P(\mu_r^2\alpha_2 - x\nu_r^2\alpha_2)\phi_2} dx, \quad (17)$$

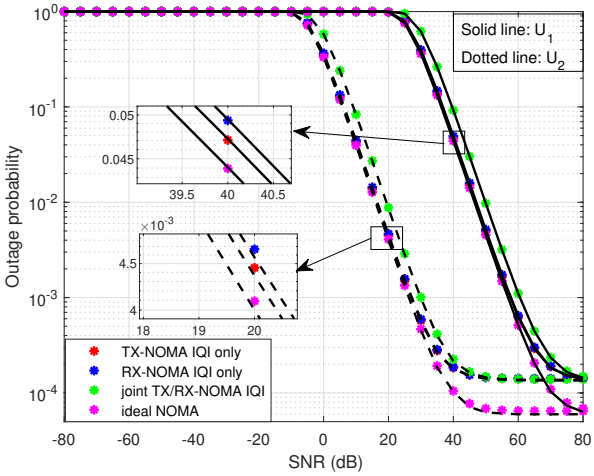


Fig. 2: OP of the uplink NOMA under IQI in function of SNR with $IRR_t = IRR_r = 20dB$.

$$C_1^{RX} = \frac{1}{ln2}\int_0^{\frac{\mu_r^2}{\nu_r^2}} \frac{\exp\left(-\frac{x(\mu_r^2+\nu_r^2)N_o}{P(\mu_r^2\alpha_1 - x\nu_r^2\alpha_1)\phi_1}\right)}{1+x} dx. \quad (18)$$

3) NOMA TX and RX impaired IQI

The EC of U_2 and U_1 when both the TX and RX are impaired can be obtained, respectively as

$$C_2^{TX/RX} = \frac{1}{ln2}\int_0^{\frac{\epsilon_1}{\epsilon_2}} \frac{\exp\left(-\frac{x(\mu_r^2+\nu_r^2)N_o}{P\Big(\epsilon_1\alpha_2 - x\epsilon_2\alpha_2\Big)\phi_2}\right)}{1+x} \frac{P\Big(\epsilon_1\alpha_2 - x\epsilon_2\alpha_2\Big)\phi_2}{x\Big((\epsilon_1+\epsilon_2)\alpha_1 P\phi_1 + P\Big(\epsilon_1\alpha_2 - x\epsilon_2\alpha_2\Big)\phi_2} dx, \quad (19)$$

$$C_1^{TX/RX} = \frac{1}{ln2}\int_0^{\frac{\epsilon_1}{\epsilon_2}} \frac{\exp\left(-\frac{x(\mu_r^2+\nu_r^2)N_o}{P\Big(\epsilon_1\alpha_1 - x\epsilon_2\alpha_1\Big)\phi_1}\right)}{1+x} dx, \quad (20)$$

IV. Numerical Results

To evaluate the performance of the uplink NOMA system, we consider the impact of IQI impairments on both the transmitter (TX) and receiver (RX), with parameters $\alpha_1 = 0.2$ and $\alpha_2 = 0.8$.

Figures 2 and 3 illustrate the influence of IQI on the uplink NOMA system as a function of SNR, assuming IRRt = IRRr = 20 dB, in terms of outage probability (OP) and ergodic capacity (EC). The results indicate that joint TX/RX IQI impairments significantly affect system performance. Additionally, U_2 outperforms $U1$ due to the uplink NOMA protocol allocating more power resources to the near user than the far user.

Figures 4 and 5 illustrate the impact of PA on uplink NOMA performance in terms of OP and EC across various SNR values. The effect of PA on user performance varies for

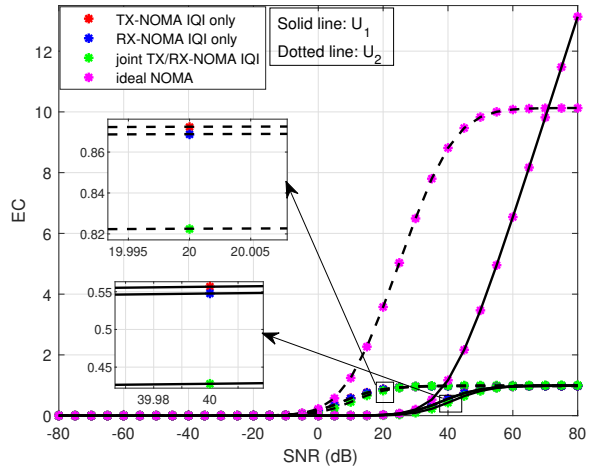


Fig. 3: EC of the uplink NOMA under IQI in function of SNR with $IRR_t = IRR_r = 20dB$.

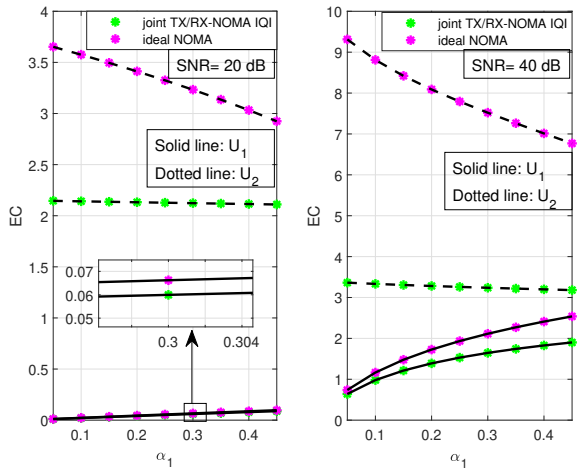


Fig. 5: EC of the uplink NOMA under IQI in function of PA with $IRR_t = IRR_r = 20dB$ and SNR= 20, 40 dB.

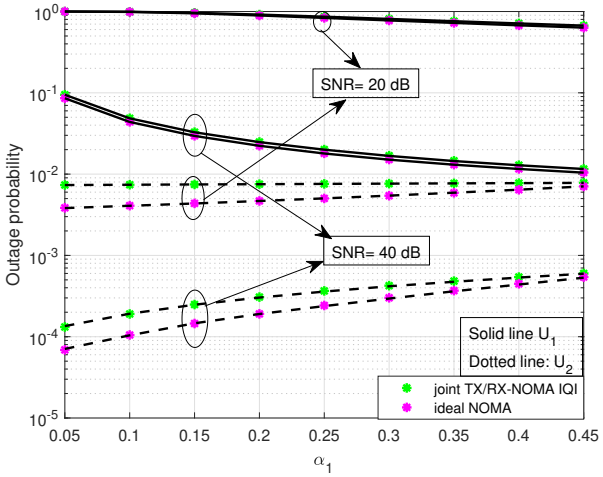


Fig. 4: OP of the uplink NOMA under IQI in function of PA with $IRR_t = IRR_r = 20dB$ and SNR= 20, 40 dB.

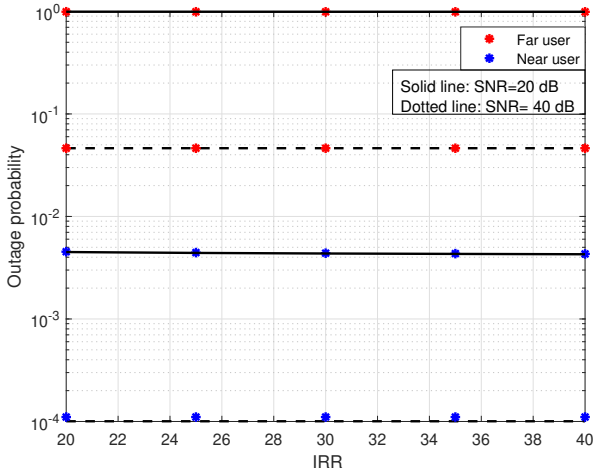


Fig. 6: OP of the uplink NOMA under IQI in the function of IRR with $IRR_t = IRR_r = 20dB$ and SNR= 20, 40 dB.

each user: increasing α_1 enhances the performance of $U1$ while causing a decline in the performance of $U2$.

Figures 6 and 7 demonstrate the impact of IRR on the uplink of direct NOMA at varying SNR levels. The results show that increasing IRR enhances the performance of both users. Moreover, despite existing imperfections, the system's performance continues to improve with higher SNR values.

V. Conclusion

This chapter examined the impact of IQI on the performance of an uplink NOMA system in terms of OP and EC over a Rayleigh fading channel. The analysis considered three scenarios: IQI impairment at the transmitter, at the receiver, and simultaneously at both the transmitter and receiver. The results demonstrated that IQI impairments significantly degrade the OP and EC performance of uplink NOMA systems, emphasizing the need for effective compensation and optimization

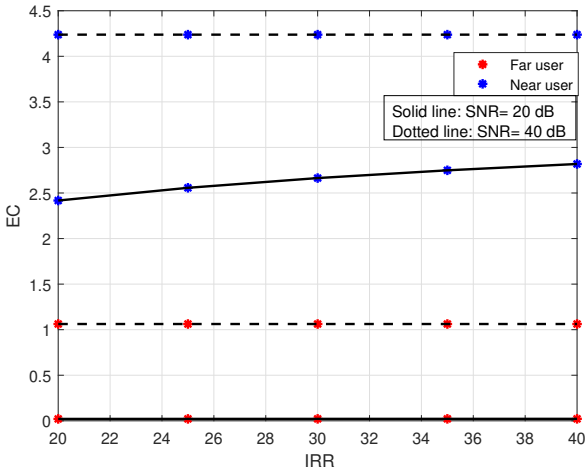


Fig. 7: OP of the uplink NOMA under IQI w.r.t of IRR with $IRR_t = IRR_r = 20dB$ and SNR= 20, 40 dB.

techniques to improve the overall performance and efficiency of NOMA systems.

REFERENCES

[1] Ö. Özdemir, R. Hamila, and N. Al-Dhahir, "Exact average OFDM subcarrier SINR analysis under joint transmit–receive I/Q imbalance," *IEEE Trans. Veh. Tech.*, vol. 63, no. 8, pp. 4125–4130, 2014.

[2] O. Ozdemir, R. Hamila, and N. Al-Dhahir, "I/Q imbalance in multiple beamforming {OFDM} transceivers: SINR analysis and digital baseband compensation," *IEEE transactions on communications*, vol. 61, no. 5, pp. 1914–1925, 2013.

[3] B. Selim, P. C. Sofotasios, S. Muhaidat, and G. K. Karagiannidis, "The effects of I/Q imbalance on wireless communications: A survey," in *2016 IEEE 59th International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, 2016, pp. 1–4.

[4] R. Hamila, Ö. Özdemir, and N. Al-Dhahir, "Beamforming OFDM performance under joint phase noise and I/Q imbalance," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2978–2989, 2015.

[5] M. Mokhtar, A. Gomaa, and N. Al-Dhahir, "OFDM AF relaying under I/Q imbalance: Performance analysis and baseband compensation," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1304–1313, 2013.

[6] B. Selim, S. Muhaidat, P. C. Sofotasios, A. Al-Dweik, B. S. Sharif, and T. Stouraitis, "Radio-frequency front-end impairments: Performance degradation in nonorthogonal multiple access communication systems," *IEEE Vehicular Technology Magazine*, vol. 14, no. 1, pp. 89–97, 2019.

[7] X. Tian, Q. Li, X. Li, H. Peng, C. Zhang, K. M. Rabie, and R. Kharel, "I/Q imbalance and imperfect SIC on two-way relay NOMA systems," *Electronics*, vol. 9, no. 2, p. 249, 2020.

[8] A.-A. A. Boulogeorgos, P. C. Sofotasios, B. Selim, S. Muhaidat, G. K. Karagiannidis, and M. Valkama, "Effects of RF impairments in communications over cascaded fading channels," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 11, pp. 8878–8894, 2016.

[9] B. Selim, S. Muhaidat, P. C. Sofotasios, B. S. Sharif, T. Stouraitis, G. K. Karagiannidis, and N. Al-Dhahir, "Outage probability of single carrier NOMA systems under I/Q imbalance," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2018, pp. 1–6.

[10] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE signal processing letters*, vol. 21, no. 12, pp. 1501–1505, 2014.

[11] X. Li, M. Zhao, C. Zhang, W. U. Khan, J. Wu, K. M. Rabie, and R. Kharel, "Security analysis of multi-antenna NOMA networks under I/Q imbalance," *Electronics*, vol. 8, no. 11, p. 1327, 2019.

[12] X. Li, M. Zhao, Y. Liu, L. Li, Z. Ding, and A. Nallanathan, "Secrecy analysis of ambient backscatter NOMA systems under I/Q imbalance," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 12 286–12 290, 2020.

[13] M. M. Alsmadi, N. A. Ali, M. Hayajneh, and S. S. Ikki, "Down-link NOMA networks in the presence of IQI and imperfect SIC: Receiver design and performance analysis," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6793–6797, 2020.

[14] S. A. Mohajeran and S. M. S. Sadough, "Outage performance analysis of secondary user under Tx/Rx IQI in cognitive radio systems," in *2015 5th International Conference on Computer and Knowledge Engineering (ICCKE)*. IEEE, 2015, pp. 246–250.

[15] A.-A. A. Boulogeorgos, P. C. Sofotasios, S. Muhaidat, M. Valkama, and G. K. Karagiannidis, "The effects of RF impairments in vehicle-to-vehicle communications."

Effect of Hardware Impairments on Cooperative NOMA with Direct Links

Beddiaf Safia[1], Abdellatif Khelil[1], Faical Khennoufa[1], and Beddiaf Walid[2]

[1]*LGEERE Laboratory, Department of Electrical Engineering, Echahid Hamma Lakhdar University, El-Oued, Algeria*
[2]*LEVE Laboratory, Department of Architecture, University Mouhamed Khidar, Biskra, Algeria*
Email: beddiaf-safia@univ-eloued.dz, abdellatif-khelil@univ-eloued.dz, khennoufa-faical@univ-eloued.dz, w.beddiaf@univ-biskra.dz.

***Abstract*—In this paper, we investigate the impact of the hardware impairment (HWI) on the non-orthogonal multiple access (NOMA) assisted cooperative fixed relay using Maximum Ratio Combining (MRC). To be practical, the imperfect channel state information (ICSI) is taken into account. The base station (BS) communicates with users via the direct transmission and the deployment of a fixed relay using MRC. To evaluate the performance of HWI on the cooperative relying NOMA in the presence of ICSI, the exact closed-form expressions of the bit error rate (BER) are derived. The analytical expressions are validated through computer simulations. The results indicate that the level of the HWI causes an error floor at the high signal-to-noise ratio (SNR), which degrade significantly the system's performance. The HWI and ICSI reduce the performance of the successive interference cancellation (SIC) process for each node.**

***Index Terms*—HWI, ICSI, cooperative, NOMA, BER.**

I. Introduction

Due to the increasing demand for mobile Internet and the Internet of Things (IoT), beyond fifth generation (5G) wireless communications must have the capacity to meet challenging requirements [1]. Non-orthogonal multiple access (NOMA) is a solution to achieve massive connectivity and high spectral efficiency [2]. The performance of the NOMA network in terms of capacity, ergodic sum rate, power allocation (PA) and outage probability has been studied extensively [3]. The exact closed-form expressions of bit error rate (BER) of the NOMA over fading channel under imperfect SIC (ipSIC) process were derived in [4]. Cooperative NOMA is an effective strategy to improve spectral efficiency and transmission reliability [5]. The NOMA-based cooperative system has been investigated in [6]–[8]. The closed-form expressions for the ergodic sum rate and outage probability were derived for the cooperative decode-and-forward (DF) relaying NOMA [6]. In [7], the end-to-end (e2e) bit error probability (BEP) expressions are obtained for DF relay aided NOMA when there is no direct communication between the users and base station (BS). The authors in [8] derived the exact e2e BEP of the cooperative NOMA using MRC to improve the performance of the far user with the help of the near user. The radio-frequency (RF) is one of the most significant components of the massive network of interconnected devices that facilitates communication between the individual devices or/ and their BS. Therefore, the direct-conversion transceivers seem to be the RF front-end solution to stringent design goals such as low cost, low power dissipation, improved efficiency and performance [9]. Although, in a realistic communication scenario, the RF component suffers from various types of hardware impairment (HWI) that limit the performance of the overall system, such as oscillator phase noise (PN), high power amplifier (HPA) and in-phase and quadrature-phase imbalance (IQI) [10], [11]. The effect of individual HWI on the system's performance has been examined. The performance of NOMA dual-hop (DH) AF relaying networks is evaluated in terms of outage probability and ergodic sum rate over the Nakagami-m fading channel in [12]. Another limitation of the previous research is the imperfect channel state information (ICSI). The related studies of the HWI with ICSI have been considered in [13], [14]. The ipSIC is unlikely to exist due to the HWI and channel fading coefficient [15]. Based on the above discussions, the impact of HWI on the BER of the cooperative NOMA system for enhancing the performance of the near and far users using an MRC has not been performed to the best of knowledge of the authors. In this paper, we analyze the impact of HWI on BER of the cooperative fixed relay NOMA in the presence of ICSI. The main contributions of the paper will be as follows:

- We present transceiver HWI on NOMA assisted cooperative relay scheme using the MRC under ICSI over the Rayleigh fading channel.
- We derive the exact e2e BER for each user in the HWI cooperative DF fixed relay NOMA using MRC under ICSI.
- We investigate the impact of the PA, channel estimation error, HWI factor and the distance of the relay for each user on the performance of the considered system. enditemize

II. System Model

We consider a downlink NOMA system based on cooperative DF fixed relay. The model consists of a base station (BS), DF relay and two users: user 1 (U1) and user 2 (U2). The BS, Relay and users are equipped with a single antenna. The complex flat fading channel coefficient between BS-users is represented as $\mathbf{h}_i \sim \mathcal{CN}(0,\ h_i^2)$ and between Relay-users as $\mathbf{g}_j \sim \mathcal{CN}(0,\ g_j^2)$, where $\{i\} \in [\{bs,1\},\{bs,2\},\{bs,r\}]$ and $\{j\} \in [\{r,1\},\{r,2\}]$. We assume an ICSI for each node. The estimation channel coefficient is given as $\tilde{h_i} = h_i - e$ and $\tilde{g_j} = h_i - e$,

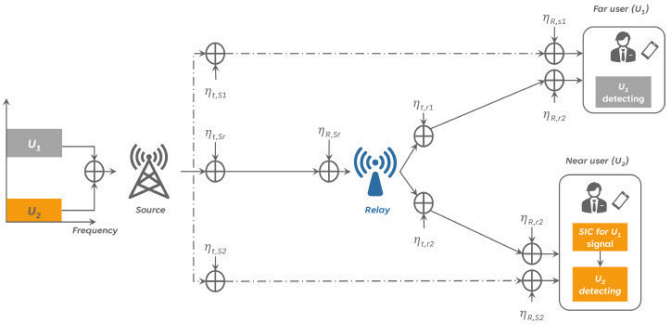


Fig. 1: System model of the cooperative NOMA in the presence of network.

where the channel estimation error is presented as $\mathbf{e} \sim \mathcal{CN}(0,\ \sigma_\epsilon^2)$. Since $\tilde{h_i}$ and e are independent, then it can be modeled as $\sigma^2_{\tilde{h_i}} = \sigma^2_{h_i} - \sigma^2_\epsilon$ and $\sigma^2_{\tilde{g_j}} = \sigma^2_{g_j} - \sigma^2_\epsilon$ [5], [16].

We use the BPSK modulation to encode the message of U1 and U2. As presented in Fig 1, the BS transmits a superimposed coding (SC) signal $s_{sc,bs} = \sqrt{\alpha_1}s_1 + \sqrt{\alpha_2}s_2$ with different power coefficients according to channel gain $|h_{bs,1}|^2 < |h_{bs,2}|^2$. The relay and users receive the BS's signal in the first phase, the U1 decodes its message directly by Maximum Likelihood Detection (MLD). The U2 and relay use the SIC to extract s_1 and s_2. The received signals at U1, U2 and Relay are given as [5]

$$y_i = (\tilde{h_i} + e)\sqrt{P_s}(s_{sc,bs} + \eta_{t,i}) + \eta_{R,i} + n_i \quad (1)$$

where, s_1, s_2 are the message of the U1 and 2, and $_1$, $_2$ are the PA coefficient of U1 and U2,respectively. $\alpha_1 + \alpha_2 = 1$, $s_{sc,bs}$ the superimposed coding signal. P_s is the BS transmits power, $\eta_{t,i}$ and $\eta_{R,i}$ are distortion noises in the transmitter and receiver, respectively. n_i is the Additive White Gaussian Noise (AWGN) $\mathbf{n}_i \sim \mathcal{CN}(0,\ \frac{N_o}{2})$.

$$\eta_{t,i} \sim \mathcal{CN}(0,\ k^2_{t,i}P_s)\ \eta_{R,i} \sim \mathcal{CN}(0,\ k^2_{R,i}P_s|h_i|^2) \quad (2)$$

where, $k^2_{t,i}$ and $k^2_{R,i}$ represent the level of impairment at the transmitter and receiver, respectively. Then, (1) can be regarded as

$$y_i = (\tilde{h_i} + e)\sqrt{P_s}(s_{sc,bs} + \eta_i) + n_i \quad (3)$$

where, η_i is the independent distortion noise defined as $\eta_i \sim \mathcal{CN}(0,\ k_i^2 P_s)$, k_i^2 is the HWI level at the transceiver, with $k_i^2 = \sqrt{k^2_{t,i} + k^2_{R,i}}$.
Subsequently, the relay creates a new SC signal and forwards it by the power of the relay to the users in the second phase. Again, the relay distributes the values of the power coefficient according to channel gain as in BS. We assume that the distribution of the PA of the messages at the relay with the same BS parameters. The received signal at U1 and U2 in the second phase is given as

$$y_i = (\tilde{g_j} + e)\sqrt{P_r}(s_{sc,bs} + \eta_j) + n_j \quad (4)$$

where P_r is the relay transmit power, n_j is the Additive White Gaussian Noise (AWGN) $\mathbf{n}_j \sim \mathcal{CN}(0,\ \frac{N_o}{2})$.
Hence, the users receive two signals, one from BS and the other from the relay. So, the users have implemented the MRC to combine the received signals from the two phases.

III. Performance Analysisl

In this section, under the Rayleigh fading, we express the e2e BER expressions of the HWI on the cooperative NOMA using MRC at U1 and U2 under the ICSI. The BPSK modulation is used to encode the message of U1 and U2.

A. BER at relay

In this subsection, we derive the exact closed-form expressions BER of the users' signal at the relay where the HWI and ICSI are taken into consideration.
1) BER of s1: We consider the ICSI at the relay that causes an error to detect the users' symbols [8] . The average BER of s1 at the relay can be expressed as [4, Eq.(12)]

$$P_{s_1,bs,r}\{s_{sc,bs} \in z\} = \frac{1}{2}\left(1 - \sqrt{\frac{\bar{\delta}^r_{s_1,bs,z}}{1+\bar{\delta}^r_{s_1,bs,z}}}\right), z = a, b. \quad (5)$$

where

$$\delta^r_{s_1,bs,a} = \frac{P_s(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2|\tilde{h}_{bs,r}|^2}{N_o+2P_sk^2_{bs,r}|\tilde{h}_{bs,r}|^2+2(k^2_{bs,r}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$$
$$\delta^r_{s_1,bs,b} = \frac{P_s(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2|\tilde{h}_{bs,r}|^2}{N_o+2P_sk^2_{bs,r}|\tilde{h}_{bs,r}|^2+2(k^2_{bs,r}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$$
$$\bar{\delta}^r_{s_1,bs,a} = E[\delta_{s_1,bs,a}],\ \bar{\delta}^r_{s_1,bs,b} = E[\delta_{s_1,bs,b}]$$

2) BER of s2: Using the SIC process and considering the ICSI, the relay detects the symbols of U2 by subtracting U1's symbols from the superimposed signal received. However, the ICSI affects the performance of the SIC process. The average BER of the s_2 can be defined as

$$P_{s_2,bs,r}\{s_{sc,bs} \in v\} = \frac{1}{2}\left(1 - \sqrt{\frac{\bar{\delta}^r_{s_2,bs,v}}{1+\bar{\delta}^r_{s_2,bs,v}}}\right), \quad v = a, b, c, d, e, f. \quad (6)$$

where

$$\delta^r_{s_2,bs,a} = \frac{P_s\alpha_2|\tilde{h}_{bs,r}|^2}{N_o+2P_sk^2_{bs,r}|\tilde{h}_{bs,r}|^2+2(k^2_{bs,r}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$$
$$\delta^r_{s_2,bs,b} = \frac{P_s\alpha_2|\tilde{h}_{bs,r}|^2}{N_o+2P_sk^2_{bs,r}|\tilde{h}_{bs,r}|^2+2(k^2_{bs,r}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$$
$$\delta^r_{s_2,bs,c} = \frac{P_s(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2|\tilde{h}_{bs,r}|^2}{N_o+2P_sk^2_{bs,r}|\tilde{h}_{bs,r}|^2+2(k^2_{bs,r}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$$
$$\delta^r_{s_2,bs,d} = \frac{P_s(2\sqrt{\alpha_1}+\sqrt{\alpha_2})^2|\tilde{h}_{bs,r}|^2}{N_o+2P_sk^2_{bs,r}|\tilde{h}_{bs,r}|^2+2(k^2_{bs,r}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$$
$$\delta^r_{s_2,bs,e} = \frac{P_s(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2|\tilde{h}_{bs,r}|^2}{N_o+2P_sk^2_{bs,r}|\tilde{h}_{bs,r}|^2+2(k^2_{bs,r}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$$

$$P_{u_1,coop-mrc}\{s_{sc,bs} \in z\} = \frac{1}{2}\left(1 - \frac{1}{\bar{\delta}_{s_1,bs,z} - \bar{\delta}_{s_1,r,z}}\left(\bar{\delta}_{s_1,bs,z}\sqrt{\frac{\bar{\delta}_{s_1,bs,z}}{1+\bar{\delta}_{s_1,bs,z}}} - \bar{\delta}_{s_1,r,z}\sqrt{\frac{\bar{\delta}_{s_1,r,z}}{1+\bar{\delta}_{s_1,r,z}}}\right)\right) \quad (7)$$

$$P_{u_2,coop-mrc}\{s_{sc,bs} \in v\} = \frac{1}{2}\left(1 - \frac{1}{\bar{\delta}_{s_2,bs,v} - \bar{\delta}_{s_2,r,v}}\left(\bar{\delta}_{s_2,bs,v}\sqrt{\frac{\bar{\delta}_{s_2,bs,v}}{1+\bar{\delta}_{s_2,bs,v}}} - \bar{\delta}_{s_2,r,v}\sqrt{\frac{\bar{\delta}_{s_2,r,v}}{1+\bar{\delta}_{s_2,r,v}}}\right)\right) \quad (8)$$

$\delta^r_{s_2,bs,f} = \frac{P_s(2\sqrt{\alpha_1}-\sqrt{\alpha_2})^2|\tilde{h}_{bs,r}|^2}{N_o+2P_sk^2_{bs,r}|\tilde{h}_{bs,r}|^2+2(k^2_{bs,r}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$
$\bar{\delta}^r_{s_2,bs,a} = E[\delta^r_{s_2,bs,a}]$, $\bar{\delta}^r_{s_2,bs,b} = E[\delta^r_{s_2,bs,b}]$, $\bar{\delta}^r_{s_2,bs,c} = E[\delta^r_{s_2,bs,c}]$, $\bar{\delta}^r_{s_2,bs,d} = E[\delta^r_{s_2,bs,d}]$, $\bar{\delta}^r_{s_2,bs,e} = E[\delta^r_{s_2,bs,e}]$, $\bar{\delta}^r_{s_2,bs,f} = E[\delta^r_{s_2,bs,f}]$.

B. BER of cooperative MRC

1) At U1: Based on the detected signals in the first phase, the relay creates a new SC signal again and sends it to U1 and U2 by its power. Thus, U1 and U2 received two signals from different sources, one from BS and one from the relay. The BER of the U1 with diversity combining of the two branches using MRC is given as in (7) (in the top of the next page)
where
$\delta_{s_1,bs,a} = \frac{P_s(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2|\tilde{h}_{bs,1}|^2}{N_o+2P_sk^2_{bs,1}|\tilde{h}_{bs,1}|^2+2(k^2_{bs,1}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$
$\delta_{s_1,bs,b} = \frac{P_s(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2|\tilde{h}_{bs,1}|^2}{N_o+2P_sk^2_{bs,1}|\tilde{h}_{bs,1}|^2+2(k^2_{bs,1}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$
$\bar{\delta}_{s_1,bs,a} = E[\delta_{s_1,bs,a}]$, $\bar{\delta}_{s_1,bs,b} = E[\delta_{s_1,bs,b}]$
$\delta_{s_1,r,a} = \frac{P_r(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2|\tilde{g}_{bs,1}|^2}{N_o+2P_rk^2_{bs,1}|\tilde{g}_{bs,1}|^2+2(k^2_{bs,1}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_r\sigma^2_\epsilon}$
$\delta_{s_1,r,b} = \frac{P_r(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2|\tilde{g}_{bs,1}|^2}{N_o+2P_rk^2_{bs,1}|\tilde{g}_{bs,1}|^2+2(k^2_{bs,1}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_r\sigma^2_\epsilon}$
$\bar{\delta}_{s_1,r,a} = E[\delta_{s_1,r,a}]$, $\bar{\delta}_{s_1,r,b} = E[\delta_{s_1,r,b}]$
2) At U2: The U2 detects s_1 firstly after combined the received signals with MRC. Thereafter, using the SIC detects its symbols s_2. The ABER for cooperative MRC is expressed as in (8) (in the top of the next page).
where
$\delta_{s_2,bs,a} = \frac{P_s\alpha_2|\tilde{h}_{bs,2}|^2}{N_o+2P_sk^2_{bs,2}|\tilde{h}_{bs,2}|^2+2(k^2_{bs,2}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$
$\delta_{s_2,bs,b} = \frac{P_s\alpha_2|\tilde{h}_{bs,2}|^2}{N_o+2P_sk^2_{bs,2}|\tilde{h}_{bs,2}|^2+2(k^2_{bs,2}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$
$\delta_{s_2,bs,c} = \frac{P_s(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2|\tilde{h}_{bs,2}|^2}{N_o+2P_sk^2_{bs,2}|\tilde{h}_{bs,2}|^2+2(k^2_{bs,2}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$
$\delta_{s_2,bs,d} = \frac{P_s(2\sqrt{\alpha_1}+\sqrt{\alpha_2})^2|\tilde{h}_{bs,2}|^2}{N_o+2P_sk^2_{bs,2}|\tilde{h}_{bs,2}|^2+2(k^2_{bs,2}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$
$\delta_{s_2,bs,e} = \frac{P_s(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2|\tilde{h}_{bs,2}|^2}{N_o+2P_sk^2_{bs,2}|\tilde{h}_{bs,2}|^2+2(k^2_{bs,2}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$
$\delta_{s_2,bs,f} = \frac{P_s(2\sqrt{\alpha_1}-\sqrt{\alpha_2})^2|\tilde{h}_{bs,2}|^2}{N_o+2P_sk^2_{bs,2}|\tilde{h}_{bs,2}|^2+2(k^2_{bs,2}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$
$\bar{\delta}_{s_2,bs,a} = E[\delta_{s_2,bs,a}]$, $\bar{\delta}_{s_2,bs,b} = E[\delta_{s_2,bs,b}]$,
$\bar{\delta}_{s_2,bs,c} = E[\delta_{s_2,bs,c}]$, $\bar{\delta}_{s_2,bs,d} = E[\delta_{s_2,bs,d}]$,
$\bar{\delta}_{s_2,bs,e} = E[\delta_{s_2,bs,e}]$, $\bar{\delta}_{s_2,bs,f} = E[\delta_{s_2,bs,f}]$.
$\delta_{s_2,r,a} = \frac{P_r\alpha_2|\tilde{g}_{r,2}|^2}{N_o+2P_rk^2_{r,2}|\tilde{g}_{r,2}|^2+2(k^2_{r,2}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_r\sigma^2_\epsilon}$
$\delta_{s_2,r,b} = \frac{P_r\alpha_2|\tilde{g}_{r,2}|^2}{N_o+2P_rk^2_{r,2}|\tilde{g}_{r,2}|^2+2(k^2_{r,2}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_r\sigma^2_\epsilon}$
$\delta_{s_2,r,c} = \frac{P_r(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2|\tilde{g}_{r,2}|^2}{N_o+2P_rk^2_{r,2}|\tilde{g}_{r,2}|^2+2(k^2_{r,2}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_r\sigma^2_\epsilon}$
$\delta_{s_2,r,d} = \frac{P_r(2\sqrt{\alpha_1}+\sqrt{\alpha_2})^2|\tilde{g}_{r,2}|^2}{N_o+2P_rk^2_{r,2}|\tilde{g}_{r,2}|^2+2(k^2_{r,2}+(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2)P_r\sigma^2_\epsilon}$
$\delta_{s_2,r,e} = \frac{P_r(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2|\tilde{g}_{r,2}|^2}{N_o+2P_rk^2_{r,2}|\tilde{g}_{r,2}|^2+2(k^2_{r,2}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_r\sigma^2_\epsilon}$
$\delta_{s_2,r,f} = \frac{P_r(2\sqrt{\alpha_1}-\sqrt{\alpha_2})^2|\tilde{g}_{r,2}|^2}{N_o+2P_rk^2_{r,2}|\tilde{g}_{r,2}|^2+2(k^2_{r,2}+(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2)P_s\sigma^2_\epsilon}$
$\bar{\delta}_{s_2,r,a} = E[\delta_{s_2,r,a}]$, $\bar{\delta}_{s_2,r,b} = E[\delta_{s_2,r,b}]$, $\bar{\delta}_{s_2,r,c} = E[\delta_{s_2,r,c}]$, $\bar{\delta}_{s_2,r,d} = E[\delta_{s_2,r,d}]$, $\bar{\delta}_{s_2,r,e} = E[\delta_{s_2,r,e}]$, $\bar{\delta}_{s_2,r,f} = E[\delta_{s_2,r,f}]$.

C. BER of cooperative MRC in the error propagation

After the implantation of the MRC, we should describe the total received signal in order to acquire the BER in case of error propagation from U1 and U2.
1) At U1 : Consequently, the sum of the received signals at the U1 by MRC [17], [18] is given as

$$\begin{aligned}\varphi_{u1,err-coop} &= \tilde{h}^*_{bs,1}y_1 + \tilde{g}^*_{r,1}y_{r1} \\ &= (\varphi_{s_1,bs,z} - \varphi_{s_1,r,z}) + \tilde{n}_a\end{aligned} \quad (9)$$

where $\varphi_{s_1,bs,a} = P_s(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2|\tilde{h}_{bs,1}|^2$ and $\varphi_{s_1,bs,b} = P_s(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2|\tilde{h}_{bs,1}|^2$ and we define $\varphi_{s_1,r,a} = P_r(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2|\tilde{g}_{r,1}|^2$ and $\varphi_{s_1,r,b} = P_r(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2|\tilde{g}_{r,1}|^2$. However, $\tilde{n}_a$ is the effective noise with $\mathbf{E}[\tilde{\mathbf{n}}_a] \sim \mathcal{CN}(0, \frac{N_o}{2}[\sigma^2_{\tilde{h}_{bs,1}} + \sigma^2_{\tilde{g}_{r,1}}] + k^2_{bs,1}P_s\sigma_{\tilde{h}^2_{bs,1}}\sigma^2_\epsilon + k^2_{bs,1}P_s\sigma^2_{\tilde{h}_{bs,1}} + k^2_{r,1}P_r\sigma^2_{\tilde{g}_{r,1}}{\sigma_\epsilon}^2 + k^2_{r,1}P_r\sigma^2_{\tilde{g}_{r,1}} + \tau_{s_1,bs,z}\sigma^2_{\tilde{h}_{bs,1}}\sigma^2_\epsilon + \tau_{s_1,r,z}\sigma^2_{\tilde{g}_{r,1}}\sigma^2_\epsilon)$, where $\tau_{s_1,bs,a} = P_s(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2$, $\tau_{s_1,bs,b} = P_s(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2$ and $\tau_{s_1,r,a} = P_r(\sqrt{\alpha_1}+\sqrt{\alpha_2})^2$, $\tau_{s_1,r,b} = P_r(\sqrt{\alpha_1}-\sqrt{\alpha_2})^2$. Through the MLD decision rule at the U1 $s_1 = +1$ is declared if $\varphi_{u1,err-coop} \geq 0$. Thus, the BER of the error propagation for the U1 symbols is defined as

$$\begin{aligned}P_{u1,prop-mrc}\{s_{sc,bs} \in z\} &= P(\varphi_{s_1,bs,z} - \varphi_{s_1,r,z} < \tilde{n}_a) \\ &= Q\left(\frac{\varphi_{s_1,bs,z} - \varphi_{s_1,r,z}}{\sqrt{\omega_a}}\right)\end{aligned} \quad (10)$$

where $\omega_a = \frac{N_o}{2}[\sigma^2_{\tilde{h}_{bs,1}} + \sigma^2_{\tilde{g}_{r,1}}] + k^2_{bs,1}P_s\sigma_{\tilde{h}^2_{bs,1}}\sigma^2_\epsilon + k^2_{bs,1}P_s\sigma^2_{\tilde{h}_{bs,1}} + k^2_{r,1}P_r\sigma^2_{\tilde{g}_{r,1}}{\sigma_\epsilon}^2 + k^2_{r,1}P_r\sigma^2_{\tilde{g}_{r,1}} + \tau_{s_1,bs,z}\sigma^2_{\tilde{h}_{bs,1}}\sigma^2_\epsilon + \tau_{s_1,r,z}\sigma^2_{\tilde{g}_{r,1}}\sigma^2_\epsilon$. From (10), if the relay forwards an incorrect symbol, this has a significant impact instead of noise on the decision variable MLD. Accordingly to [17], [18] , we approximate the BER as $\varphi_{s_1,bs,z} - \varphi_{s_1,r,z} < 0$. the BER under error propagation is given as

$$P_{u1,prop-mrc}\{s_{sc,bs} \in z\} = P(\varphi_{s_1,bs,z} - \varphi_{s_1,r,z} < 0) \quad (11)$$

$$P_{u1,prop-mrc} = \int_{\varphi_{s_1,bs,z}}^{\infty} \int_{0}^{\varphi_{s_1,r,z}} \frac{1}{\bar{\varphi}_{s_1,bs,z}\bar{\varphi}_{s_1,r,z}} e^{\frac{-\varphi_{s_1,bs,z}}{\bar{\varphi}_{s_1,bs,z}}} e^{\frac{-\varphi_{s_1,r,z}}{\bar{\varphi}_{s_1,r,z}}} d\varphi_{s_1,bs,z} d\varphi_{s_1,r,z} = \frac{\bar{\varphi}_{s_1,r,z}}{\bar{\varphi}_{s_1,r,z} + \bar{\varphi}_{s_1,bs,z}} \tag{12}$$

By averaging (11) over $\varphi_{s_1,bs,z}$and $\varphi_{s_1,r,z}$ the average error probability under error propagation is obtained as in (12) (in the top of the next page).

2) At U2: Also, after combining the BS and relay signals at the U2, the total received signal by U2 is given as

$$\varphi_{u2,err-coop} = \tilde{h}^*_{bs,2} y_2 + \tilde{g}^*_{r,2} y_{r2} \tag{13}$$

At the U2, after the implantation of the MRC, it detects the U1 symbols firstly then detects its symbols through the SIC process. Therefore, the sum of the received signal at U2 can be indicated as

$$\varphi_{u2,err-coop} = (\varphi_{s_2,bs,v} - \varphi_{s_2,r,v}) + \tilde{n}_b \tag{14}$$

where $\varphi_{s_2,bs,a} = P_s\alpha_2|\tilde{h}_{bs,2}|^2$, $\varphi_{s_2,bs,b} = P_s\alpha_2|\tilde{h}_{bs,2}|^2$, $\varphi_{s_2,bs,c} = P_s(\sqrt{\alpha_1} + \sqrt{\alpha_2})^2|\tilde{h}_{bs,2}|^2$, $\varphi_{s_2,bs,d} = P_s(2\sqrt{\alpha_1} + \sqrt{\alpha_2})^2|\tilde{h}_{bs,2}|^2$, $\varphi_{s_2,bs,e} = P_s(\sqrt{\alpha_1} - \sqrt{\alpha_2})^2|\tilde{h}_{bs,2}|^2$, $\varphi_{s_2,bs,f} = P_s(2\sqrt{\alpha_1} - \sqrt{\alpha_2})^2|\tilde{h}_{bs,2}|^2$ and $\varphi_{s_2,r,a} = P_r\alpha_2|\tilde{g}_{r,2}|^2$, $\varphi_{s_2,r,b} = P_r\alpha_2|\tilde{g}_{r,2}|^2$, $\varphi_{s_2,r,c} = P_r(\sqrt{\alpha_1} + \sqrt{\alpha_2})^2|\tilde{g}_{r,2}|^2$, $\varphi_{s_2,r,d} = P_r(2\sqrt{\alpha_1} + \sqrt{\alpha_2})^2|\tilde{g}_{r,2}|^2$, $\varphi_{s_2,r,e} = P_r(\sqrt{\alpha_1} - \sqrt{\alpha_2})^2|\tilde{g}_{r,2}|^2$, $\varphi_{s_2,r,f} = P_r(2\sqrt{\alpha_1} - \sqrt{\alpha_2})^2|\tilde{g}_{r,2}|^2$. $\tilde{n}_b$ is the effective noise with $\mathbf{E}[\tilde{\mathbf{n}}_b] \sim \mathcal{CN}(0, \frac{N_o}{2}[\sigma^2_{\tilde{h}_{bs,2}} + \sigma^2_{\tilde{g}_{r,2}}] + k^2_{bs,2}P_s\sigma_{\tilde{h}^2_{bs,2}}\sigma^2_\epsilon + k^2_{bs,2}P_s\sigma^2_{\tilde{h}_{bs,2}} + k^2_{r,2}P_r\sigma^2_{\tilde{g}_{r,2}}{\sigma_\epsilon}^2 + k^2_{r,2}P_r\sigma^2_{\tilde{g}_{r,2}} + \tau_{s_2,bs,v}\sigma^2_{\tilde{h}_{bs,2}}\sigma^2_\epsilon + \tau_{s_2,r,v}\sigma^2_{\tilde{g}_{r,2}}\sigma^2_\epsilon)$, where $\tau_{s_2,bs,a} = P_s\alpha_2$, $\tau_{s_2,bs,b} = P_s\alpha_2$, $\tau_{s_2,bs,c} = P_s(\sqrt{\alpha_1} + \sqrt{\alpha_2})^2$, $\tau_{s_2,bs,d} = P_s(2\sqrt{\alpha_1} + \sqrt{\alpha_2})^2$, $\tau_{s_2,bs,e} = P_s(\sqrt{\alpha_1} - \sqrt{\alpha_2})^2$, $\tau_{s_2,bs,f} = P_s(2\sqrt{\alpha_1} - \sqrt{\alpha_2})^2$ and $\tau_{s_2,r,a} = P_r\alpha_2$, $\tau_{s_2,r,b} = P_r\alpha_2$, $\tau_{s_2,r,c} = P_r(\sqrt{\alpha_1} + \sqrt{\alpha_2})^2$, $\tau_{s_2,r,d} = P_r(2\sqrt{\alpha_1} + \sqrt{\alpha_2})^2$, $\tau_{s_2,r,e} = P_r(\sqrt{\alpha_1} - \sqrt{\alpha_2})^2$, $\tau_{s_2,r,f} = P_r(2\sqrt{\alpha_1} - \sqrt{\alpha_2})^2$. Through the MLD decision rule and the SIC process at the U2 are declared as $s_1 = +1$ and $s_2 = +1$. Thus, the BER of the error propagation for the U2 symbols is defined as

$$P_{u2,prop-mrc}\{s_{sc,bs} \in v\} = P(\varphi_{s_2,bs,v} - \varphi_{s_1,r,z} < \tilde{n}_b) = Q\left(\frac{\varphi_{s_2,bs,v} - \varphi_{s_2,r,v}}{\sqrt{\omega_b}}\right) \tag{15}$$

where $\omega_b = \frac{N_o}{2}[\sigma^2_{\tilde{h}_{bs,2}} + \sigma^2_{\tilde{g}_{r,2}}] + k^2_{bs,2}P_s\sigma_{\tilde{h}^2_{bs,2}}\sigma^2_\epsilon + k^2_{bs,2}P_s\sigma^2_{\tilde{h}_{bs,2}} + k^2_{r,2}P_r\sigma^2_{\tilde{g}_{r,2}}{\sigma_\epsilon}^2 + k^2_{r,2}P_r\sigma^2_{\tilde{g}_{r,2}} + \tau_{s_2,bs,v}\sigma^2_{\tilde{h}_{bs,2}}\sigma^2_\epsilon + \tau_{s_2,r,v}\sigma^2_{\tilde{g}_{r,2}}\sigma^2_\epsilon$ From (15), if the relay forwards an incorrect symbol, this has a significant impact instead of noise on the decision variable MLD and SIC. Accordingly, we approximate the BER for the SIC to detect s_2 as$\varphi_{s_2,bs,v} - \varphi_{s_2,r,v} < 0$ the BER under error propagation is presented as [17], [18].

$$P_{u2,prop-mrc}\{s_{sc,bs} \in v\} = P(\varphi_{s_2,bs,v} - \varphi_{s_2,r,v} < 0) \tag{16}$$

By averaging (16) over $\varphi_{s_2,bs,v}$ and $\varphi_{s_2,r,v}$, the average error probability under error propagation is obtained as in (17) (in the top of the page).

D. E2e BER of cooperative MRC

The e2e BER of HWI of the cooperative NOMA with MRC for U1 [17], [18] is given by

$$P^{avg}_{u1,e2e} = \frac{1}{2}[P_{u1,e2e,a} + P_{u1,e2e,b}] \tag{18}$$

where $P_{u1,e2e,a} = P_{u1,prop-mrc,a} \times P_{s1,bs,r,a} + (1 - P_{s1,bs,r,a})P_{u2,prop-mrc,a}$ and $P_{u1,e2e,b} = P_{u2,prop-mrc,b} \times P_{s1,bs,r,b} + (1 - P_{s1,bs,r,b})P_{u1,prop-mrc,b}$ Also, the e2e BER of HWI of the cooperative NOMA with MRC for U2 is given by

$$P^{avg}_{u2,e2e} = \frac{1}{2}[P_{u2,e2e,a} + P_{u2,e2e,b} - P_{u2,e2e,c} + P_{u2,e2e,d} + P_{u2,e2e,e} - P_{u2,e2e,f}] \tag{19}$$

where $P_{u2,e2e,a} = P_{u2,prop-mrc,a} \times P_{s2,bs,r,a} + (1 - P_{s2,bs,r,a}P_{u2,prop-mrc,a}$ and $P_{u2,e2e,b} = P_{u2,prop-mrc,b} \times P_{s2,bs,r,b} + (1 - P_{s1,bs,r,b}P_{u2,prop-mrc,b}$ $P_{u2,e2e,c} = P_{u2,prop-mrc,c} \times P_{s2,bs,r,c} + (1 - P_{s1,bs,r,c}P_{u2,prop-mrc,c}$ $P_{u2,e2e,d} = P_{u2,prop-mrc,d} \times P_{s2,bs,r,d} + (1 - P_{s1,bs,r,d}P_{u2,prop-mrc,d}$ $P_{u2,e2e,e} = P_{u2,prop-mrc,e} \times P_{s2,bs,r,e} + (1 - P_{s1,bs,r,e}P_{u2,prop-mrc,e}$ $P_{u2,e2e,f} = P_{u2,prop-mrc,f} \times P_{s2,bs,r,f} + (1 - P_{s1,bs,r,f}P_{u2,prop-mrc,f}$ By substituting for (5), (7) and (12) in (18), then substituting for (6), (8) and (17) in (19). We find the exact e2e BER for the U1 and U2, respectively. Assume the perfect SIC case at the relay: the relay forwards only the correct symbols to the users, i.e., when the relay detects the symbols incorrectly, the relay does not forward these symbols to the users, this means the diversity combining does not occur at the users, the users will receive only the BS signal. Hence, we assume that when the relay detects and forwards the correct symbols, the relay power is equal to P_r and the diversity combining occurs. If the relay detects the incorrect symbols, the relay does not forward the symbols to the users, i.e., relay power is equal to zero $P_r = 0$ and the diversity combining does not occur [19, pp 275-276] .

$$P_{u2,prop-mrc} = \int_{\varphi_{s_2,bs,v}}^{\infty} \int_{0}^{\varphi_{s_2,r,v}} \frac{1}{\bar{\varphi}_{s_2,bs,v}\bar{\varphi}_{s_2,r,v}} e^{\frac{-\varphi_{s_2,bs,v}}{\bar{\varphi}_{s_2,bs,v}}} e^{\frac{-\varphi_{s_2,r,v}}{\bar{\varphi}_{s_2,r,v}}} d\varphi_{s_2,bs,v} d\varphi_{s_2,r,z}$$
$$= \frac{\bar{\varphi}_{s_2,r,v}}{\bar{\varphi}_{s_2,r,v} + \bar{\varphi}_{s_2,bs,v}} \tag{17}$$

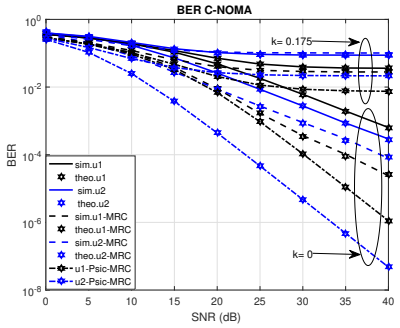
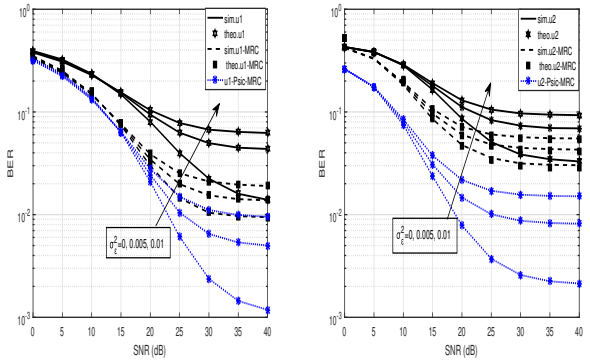


Fig. 2: BER of effect of HWI on the NOMA system with and without cooperative MRC.



(a) U1 (b) U2

Fig. 3: BER of U1 and U2 with different level of σ_ϵ^2.

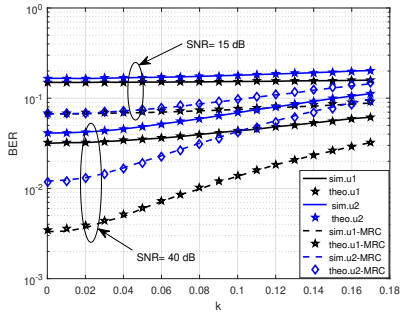
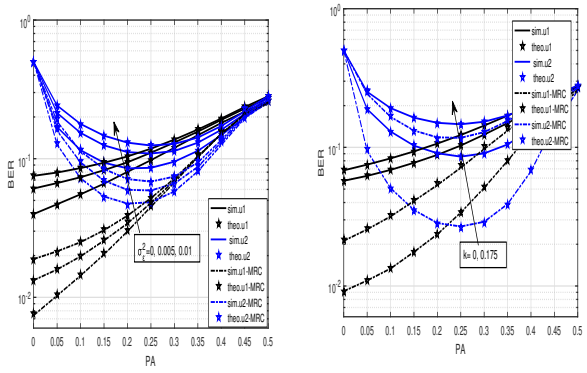


Fig. 4: BER in function of HWI factor with different level of SNR.

IV. Numerical results

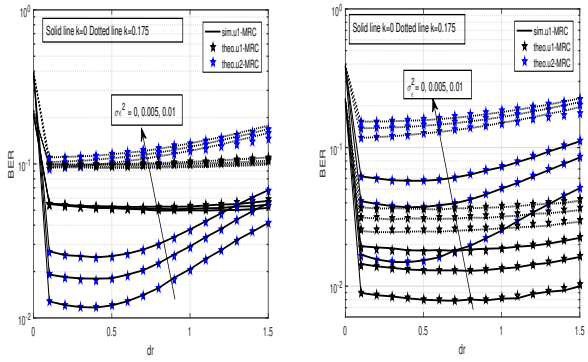
In this section, the performance of the HWI on the cooperative NOMA fixed relay using MRC is evaluated in terms of BER. The simulation results are obtained to validate the accuracy of the analytical derivations. We set the parameters to $d_{bs,1} = 4m$,$d_{bs,2} = 2m$,$d_{bs,r} = 1m$,$d_{r,1} = 3m$,$d_{r,2} = 1m$, transmit power of the relay $P_r = P_s$, $a = 2$, $\alpha_1 = 0.8$, $\alpha_2 = 0.2$ and HWI is defined as [30] $k_{bs,1} = k_{bs,2} = k_{r,1} = k_{r,2} = k$. Fig 2 examines the effects of the HWI on the NOMA system as a function of the SNR for two scenarios without and with cooperative NOMA using MRC. Perfect hardware is provided for comparison. The HWI factors are set at $k = 0$ and $k = 0.175$ representing the best and the worst value of HWI [38]. With $\sigma_\epsilon^2 = 0$, the numerical results of U1 and U2 in conventional NOMA through (6), (8) and (22), (23) for the cooperative NOMA using MRC are obtained, respectively. It is observed that the numerical results match perfectly with the simulation results. The presence of the HWI decreases the performance of both U1 and U2. The U1 and U2 with cooperative NOMA employing MRC achieve better performance than conventional NOMA when both of the system is affected by different levels of HWI ($k = 0$ and $k = 0.175$). Further, the perfect SIC case reaches maximal performance. Fig 3 illustrates the BER performance of the far and the near users with different levels of channel estimation error and $k = 0.1$ versus SNR. We can observe that increasing channel estimation error decreases the performance of both users, which means that the system's performance depends clearly on channel estimation error. It is also noticed that at the high SNR, the presence of the HWI and ICSI cause an error floor. Fig 4 presents the effect of HWI on the system in terms of BER with different levels of SNR versus HWI factor. It is seen that the BER increases continuously as the levels of HWI increase. Thus, the BER performance of the conventional and cooperative NOMA degrade as the level of HWI increase. It means that the hardware imperfection causes significant degradation of the BER performance. Fig 5 shows the BER of PA's effect on U1 and U2 performances, respectively, with different levels of HWI and ICSI. It is observed that the PA affects the BER performance of one user at the expense of the other. In addition to the channel estimate error and HWI degrading the SIC process' performance. increasing the PA factor makes also the system more susceptible to the SIC process. Fig 6 shows the impact of the distance of the relay on NOMA's performance system with different levels of HWI and ICSI. It is observed that when the relay is placed closer to the BS, the near user achieves better performance. Increasing the distance

FT/Boumerdes/NCASEE-24-Conference



(a) U1 (b) U2

Fig. 5: BER in function of PA with value of HWI and σ_ϵ^2 .



(a) $\alpha_1 = 0.7$ (b) $\alpha_2 = 0.9$

Fig. 6: BER of the HWI on C-NOMA versus d_r with different scenarios.

between the BS and the relay decreases the near user's performance. To be more specific, we present scenario 2 with $\alpha_1 = 0.9$ and $\alpha_2 = 0.1$, we can see that the far user's performance improves compared to scenario 1, which means that the Power allocation has a crucial role in the performance of the users.

V. CONCLUSION

In this paper, the performance of the HWI of the cooperative fixed relay NOMA system using MRC in the presence of ICSI over the Rayleigh fading channel is analyzed. The BS communicates with users through the direct and indirect channels where the existence of the HWI for each node is considered. The exact closed-form expressions of the BER are derived and validated by the simulation results. However, the impact of the HWI, channel estimation error, power allocation and the distance of the relay are examined. Based on the results, it can be shown that the HWI has a negative impact on the system's performance. The presence of the HWI and ICSI decrease the performance of the SIC process leading to severe degradation.

REFERENCES

[1] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5g and beyond," *Proceedings of the IEEE*, 2017.

[2] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5g: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, 2015.

[3] Q. Sun, S. Han, I. Chin-Lin, and Z. Pan, "On the ergodic capacity of mimo noma systems," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 405–408, 2015.

[4] F. Kara and H. Kaya, "Ber performances of downlink and uplink noma in the presence of sic errors over fading channels," *IET Communications*, vol. 12, no. 15, pp. 1834–1844, 2018.

[5] X. Li, J. Li, P. T. Mathiopoulos, D. Zhang, L. Li, and J. Jin, "Joint impact of hardware impairments and imperfect csi on cooperative swipt noma multi-relaying systems," in *2018 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE, 2018, pp. 95–99.

[6] M. F. Kader and S. Y. Shin, "Cooperative relaying using space-time block coded non-orthogonal multiple access," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 5894–5903, 2016.

[7] F. Kara and H. Kaya, "The error performance analysis of the decode-forward relay-aided-noma systems and a power allocation scheme for user fairness," *J. Fac. Eng. Archit. Gazi Univ*, vol. 35, no. 1, pp. 97–108, 2020.

[8] ——, "On the error performance of cooperative-noma with statistical csit," *IEEE Communications Letters*, vol. 23, no. 1, pp. 128–131, 2018.

[9] B. Selim, S. Muhaidat, P. C. Sofotasios, B. S. Sharif, T. Stouraitis, G. K. Karagiannidis, and N. Al-Dhahir, "Performance analysis of non-orthogonal multiple access under i/q imbalance," *IEEE Access*, vol. 6, pp. 18 453–18 468, 2018.

[10] T. Schenk, *RF imperfections in high-rate wireless systems: impact and digital compensation*. Springer Science & Business Media, 2008.

[11] L. Smaini, *Rf analog impairments modeling for communication systems simulation: application to OFDM-based transceivers*. John Wiley & Sons, 2012.

[12] C. Deng, X. Zhao, D. Zhang, X. Li, J. Li, and C. C. Cavalcante, "Performance analysis of noma-based relaying networks with transceiver hardware impairments," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 12, no. 9, pp. 4295–4316, 2018.

[13] A. K. Mishra, D. Mallick, and P. Singh, "Combined effect of rf impairment and cee on the performance of dual-hop fixed-gain af relaying," *IEEE Communications Letters*, vol. 20, no. 9, pp. 1725–1728, 2016.

[14] X. Li, M. Liu, C. Deng, D. Zhang, X.-C. Gao, K. M. Rabie, and R. Kharel, "Joint effects of residual hardware impairments and channel estimation errors on swipt assisted cooperative noma networks," *IEEE Access*, vol. 7, pp. 135 499–135 513, 2019.

[15] X. Li, M. Liu, C. Deng, P. T. Mathiopoulos, Z. Ding, and Y. Liu, "Full-duplex cooperative noma relaying systems with i/q imbalance and imperfect sic," *IEEE Wireless Communications Letters*, vol. 9, no. 1, pp. 17–20, 2019.

[16] A. Afana, N. Abu-Ali, and S. Ikki, "On the joint impact of hardware and channel imperfections on cognitive spatial modulation mimo systems: Cramer–rao bound approach," *IEEE Systems Journal*, vol. 13, no. 2, pp. 1250–1261, 2018.

[17] M. Jain, S. Soni, N. Sharma, and D. Rawal, "Performance analysis at far and near user in noma based system in presence of sic error," *AEU-International Journal of Electronics and Communications*, vol. 114, p. 152993, 2020.

[18] F. A. Onat, A. Adinoyi, Y. Fan, H. Yanikomeroglu, J. S. Thompson, and I. D. Marsland, "Threshold selection for snr-based selective digital relaying in cooperative wireless networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4226–4237, 2008.

[19] R. S. Kshetrimayum, *Fundamentals of MIMO wireless communications*. Cambridge University Press, 2017.

Design of a Triple-Band High-Gain Fractal MiMo Antenna with Low ECC for Satellite and Radar Systems

Abdelbasset Azzouz*, Rachid Bouhmidi*, Mohammed Chetioui*
* LESM, University of Saida, Saida, Algeria
abdelbasset.azzouz@univ-saida.dz; rachid.bouhmidi@univ-saida.dz; mohammed.chetioui@univ-saida.dz

***Abstract*—This paper presents a high-performance two-element MIMO circular fractal antenna designed for broadband applications in the 2.5 to 10 GHz frequency range. The compact antenna, with dimensions of 50x80x0.8 mm³, is designed using FR4 substrate with a dielectric constant of 4.4 and a loss tangent of 0.02. The antenna demonstrates remarkable performance at three resonating frequencies: 4.07 GHz, 7.06 GHz, and 9.9 GHz, corresponding to C-band satellite communications and X-band radar applications. The proposed antenna achieves high gains of 9.64 dBi, 25.89 dBi, and 16.75 dBi at these respective frequencies, making it highly suitable for high-gain communication systems. The return losses at the resonating frequencies are -23.75 dB, -22.90 dB, and -27.74 dB, with corresponding bandwidths of 100 MHz, 110 MHz, and 120 MHz, ensuring efficient impedance matching and wideband operation. The envelope correlation coefficient (ECC) remains mostly below 0.02, indicating excellent MIMO performance with minimal signal interference between the antenna elements. The superior gain, broad bandwidth, and low ECC values highlight the antenna's potential for high-frequency, high-reliability communication systems.**

***Index Terms*—Envelope correlation coefficient, Fractal antenna, High gain, MIMO antenna, Radar systems, Satellite communications.**

I. Introduction

The rapid evolution of wireless communication systems, particularly those utilizing Multiple-Input Multiple-Output (MIMO) technology, has driven the demand for antennas that offer enhanced performance across multiple frequency bands. MIMO systems, which leverage multiple antennas at both the transmitter and receiver ends, require antennas with low mutual coupling and high diversity performance to maximize data throughput and reliability. In response to these needs, the design of fractal antennas has gained significant attention due to their unique ability to operate efficiently over a broad range of frequencies while maintaining a compact size. Jeet Banerjee et al [1] presented a multiband MIMO antenna with orthogonal orientation, tailored for WLAN, C-Band, and X-Band wireless applications. Mandeep Singh and Simranjit Singh [2] proposed a 2×2 slotted MIMO printed antenna with triple-band capability, designed for next-generation wireless communication. Ishita Agarwal et al [3] developed a multi-band MIMO antenna with an orthogonal, bird-shaped design, featuring high isolation and efficiency for 5G technology.

979-8-3315-3074-7/24/$31.00 © 2024 IEEE

Abdullah A. Jabber and Raad H. Thaher [4] designed a reconfigurable multiband MIMO antenna intended for wireless communication systems. Aidi Ren et al [5] introduced a triple-band antenna pair configured in an eight-element MIMO array, optimized for 5G smartphone applications. Jeet Banerjee et al [6] also proposed enhancements in isolation for UWB MIMO antennas using fractal and hybrid-fractal designs.

Ranjana Kumari and V.K. Tomar [7] optimized a reconfigurable MIMO antenna with selectable notches for WLAN, WiMAX, and X-Band applications. Hemalatha T et al [8] presented a UWB MIMO antenna with semi-elliptical slots, featuring multiple notch characteristics and improved isolation. Vishrut Sharma et al [9] proposed a four-element MIMO antenna with CPW feeding and tri-band operation for wireless communication. Lastly, Praveen V. Naidu et al [10] designed a MIMO antenna with a tree branch design and triple-band ACS feeding, aimed at high-speed data applications.

This paper presents the design of a triple-band high-gain fractal antenna specifically tailored for advanced wireless MIMO communication systems. The proposed antenna not only achieves high gain across the targeted frequency bands but also maintains a low Envelope Correlation Coefficient (ECC), ensuring excellent diversity performance and reliable signal transmission, making it an ideal candidate for next-generation MIMO communication systems. The remainder of this paper is organized as follows: Section II provides a detailed overview of the design methodology for the proposed triple-band fractal antenna, including the mathematical foundations and simulation setup. Section III presents the performance analysis of the antenna, with a focus on gain, return loss, and radiation patterns across the three targeted frequency bands. Also examines the antenna's MIMO characteristics, highlighting the Envelope Correlation Coefficient (ECC) and diversity gain. Finally, Section IV concludes the paper, summarizing the key findings and suggesting potential areas for future research.

II. Antenna Design

A. Single Element

Figure 1 shows the dimensions of the single-element antenna, which features a circular radiating patch with a central circular slot. This patch is printed over a microstrip feed line, which is etched onto an FR-4 substrate. The substrate has a

relative permittivity of $\varepsilon_r = 4.4$, a loss tangent of 0.02, and a thickness of 0.8 mm. The optimized design parameters and corresponding dimensions are listed in Table I.

B. 2x2 MIMO antenna

To further increase the data rate and improve both the range and coverage of the antenna, the single-element antenna is transformed into a 2-element MIMO antenna. The proposed MIMO antenna configuration is illustrated in Figure 2, and the dimensions of the 2-port MIMO antenna are provided in Table II.

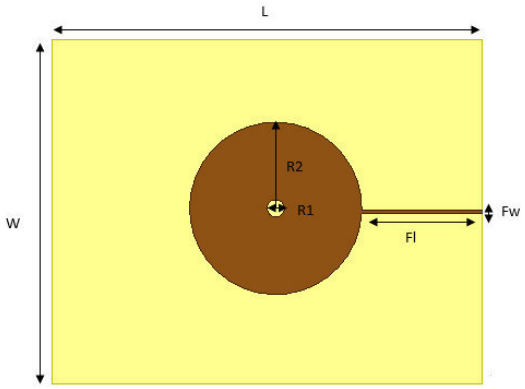


Fig. 1. Top view of the proposed antenna.

TABLE I
PARAMETERS AND THEIR VALUES OF THE PROPOSED ANTENNA

Parameters	Values (mm)	Parameters	Values (mm)
L	40	Fw	0.3
W	50	Fl	14.02
R1	1	R2	10

TABLE II
PARAMETERS AND THEIR VALUES OF THE PROPOSED TWO-ELEMENT MIMO ANTENNA

Parameters	Values (mm)	Parameters	Values (mm)
L	80	Fw	0.3
W	68	Fl	14.02
R1	1	R2	10

The resonant frequency f_r of a circular patch antenna is given by:

$$f_r = \frac{1.8412 \times c}{2\pi a\sqrt{\epsilon_r}} \tag{1}$$

The effective radius a_{eff} is:

$$a_{\text{eff}} = a\left[1 + \frac{2h}{\pi a \epsilon_r}\left(\ln\left(\frac{\pi a}{2h}\right) + 1.7726\right)\right]^{\frac{1}{2}} \tag{2}$$

The input impedance Z_{in} at the edge of the patch is:

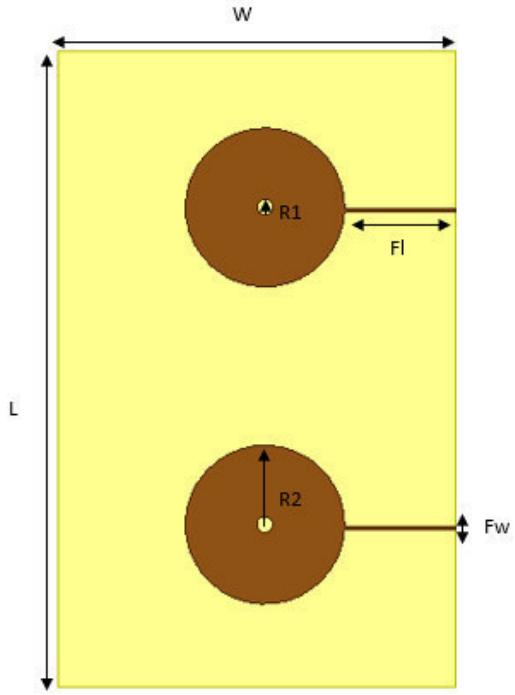


Fig. 2. Top view of the proposed MIMO antenna.

$$Z_{\text{in}} = \frac{120\pi}{2\epsilon_r}\frac{J_1(k_0 a_{\text{eff}})}{k_0 a_{\text{eff}} J_0(k_0 a_{\text{eff}})} \tag{3}$$

The radiation pattern $F(\theta)$ in the θ-plane is:

$$F(\theta) = \frac{2J_1(k_0 a \sin\theta)}{k_0 a \sin\theta} \tag{4}$$

III. RESULTS AND DISCUSSIONS

All simulations were performed using High-Frequency Structure Simulator.(HFSS) version 15. In addition to the reflection coefficient and transmission coefficient analysis, further evaluations are presented, including gain, radiation pattern and voltage standing wave ratio (VSWR).

A. Return Loss

The reflection coefficient and transmission coefficient analysis presented in Figure 3 indicates that the designed MIMO antenna exhibits triple band characteristics at the frequencies of 4.07 GHz, 7.06 GHz, and 9.60 GHz, with corresponding return losses of -23.74 dB, -22.91 dB, and -27.74 dB, respectively, It is evident that all the transmission coefficients are significantly below -16 dB, indicating that the simulated results are highly suitable for the MIMO antenna configuration.

The return loss RL in dB is given by:

$$RL = -20\log_{10}|\Gamma| \tag{5}$$

where the reflection coefficient Γ is:

$$\Gamma = \frac{Z_{in} - Z_0}{Z_{in} + Z_0} \tag{6}$$

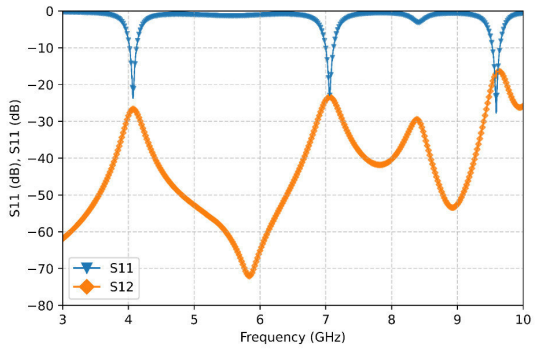


Fig. 3. Transmission and reflection coefficients of the proposed antenna.

Here, Z_{in} is the input impedance of the antenna, and Z_0 is the characteristic impedance of the transmission line.

B. VSWR

Ideally, the voltage standing wave ratio (VSWR) should be equal to 1. The proposed antenna exhibits VSWR values of 1.13, 1.15, and 1.06 at the resonant frequencies of 4.07 GHz, 7.06 GHz, and 9.60 GHz, respectively, the plot of VSWR is shown in in Figure 4 .

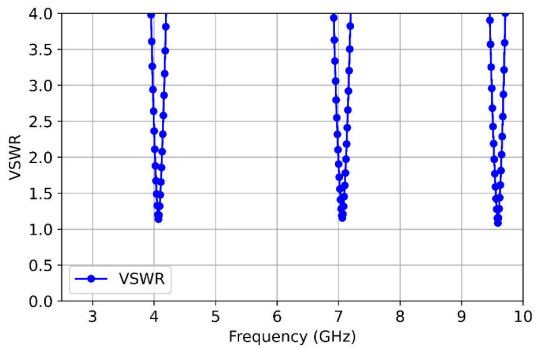


Fig. 4. Voltage Standing Wave Ratio (VSWR) of the proposed antenna.

The Voltage Standing Wave Ratio (VSWR) is given by:

$$\text{VSWR} = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{7}$$

where Γ is the reflection coefficient.

C. Radiation Pattern

The radiation patterns of the presented MIMO antenna are shown in Figure 5. The antenna exhibits good radiation patterns, evaluated at the operating frequencies in the E-plane and H-plane.

The radiation pattern $P(\theta, \phi)$ of the antenna is given by:

$$P(\theta, \phi) = |F(\theta, \phi)|^2 \tag{8}$$

where $F(\theta, \phi)$ is the far-field radiation function.

D. Gain

Figure 6 illustrates the three-dimensional gain patterns at the resonant frequencies of 4.07 GHz, 7.06 GHz, and 9.60 GHz, with corresponding high gain values of 9.64 dBi, 25.89 dBi, and 16.75 dBi, respectively. Ideally, the gain of an antenna should be optimized for the specific application, with higher gain values generally preferred for increased signal strength and coverage.

The gain $G(\theta, \phi)$ of the antenna is given by:

$$G(\theta, \phi) = \frac{4\pi U(\theta, \phi)}{P_{\text{in}}} \tag{9}$$

where $U(\theta, \phi)$ is the radiation intensity in a given direction, and P_{in} is the total input power to the antenna.

E. Envelope Correlation Coefficient

The Envelope Correlation Coefficient (ECC) quantifies the degree of independence between the radiation patterns of two antennas and can be calculated using the provided equations. Figure 7 shows the ECC plotted against frequency. For the proposed configuration, the ECC is mostly below 0.02 across the operating frequency ranges.

The Envelope Correlation Coefficient (ECC) between two antennas is given by:

$$\text{ECC} = \frac{\left|S_{11}^* S_{12} + S_{21}^* S_{22}\right|^2}{\left(1 - |S_{11}|^2 - |S_{21}|^2\right)\left(1 - |S_{22}|^2 - |S_{12}|^2\right)} \tag{10}$$

where S_{11} and S_{22} are the reflection coefficients, and S_{12} and S_{21} are the transmission coefficients between the antennas.

F. Diversity Gain

The simulated values of diversity gain (DG) are shown in Figure 8, remaining close to 10 dB across the operating bandwidth. This confirms the suitability of the proposed MIMO antenna for wireless applications.

The diversity gain DG in a 2-antenna system is given by:

$$DG = 10\sqrt{1 - \text{ECC}} \tag{11}$$

where ECC is the Envelope Correlation Coefficient between the antennas.

G. Antenna Efficiency

The diagram in Figure 9 presents the simulated results for total efficiency, demonstrating that the antenna achieves an efficiency of 90%.

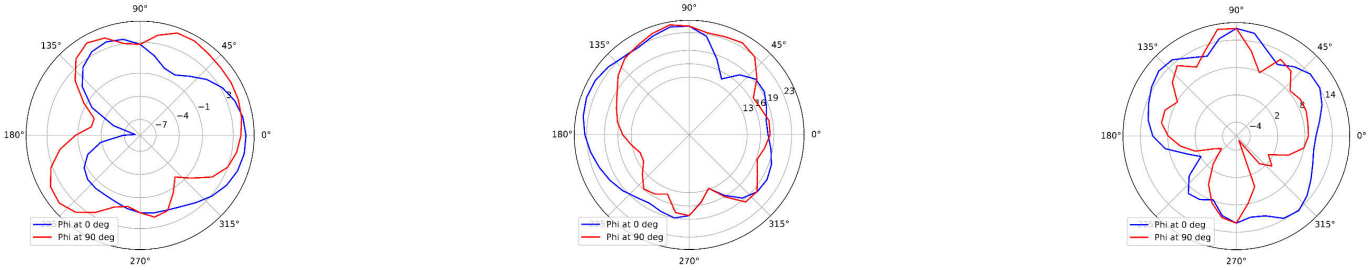
The antenna efficiency η is given by:

$$\eta = \eta_r \times (1 - |\Gamma|^2) \tag{12}$$

Alternatively, it can be expressed as:

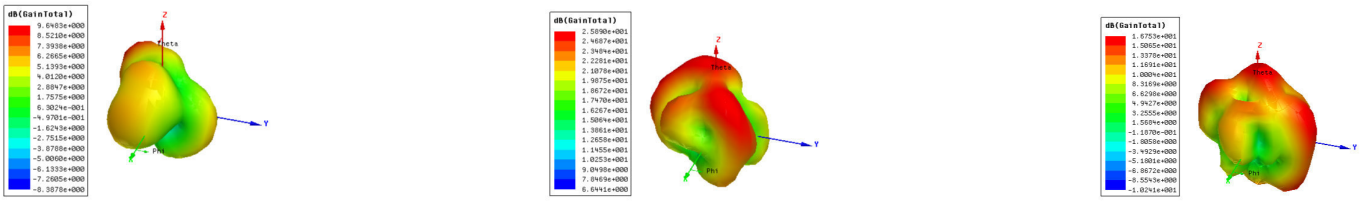
$$\eta = \frac{P_{\text{rad}}}{P_{\text{in}}} \tag{13}$$

where η_r is the radiation efficiency, $|\Gamma|^2$ is the squared reflection coefficient, P_{rad} is the radiated power, and P_{in} is the total input power to the antenna.



(a) (b) (c)

Fig. 5. Radiation pattern of the proposed antenna at (a) 4.07 GHz, (b) 7.059 GHz, (c) 9.59 GHz.



(a) (b) (c)

Fig. 6. Gain of the proposed antenna at (a) 4.07 GHz, (b) 7.059 GHz, (c) 9.59 GHz.

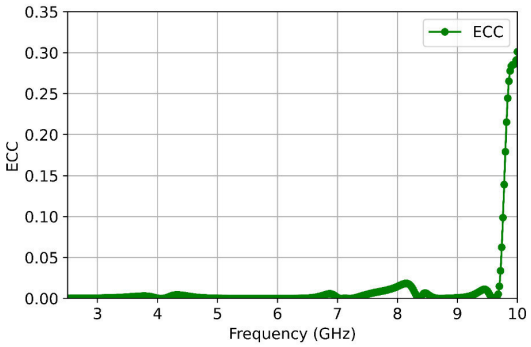


Fig. 7. Envelope Correlation Coefficient (ECC) of the proposed antenna.

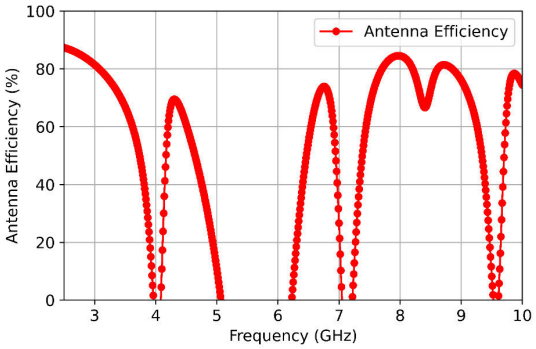


Fig. 9. Antenna efficiency of the proposed antenna.

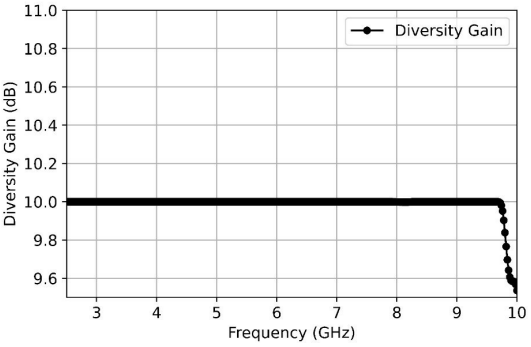


Fig. 8. Diversity gain of the proposed antenna.

H. Surface Current Distribution

To examine the current distribution on the radiating element at the operating frequency, surface current results were recorded throughout the evolution process. These results, along with their corresponding scale, are presented in Figure 10.

IV. Conclusion

In conclusion, the proposed two-element MIMO circular fractal antenna demonstrates exceptional performance across a broad frequency range of 2.5 to 10 GHz, making it an excellent candidate for advanced communication systems. The antenna's high gain, particularly at the key resonant frequencies of 4.07 GHz, 7.06 GHz, and 9.9 GHz, ensures robust signal transmission for both C-band satellite communications and X-band radar applications. The low return loss and wide bandwidth at these frequencies confirm efficient impedance matching and reliable operation. Additionally, the consistently low envelope correlation coefficient (ECC) indicates supe-

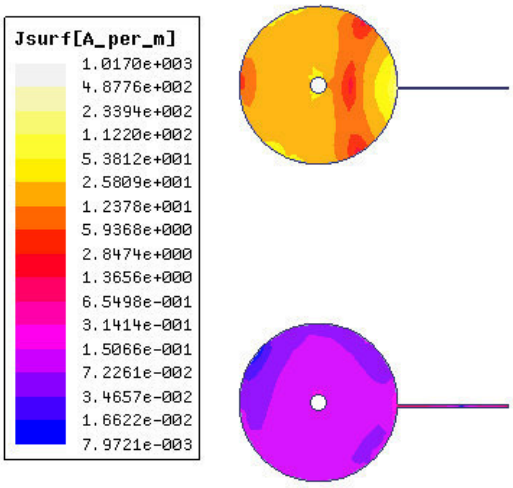


Fig. 10. Current distribution of the proposed antenna.

rior MIMO performance with minimal interference, further enhancing the antenna's suitability for high-frequency, high-reliability applications. The compact design, combined with the use of an FR4 substrate, underscores the practicality and effectiveness of this antenna for integration into modern communication devices.

REFERENCES

[1] J. Banerjee, S. Bhattacharya, A. Choudhury, A. Biswas, R. Paul, A. Dutta, and R. Ghatak, "An orthogonally oriented multiband mimo antenna for wlan, c-band and x-band wireless applications," in *2020 IEEE Calcutta Conference (CALCON)*, pp. 328–332, IEEE, 2020.

[2] M. Singh and S. Singh, "Triple band 2× 2 slotted mimo printed antenna for next generation wireless applications," in *2020 Indo–Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN)*, pp. 36–40, IEEE, 2020.

[3] I. Agarwal, P. Wadhwa, S. Vignesh, and D. Yadav, "Orthogonal bird shaped multiband mimo antenna for 5g technology with high isolation and efficiency," in *2020 4th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)*, pp. 1–5, IEEE, 2020.

[4] A. A. Jabber and R. H. Thaher, "Multiband reconfigurable mimo antenna for wireless applications," in *2020 6th International Engineering Conference "Sustainable Technology and Development"(IEC)*, pp. 36–41, IEEE, 2020.

[5] A. Ren, H. Yu, Z. Zhang, and H. Liu, "An eight-element mimo array based on triple-band antenna pair for 5g smartphone applications," in *2022 International Applied Computational Electromagnetics Society Symposium (ACES-China)*, pp. 1–2, IEEE, 2022.

[6] J. Banerjee, U. Bose, and R. Ghatak, "Improvement of isolation in uwb mimo antennas with fractal and hybrid-fractal structures," in *2023 8th International Conference on Computers and Devices for Communication (CODEC)*, pp. 1–2, IEEE, 2023.

[7] R. Kumari and V. Tomar, "On demand notched reconfigurable mimo antenna for wlan, wimax, and x-band applications," in *2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI)*, pp. 1–7, IEEE, 2023.

[8] T. Hemalatha, B. Roy, U. Maheswari, *et al.*, "Semi-elliptical slotted uwb mimo antenna with multiple notch characteristics and high isolation," in *2024 4th URSI Atlantic Radio Science Meeting (AT-RASC)*, pp. 1–4, IEEE, 2024.

[9] V. Sharma, T. Shringi, G. Sharma, D. Bhatia, and M. Kumar, "Tri-band cpw-fed four element mimo antenna for wireless applications," in *2023 Fourth International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, pp. 1–5, IEEE, 2023.

[10] P. V. Naidu, N. Kalasani, P. Hareesha, T. Dheeraj, E. A. Vardhan, and A. Kumar, "Tree branch shaped triple band acs fed mimo antenna for high speed data applications," in *2022 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, pp. 1–6, IEEE, 2022.

Design of Hexagonal Symmetry Meets Circular Polarization Antenna

1st BOUAMAMA Safi
Laboratory of Telecommunication of Tlemcen (LTT),Department of Telecommunication, University of Tlemcen, 13000, Algeria,
e-mail:bouamama.safi@gmail.com

2nd MERIAH SIDI Mohammed
Laboratory of Telecommunication of Tlemcen (LTT),Department of Telecommunication, University of Tlemcen, 13000, Algeria,
e-mail:

3rd BELGACEM Wahiba
Laboratory of Telecommunication of Tlemcen (LTT),Department of Telecommunication, University of Tlemcen, 13000, Algeria,
e-mail:

***Abstract*—** This paper presents a circularly polarized antenna with a symmetrical irregular hexagonal design, optimized for performance at 2.45 GHz. The antenna is fabricated on FR-4 substrate, achieving impressive results with an S11 reflection coefficient of -24.2 dB, an axial ratio (AR) of 2.72, and a voltage standing wave ratio (VSWR) under 2, indicating strong impedance matching. The design achieves notable gain, enhancing efficiency in practical applications. Optimization was conducted using CST Microwave Studio, leveraging a genetic algorithm to refine the antenna's parameters for optimal circular polarization and gain. This innovative approach to design and optimization demonstrates the potential of hexagonal symmetry in advancing compact, high-performance antennas for applications in wireless communication and IoT systems.

Keywords: Antenna, Circular polarization, Optimization, Performance, CST.

I. INTRODUCTION

The rapid growth of wireless communication systems has intensified the demand for compact, high-performance antennas with robust polarization characteristics. Circularly polarized (CP) antennas have emerged as a preferred choice in modern wireless applications due to their ability to mitigate polarization mismatch and multipath interference, key issues in mobile and IoT environments. Unlike linearly polarized antennas, CP antennas are less sensitive to the orientation of transmitting and receiving antennas, making them ideal for dynamic mobile applications where signal reliability is critical. Circular polarization is particularly beneficial in reducing the effects of signal fading and enhancing link stability, which are crucial for applications such as wireless sensing, satellite communication, and wearable health monitoring devices [1].

In this work, we introduce a circularly polarized antenna design with a symmetrical, irregular hexagonal shape, optimized for operation at 2.45 GHz. The choice of a hexagonal geometry represents an innovative approach, providing a balanced structure that supports circular polarization and compactness. Fabricated on a standard FR-4 substrate, this antenna demonstrates excellent performance metrics, including a reflection coefficient (S11) of -24.2 dB, an axial ratio (AR) of 2.72, and a voltage standing wave ratio (VSWR) of less than 2, indicating strong impedance matching and effective circular polarization. The optimization was conducted using CST Microwave Studio, employing a genetic algorithm to fine-tune the antenna's parameters for maximum efficiency and gain.

II. MATERIALS AND METHODS

A. Antenna design and geometry:

Figure 1 illustrates a view of the proposed antenna. This design employs a microstrip patch antenna, chosen for its compact size, affordability, low profile, and ease of fabrication . These antennas are commonly used in the ISM band and at higher frequencies, where their dimensions are determined by resonant frequency and wavelength. Here, a square microstrip patch antenna is utilized, with design specifications calling for a 2450 MHz operating frequency, a 50-ohm impedance, and circular polarization of the radiated field. Figure 2 provides a closer look at the antenna's structure, which incorporates an FR-4 substrate. This material, selected for its accessibility and suitable properties, features a thickness of 1.6 mm, a dielectric constant of 4.3, and a loss tangent of 0.0025, making it an ideal choice for this study's requirements

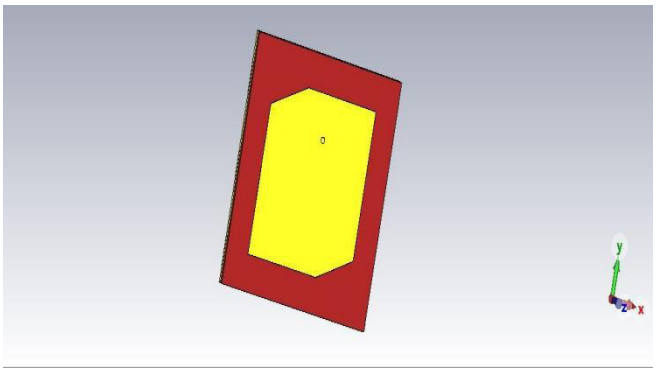


Fig 1: Symmetrical irregular hexagonal CP antenna.

B. Analytic study and dimensions

The first step in designing the circularly polarized microstrip patch antenna involves determining the dimensions of the antenna. To calculate the size of the rectangular patch antenna suitable for a resonant frequency of 2.45 GHz, and considering a substrate with a dielectric

constant of 4.3 and a thickness of 1.6 mm, the following equations are utilized [2].

$$W = \frac{c}{2 f_0 \sqrt{\frac{\varepsilon_r + 1}{2}}} \quad (1)$$

$$L = L_{eff} - 2\Delta L \quad (2)$$

$$L_{eff} = \frac{C}{2 f_0 \sqrt{\varepsilon_{reff}}} \quad (3)$$

$$\varepsilon_{reff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + 12\frac{h}{W}\right)^{-\frac{1}{2}} \quad (4)$$

$$\Delta L = 0.412h\left[\frac{\left(\varepsilon_{reff} + 0.3\right)\left(\frac{W}{h} + 0.264\right)}{\left(\varepsilon_{reff} - 0.258\right)\left(\frac{W}{h} - 0.8\right)}\right] \quad (5)$$

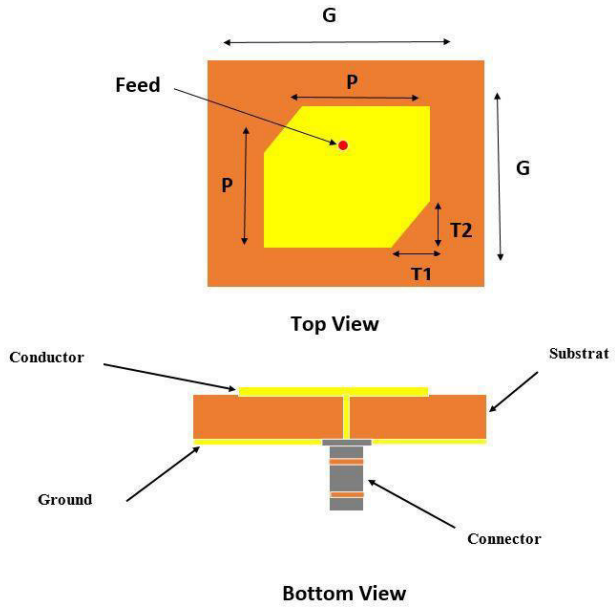


Fig.2: Geometry of the proposed CP antenna

Table 1: Optimized dimensions of the proposed antenna.

Items	Optimal Value [mm]	Symbols
Length of patch	28.48	L
Width of patch	28.48	L
Length of square substrate	40	G
Feeding point in x-axis	0	XF
Feeding point in y-axis	4	YF
Length of triangle's first slit side	4.5	T1
Length of triangle's second slit side	8	T2
Inner Radius	0.7	R1
Outer Radius	1.35	R2

III. GENETIC ALGORITHM OPTIMIZATION FOR ANTENNA DESIGN

In this study, a genetic algorithm (GA) was implemented within CST Microwave Studio to optimize key performance parameters of the proposed antenna design, including return loss, axial ratio (AR), voltage standing wave ratio (VSWR), gain, and miniaturization. Genetic algorithms, inspired by biological evolution, provide a robust and efficient approach for navigating the large design space associated with antenna parameters. This method facilitates the development of an antenna that meets strict requirements for wireless communication systems, such as impedance matching, circular polarization, and compactness [3,4].

For the proposed antenna, the genetic algorithm was set to improve multiple critical metrics:

- Return Loss: Minimizing return loss ensures maximum power transfer between the antenna and the transmission line, leading to efficient performance at the desired frequency (2.45 GHz).
- Axial Ratio (AR): An optimal axial ratio indicates effective circular polarization, which is critical for reducing signal degradation in multipath environments.
- VSWR (Voltage Standing Wave Ratio): Lowering the VSWR below 2 improves impedance matching, reducing signal reflection and increasing the reliability of data transmission.
- Gain: By maximizing gain, the antenna can cover a larger area or reach greater distances with improved signal strength, which is essential for many wireless applications.
- Miniaturization: Given the design's application needs, compactness was prioritized, making the antenna suitable for integration into various devices without compromising performance.

Using the GA, CST Studio iteratively tuned these parameters to balance the competing requirements, resulting in an optimized antenna design that offers high efficiency, compact size, and robust performance. This approach highlights the effectiveness of genetic algorithms in antenna optimization, as they allow for the exploration of numerous design variations and the achievement of a finely-tuned balance between multiple design constraints.

IV. RESULTS AND DISCUTION

This study focuses on enhancing the performance of a Circularly Polarized (CP) antenna through a symmetrical irregular hexagonal design. The proposed antenna incorporates a novel symmetrical irregular hexagon shape, which aims to optimize key performance parameters such

as return loss, gain, Voltage Standing Wave Ratio (VSWR), axial ratio, and radiation pattern. The investigation explores the effects of this unique design on the antenna's efficiency and effectiveness, demonstrating the potential of the symmetrical irregular hexagonal configuration to advance the performance of CP antennas in wireless communication applications.

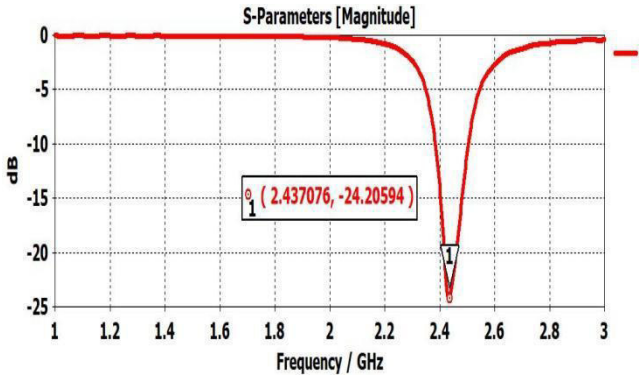


Fig 3: The Return loss of the proposed CP Antenna

A. S-parameters:

The S-parameter graph provides valuable insights into the impedance matching and overall performance of the antenna. Specifically, the S11 parameter, which represents the reflection coefficient, is critical for assessing how much of the signal is reflected back from the antenna due to impedance mismatch. In this study, the S11 value reaches -24.2 dB at the operating frequency of 2.45 GHz, indicating excellent impedance matching and minimal reflection. A low S11 value, typically below -10 dB [5], shows that most of the power is being transmitted into the antenna, which is desirable for efficient operation. The S-parameter graph also reveals the resonance frequency, where the antenna performs optimally, and provides insight into the bandwidth and efficiency of the design. By analyzing this curve, we can confirm the antenna's effective operation at the desired frequency with strong signal integrity and minimal signal loss.

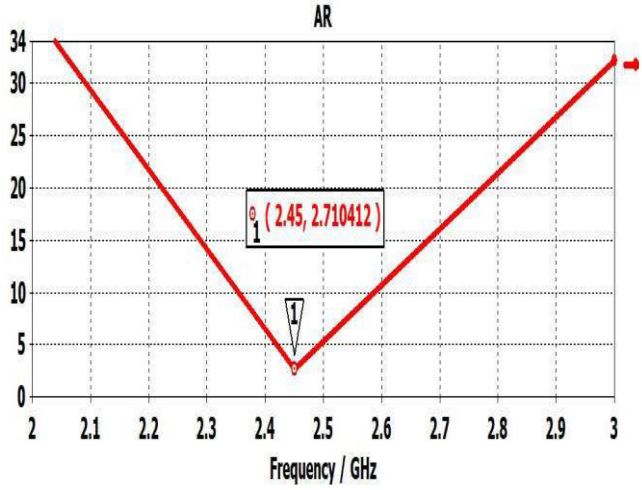


Fig 4: Axial Ratio of the proposed CP antenna

B. Axial Ratio:

The axial ratio (AR) is an important parameter for assessing the polarization characteristics of the antenna, particularly for circularly polarized antennas. It describes the quality of the circular polarization, with an ideal AR value of 1 (0 dB) indicating perfect circular polarization [6][7]. In this study, the axial ratio of the proposed antenna is measured to be 2.72, which demonstrates good circular polarization. A lower AR value signifies a more efficient polarization, as it indicates less deviation from perfect circularity. An AR value below 3 dB is generally considered acceptable for many wireless communication applications, as it ensures that the antenna maintains a relatively consistent radiation pattern in all directions. The axial ratio graph allows us to visualize how well the antenna maintains circular polarization across its operating frequency and further confirms its suitability for reliable signal transmission in environments with multipath interference.

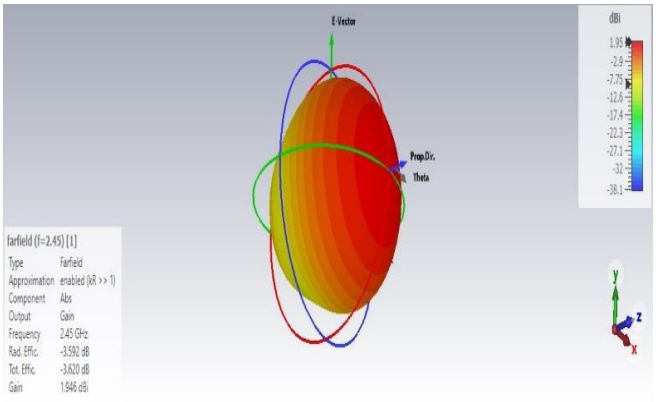


Fig 5:3D Radiation pattern of proposed model

C. Radiation pattern:

The radiation pattern of an antenna is a critical parameter that illustrates the distribution of radiated power as a function of direction in space. It provides insight into how effectively the antenna transmits or receives signals at different angles. For the proposed antenna, the radiation pattern is evaluated to understand its directional characteristics and efficiency in transmitting signals. A well-designed antenna should ideally exhibit a wide and consistent radiation pattern, ensuring effective coverage in the desired directions. In the case of a circularly polarized antenna, the radiation pattern also reveals how the antenna's polarization remains consistent across different orientations, which is crucial for minimizing signal degradation due to polarization mismatch [8]. By analyzing the radiation pattern, we can assess the antenna's performance in practical scenarios, including its ability to deliver uniform signal strength and coverage, especially in environments where the angle of arrival or reception may vary. This information is essential for determining the antenna's suitability for applications in wireless communication systems, where signal reliability and uniformity are key factors.

V. CONCLUSION

In this work, we presented a novel circularly polarized antenna featuring a symmetrical, irregular hexagonal design, optimized to operate efficiently at 2.45 GHz on an FR-4 substrate. The design achieved excellent performance metrics, including a reflection coefficient (S11) of -24.2 dB, an axial ratio (AR) of 2.7, and a VSWR below 2, indicating effective circular polarization and strong impedance matching. Optimization through CST Microwave Studio, guided by a genetic algorithm, allowed for precise adjustments, maximizing gain and enhancing the antenna's overall performance. This hexagonal geometry demonstrates a compact, reliable solution for modern wireless applications, particularly in IoT, healthcare, and other mobile settings.

V. REFERENCES

[1] A. Sayem, R. Simorangkir, K. Esselle, Lalbakhsh, D. Gawade, B. O'Flynn and J. Buckley "Flexible and Transparent Circularly Polarized Patch Antenna for Reliable Unobtrusive Wearable Wireless" Communications. Sensors 2022, 22, 1276.

[2] C. A. Balanis, Antenna Theory, Analysis and Design, John Wiley & Sons, Hoboken, New Jersey, May 2005.

[3] M. Lamsalli, A. Hamichi, M. Boussouis, N. Amar, T. Elhamadi, "Genetic Algorithm Optimization for Microstrip Patch Antenna Miniaturization," Progress In Electromagnetics Research Letters, Vol. 60, 113-120, 2016.

[4] R. P.Moroyoqui, S. R.Romo, "Genetic algorithm for the design of a multiband microstrip antenna with self-avoiding geometry obtained by backtracking," International Journal of Communication Systems Volume35, Issue18 December 2022.

[5] M. S. Ellis, Z. Zhao, J. Wu, X. Ding, Z. Nie and Q. H. Liu, "Novel Simple and Compact Micro Strip-Fed Circularly-Polarized Wide Slot Antenna with Wide Axial Ratio Bandwidth", IEEE Transactions On Antenna and Propagation, vol. 64, no. 4, pp. 1552-1555, 2016.

[6] S.S. Goa, Q. Luo, F. Zhu, Circularly Polarized Antennas, Hoboken, NJ, USA: Wiley, 2014.

[7] J. Huang, A. C. Densmore, "Microstrip Yagi array antenna for mobile satellite vehicle application," IEEE Transactions on Antennas and Propagation, vol. 39, no. 7, pp. 1024–1030, Jul. 1991.

[8] S.Pramono,F.Rahutomo,W. A. Syafei, "Axial Ratio-Bandwidth Enhancement of a Compact L Band Circularly Polarized Antenna Using a Simple Diagonal Slot and Corner Truncation," 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE) 2021.

Advanced Systems in Electrical Engineering

Artificial intellegence and its applications

SAMET KOURDI Rahma
Phd Student -Industrial Engineering
National Higher School of Technology and Engineering.
Annaba, ALGERIA
K.samet@etu.ensti-annaba

***Abstract*—Artificial intelligence (AI) is playing an increasingly significant role in the field of electrical engineering, offering innovative and effective solutions to optimize complex systems, enhance energy efficiency, and transform the management of electrical infrastructures. This article examines the main applications of AI in this domain, highlighting intelligent energy systems, electrical networks, as well as automation and process control. We will also discuss the challenges and future prospects related to the integration of AI into modern electrical systems.**

***Index Terms*—Artificial Intelligence (AI),Machine Learning,Deep Learning,Neural Networks,Smart Grids,Energy Management Systems,Predictive Maintenance .**

I. Introduction

Artificial Intelligence (AI) is transforming many sectors, and the field of electrical engineering is no exception to this technological revolution. This article explores the applications of AI in electrical engineering, highlighting its contributions to energy efficiency, predictive maintenance, and the optimization of electrical systems. Artificial intelligence encompasses techniques that enable machines to simulate human behaviors, including learning, reasoning, and self-correction. In the context of electrical engineering, AI is used to analyze large amounts of data, optimize systems, and enhance decision-making. Artificial Intelligence (AI) refers to the ability of computer systems to replicate human intelligence processes. This encompasses skills such as machine learning, logical reasoning, perception, and autonomous decision-making. In the field of electrical engineering, AI is employed to tackle complex issues in constantly evolving environments, where systems must adapt in real time and optimize their performance. With the increasing integration of renewable energy sources, the digitization of infrastructures, and the rising demand for energy efficiency, modern electrical systems are becoming increasingly sophisticated. AI plays a key role in managing these challenges by providing advanced solutions for intelligent energy management, network monitoring, and predictive maintenance.

Identify applicable funding agency here. If none, delete this.

II. Applications of AI in Electrical Engineering

A. Predictive Maintenance

AI-driven predictive maintenance techniques utilize real-time data from equipment sensors to anticipate potential failures before they happen. This proactive strategy reduces unplanned downtime and lowers maintenance costs by facilitating timely interventions. Additionally, machine learning algorithms can analyze historical maintenance data to identify patterns, leading to more precise predictions regarding equipment performance and lifespan.

B. Energy Management

AI enhances energy management in electrical systems by examining usage patterns and optimizing operations to boost efficiency. This includes the effective integration of renewable energy sources and the management of load distribution.

C. Smart Grids

AI technologies improve the functionality of smart grids by enabling real-time monitoring and control of electrical distribution networks. This capability assists in balancing supply and demand, integrating distributed energy resources, and enhancing grid reliability.

D. Fault Detection and Diagnostics

AI systems are capable of automatically detecting anomalies within electrical systems, providing early warnings for potential faults. This functionality is essential for maintaining system integrity and ensuring safety.

E. Automation and Control

AI promotes advanced automation in electrical engineering processes, allowing for autonomous decision-making in system operations. This encompasses the optimization of control strategies across various applications, ranging from manufacturing to building management systems.

F. Data Analytics

By leveraging advanced analytics, AI processes large volumes of data generated by electrical systems to derive actionable insights. These insights empower engineers to make informed decisions about system design, operation, and maintenance.

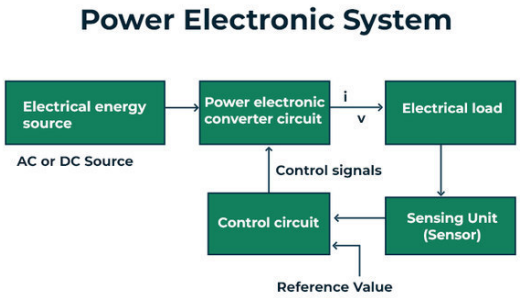


Fig. 1. Power Electronic System

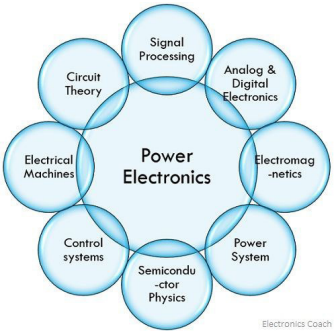


Fig. 2. Realation of power electronics with other disciplines

G. Digital Twins

AI facilitates the development of digital twins—virtual representations of physical systems—that simulate real-time performance and predict future behavior. This technology is beneficial for testing scenarios and optimizing system designs without the need for physical modifications.

H. Robotics

In the realm of electrical engineering, AI-powered robotics are employed for tasks such as inspection, maintenance, and repair of electrical infrastructure. This application enhances efficiency and safety while reducing the reliance on human labor.

These applications demonstrate how AI is transforming the field of electrical engineering by enhancing efficiency, reliability, and safety across diverse systems and processes.

III. The Role of AI in Electrical Engineering

Artificial Intelligence (AI) includes a diverse range of techniques that empower machines to execute tasks traditionally requiring human intelligence. Within the field of electrical engineering, various AI methodologies—such as machine learning, fuzzy logic, and metaheuristic algorithms—are utilized to address complex challenges and enhance system performance

IV. Main AI Methods in Power Electronics

The article titled ”An Overview of Artificial Intelligence Applications for Power Electronics” provides a statistical analysis of the application of artificial intelligence (AI) methods in the field of power electronics. This study is based on over

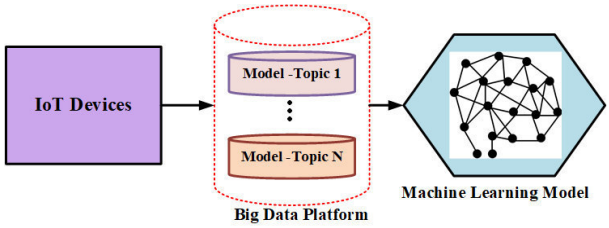


Fig. 3. integration-of Io T big data platform and ML

500 publications from the IEEE Xplore database, including journals such as IEEE Transactions on Power Electronics, IEEE Journal of Emerging and Selected Topics in Power Electronics, IEEE Transactions on Industrial Electronics, IEEE Transactions on Industrial Informatics, and IEEE Transactions on Industry Applications. According to the data presented in the article, the main AI methods used in power electronics are:

A. Machine Learning :

This method accounts for the majority of AI applications in the field, with a usage rate of 45.8Its popularity stems from its flexibility and ability to handle complex tasks. Supervised learning is the dominant form of machine learning in this sector. Machine learning (ML) has become a vital component of electrical engineering, fostering innovation across a range of applications. This synergy not only improves system performance but also optimizes processes and supports the creation of cutting-edge technologies.

1) **Major Applications :**

- Smart Grids and Energy Management: ML algorithms enhance energy distribution by accurately predicting demand, identifying faults, and enabling efficient management within smart grids.
- Predictive Maintenance: By forecasting potential equipment failures in electrical systems—such as transformers and circuit breakers—ML models facilitate timely maintenance, minimizing downtime.
- Energy Efficiency: Machine learning contributes to energy efficiency improvements in buildings and industrial operations by optimizing heating, ventilation, air conditioning (HVAC) systems, and lighting.

- Renewable Energy Integration: ML plays a crucial role in forecasting renewable energy generation, such as solar and wind power, thereby improving their integration into existing power grids.
- Power Quality Monitoring: Techniques in machine learning are utilized to monitor power quality issues like voltage sags and harmonics, ensuring a stable electrical supply.
- Robotics and Automation: In the field of robotics, ML is essential for tasks such as object recognition and path planning, significantly enhancing automation capabilities.

- Signal Processing: ML improves signal processing applications, leading to greater accuracy in data transmission for telecommunications and medical imaging.
- Autonomous Vehicles: Electrical engineers apply ML to develop algorithms that support self-driving technologies, including navigation systems and vehicle-to-vehicle communication.
- Healthcare Devices: The integration of ML into medical devices aids in tasks such as disease diagnosis and patient monitoring, improving healthcare delivery.
- Circuit Design Optimization: Machine learning assists in refining circuit designs to lower power consumption while enhancing performance.

Educational Integration Many educational institutions are now offering specialized programs that focus on the application of machine learning within electrical engineering. For example: The University of Wisconsin provides a Machine Learning and Data Science option within its Electrical Engineering curriculum, emphasizing the mathematical foundations and tools relevant to ML applications. Courses such as "Machine Learning for Electrical Engineering" address critical topics including regression, classification, neural networks, and deep learning techniques tailored for engineering purposes.

2) Future Trends

The future landscape of electrical engineering is expected to witness deeper integration of machine learning across various sectors: Adaptive Control Systems: These systems will utilize ML to improve performance in complex environments where traditional control methods may be inadequate. Cyber-Physical Systems: Ongoing research aims to create intelligent systems that merge physical processes with computational resources through ML for enhanced decision-making. Data-Driven Intelligence: The combination of big data analytics with ML is set to drive advancements in energy systems management and operational efficiency across multiple industries.

In conclusion, machine learning is transforming the field of electrical engineering by enhancing existing technologies and paving the way for innovative solutions across diverse applications. Its profound impact promises a future where electrical systems are more efficient, reliable, and intelligent.

B. Metaheuristic Methods:

Used at a rate of 32these methods are particularly effective for optimization in power electronics. The most common among them are Genetic Algorithms (GA) and Particle Swarm Optimization (PSO). Metaheuristic methods are sophisticated optimization techniques widely employed in electrical engineering to tackle complex problems that are often challenging for traditional optimization approaches. These methods are particularly effective for NP-hard problems, where obtaining exact solutions is computationally impractical.

1) **Characteristics of Metaheuristic Methods**

Stochastic Nature: Most metaheuristic algorithms incorporate randomness in their search processes, enabling them to explore the solution space more effectively without the need to exhaustively evaluate every possible option.

Near-Optimal Solutions: These methods focus on identifying satisfactory solutions rather than striving for exact optimality, which is especially beneficial in practical scenarios where perfect solutions are not required.

No Gradient Requirement: Unlike conventional optimization techniques, metaheuristics do not depend on gradient information, making them suitable for non-analytic or black-box objective functions.

Local Optima Recovery: Many metaheuristic algorithms possess mechanisms that allow them to escape local optima, thanks to their inherent randomness or specific heuristics, enabling them to manage uncertainties in objectives more effectively.

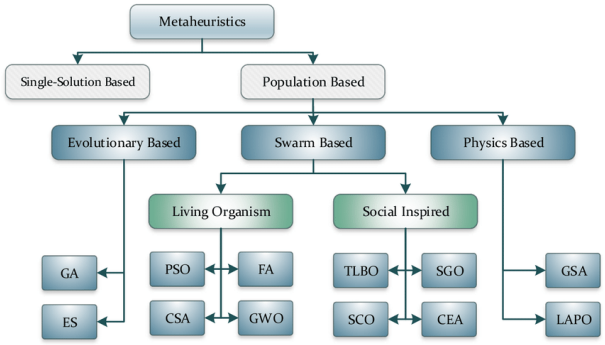


Fig. 4. Classification of methaheuristic methods

2) **Common Metaheuristic Techniques**

Metaheuristic methods can be categorized according to their approach to modifying solutions:

Single-Solution Methods: These begin with a single candidate solution and iteratively refine it.

Population-Based Methods These involve a group of solutions, simultaneously modifying multiple candidates to comprehensively explore the search space.

3) Applications in Electrical Engineering

Power System Optimization: Metaheuristic algorithms are utilized to optimize various components of power systems, such as optimal power flow, economic dispatch, and the placement of distributed generation sources. They contribute to reducing power losses and enhancing system reliability. Renewable Energy Forecasting: When combined with machine learning models, metaheuristics improve the accuracy of forecasting renewable energy generation (e.g., solar and wind power) by optimizing parameters in models like artificial neural networks (ANNs) and support vector machines (SVMs). Design Optimization: Metaheuristics aid the design process in electrical engineering by optimizing parameters in control systems, circuit designs, and other applications to achieve performance targets while minimizing costs. Smart Grid Management: These techniques play a role in managing smart grids by optimizing load distribu-

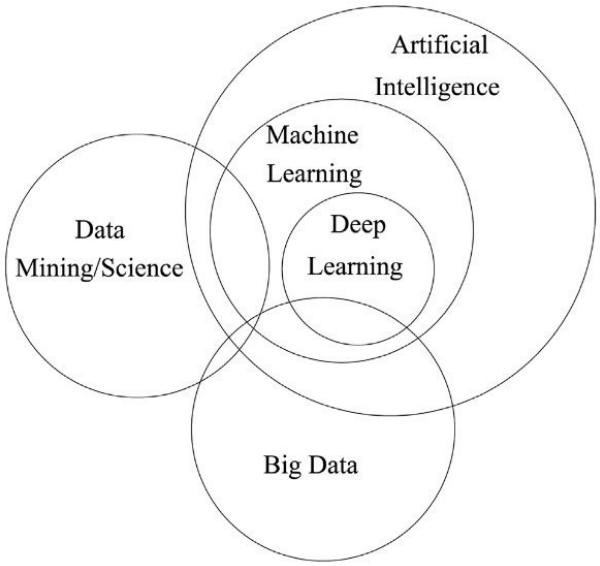


Fig. 5. Relation between Data - Articial Intelligence

tion and boosting energy efficiency through intelligent decision-making.

4) Future Directions

The application of metaheuristic methods in electrical engineering is anticipated to expand as system complexity increases and the demand for efficient solutions grows. Current research is focused on developing hybrid approaches that integrate metaheuristics with other optimization techniques and machine learning models to further enhance performance across diverse applications. In conclusion, metaheuristic methods are essential for addressing complex optimization challenges in electrical engineering. They offer flexible and efficient solutions that traditional methods may find difficult to achieve. Their capacity to navigate uncertainties and escape local optima makes them invaluable tools in contemporary engineering practices.

C. Fuzzy Logic:

Representing 21.3 of AI applications, fuzzy logic is essential for managing uncertainty and imprecision within systems.

D. Expert Systems:

Although their usage is limited to 0.9these systems are less widespread due to their specificity and dependence on precise rules.

V. Conclusion

Artificial intelligence (AI) is transforming the electrical engineering sector by enabling the development of smarter, more efficient, and more resilient systems. By optimizing energy management, equipment maintenance, and the automation of industrial processes, AI contributes to making electrical networks more reliable and sustainable. However, to fully leverage these benefits, it is crucial to overcome the challenges related to the implementation of these technologies and the training of industry professionals. With the rapid evolution of technologies and ongoing research in this field, AI is poised to play a central role in the future of electrical engineering.

References

[1] A. S. Albahri et al., A systematic review of trustworthy artificial intelligence applications in natural disasters , Computers and Electrical Engineering, vol. 118, p. 109409, sept. 2024, doi: 10.1016/j.compeleceng.2024.109409.

[2] S. Zhao, F. Blaabjerg, et H. Wang, An Overview of Artificial Intelligence Applications for Power Electronics , IEEE Trans. Power Electron., vol. 36, no 4, p. 4633-4658, avr. 2021, doi: 10.1109/TPEL.2020.3024914..

[3] S. A. Kalogirou, Artificial intelligence for the modeling and control of combustion processes: a review , Progress in Energy and Combustion Science, vol. 29, no 6, p. 515-566, janv. 2003, doi: 10.1016/S0360-1285(03)00058-3.

[4] B. K. Bose, Expert system, fuzzy logic, and neural network applications in power electronics and motion control , Proc. IEEE, vol. 82, no 8, p. 1303-1323, août 1994, doi: 10.1109/5.301690.

[5] H. Salehi et R. Burgueño, Emerging artificial intelligence methods in structural engineering , Engineering Structures, vol. 171, p. 170-189, sept. 2018, doi: 10.1016/j.engstruct.2018.05.084.

A deep learning-assisted MCDM Service Selection Strategy in UAV Networks

1st Abdessalam Mohammed Hadjkouider
Department of Computer Science and IT, Faculty of Sciences and Technology
Kasdi Merbah University of Ouargla, LINATI Laboratory
Ouargla, Algeria
abdessalam.hadjkouider@univ-ouargla.dz

2nd Abdelhammid Bouazza
Department of Computer Science, LIAM Laboratory
University of M'sila
M'sila, Algeria
abdlhammid.bouazza@univ-msila.dz

3rd Malika Abid
Department of Computer Science and IT, Faculty of Sciences and Technology
Kasdi Merbah University of Ouargla, LINATI Laboratory
Ouargla, Algeria
abid.malika@univ-ouargla.dz

***Abstract*—Heres a rephrased version with "deep learning" replacing "machine learning" (ML):**

"The rapid proliferation of Unmanned Aerial Vehicles (UAVs) has unlocked new opportunities across various application domains. However, selecting optimal UAV services remains a complex challenge that requires a nuanced approach considering multiple criteria. This article introduces a novel framework that integrates deep learning (DL) techniques with multi-criteria decision-making (MCDM) methods to tackle this challenge. The framework evaluates UAV services comprehensively based on critical metrics such as delay, packet loss, throughput, and residual energy, offering a holistic perspective for informed decision-making.

Our findings demonstrate the effectiveness of the proposed approach, achieving a classification accuracy of 99.9% using the RF model. These results highlight the frameworks capability to optimize UAV service delivery, thereby improving overall network performance and user satisfaction. By combining deep learning and MCDM, the framework provides a robust and efficient solution for UAV service selection, addressing the diverse requirements of users and service providers in the dynamic landscape of UAV applications."

***Index Terms*—Service selection; MCDM; TOPSIS; Machine learning; UAVs.**

I. Introduction

Unmanned Aerial Vehicles (UAVs) have proven to be highly effective technologies, offering cost-efficiency and unparalleled mobility that enable their application across diverse domains. Their versatility has transformed numerous industries and sectors, including surveillance, agriculture, logistics, service delivery, and disaster remote sensing.[1].

The recent advancements in Artificial Intelligence (AI) algorithms have enabled swarms of UAVs to play a significant role in establishing robust communication networks, executing diverse tasks, and delivering services [2]. For instance, utilizing UAV swarms for service selection offers considerable potential. However, their deployment faces challenges, particularly energy constraints, which limit the full realization of their capabilities.

To maximize the potential of UAVs in service selection, integrating deep learning (DL) into various decision-making platforms has proven to be a key technology for overcoming limitations and enabling more efficient and effective service selection. However, implementing DL within UAV systems poses challenges, particularly in data management and processing. Overcoming these obstacles is crucial to fully unlocking the benefits of DL for enhancing UAV-based service selection processes

Multi-Criteria Decision-Making (MCDM) plays a vital role in service selection by providing a structured framework for evaluating and comparing alternatives based on multiple criteria simultaneously [3], such as cost, quality, and reliability. When combined with deep learning, MCDM offers a robust approach to analyzing and understanding the strategic interactions between service providers and consumers. This integration has gained significant recognition for its potential to improve service consumer satisfaction.

This study investigates the potential of employing deep learning techniques to enhance the decision-making process, with a particular emphasis on service selection in UAV-based

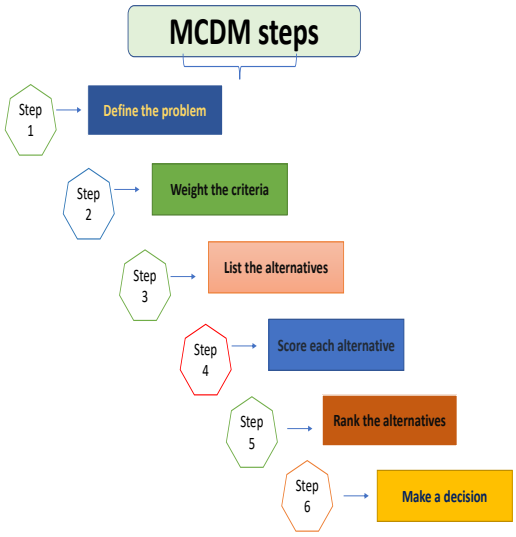


Fig. 1. Multi-criteria Decision-Making steps

communication networks. Labels generated through Multi-Criteria Decision Making (MCDM) are utilized as input for deep learning models. Furthermore, we explore the application of various deep learning architectures to analyze UAV service metrics and predict optimal service configurations. The primary goals are to optimize overall network performance, improve user experience, and provide valuable insights for deploying diverse UAV applications. The key contributions of this paper can be summarized as follows:

- Proposing a decision-making framework for UAVs that incorporates deep learning (DL) labelsclassified as Good Service, Acceptable Service, and Ugly Servicebased on UAV Quality of Service (QoS) metrics and residual energy. This framework enhances the classification accuracy of service selection in dynamic scenarios.
- Conducting extensive simulation experiments in UAV environments to validate the effectiveness of our proposed DL-based decision-making model. The experimental results demonstrate the advantages of coupling deep learning with decision-making analysis to enhance service selection accuracy.

The remainder of this paper is organized afollows: . Section II reviews related works. Section III presents the overall framework and design components of the DL-based decision-making model. Section IV details the extensive experiments conducted. Finally, Section V concludes the paper and outlines future directions.

II. Literature Review

The academic literature has extensively explored methods for service selection, particularly in distributed frameworks like cloud computing, online platforms, and, more recently, in emerging fields such as the Internet of Things (IoT) and UAV-based systems

- **multi-criteria decision-making (MCDM)**

  Multi-Criteria Decision-Making (MCDM) is a field within operations research and decision science focused on making choices when multiple, often conflicting, criteria are involved. It provides a systematic approach to evaluating and selecting alternatives based on their performance across various attributes or factors. MCDM finds applications across diverse fields, such as engineering, business, finance, healthcare, and environmental planning, where decisions must account for multiple considerations simultaneously [4].
- **Service Selection**: Service selection in UAVs involves identifying and choosing specific services or capabilities that UAVs can offer or utilize across a wide range of applications. This process is critical as it directly impacts the UAV's effectiveness in tasks such as surveillance, delivery, agricultural monitoring, disaster response, and communication support. The selection is guided by criteria such as the UAV's payload capacity, flight endurance, communication range, autonomy, security, and responsiveness, as well as the unique requirements of the application, such as real-time data processing or operation in challenging environments [2]. Various approaches, including deep learning, Multi-Criteria Decision Making, and game theory, are employed to optimize the service selection process.
- **Deep learning**: Deep learning, a specialized branch of machine learning, represents a significant advancement in the quest for artificial intelligence, striving to achieve AI's primary goal of mimicking human-like intelligence. It is inspired by the structure and function of the human brain, utilizing artificial neural networks with multiple layers to process and analyze complex data.[**?**]

A. Service Selection Based Machine Learning

This section delivers a succinct review of contemporary literature within the context of UAVs, focusing on the utilization of diverse machine learning algorithms for forecasting optimal service selection.

The authors in [5], integrate ML with digital surface model(DSM) terrain data to automate an air-to-ground (A2G) channel model selection, crucial for precise communication network simulation like Urban Air Mobility (UAM). It autonomously categorizes terrain types and improves data selection accuracy compared to manual methods. Leveraging publicly available DSM data enhances understanding of ML-based service selection in UAV networks. This approach holds significance for analyzing emerging communication networks, including UAV communication.

Khan *et al.* [6] integrate blockchain, machine learning, and auctions for optimal UAV selection. They establish real-time monitoring for enforcing service-level agreements and introduce a novel auction process based on UAV reputation, resulting in a secure and efficient UAV network coverage system.

Manogaran *et al.* [7], present deep learning (DL) for resource allocation in UAV networks, tailored for 6G network demands. This approach aims to enhance efficiency and reliability by adeptly managing resources with advanced machine learning (ML) algorithms. It offers an efficient method for managing communication, resource allocation, and service selection within UAV systems in 6G networks.

The authors in [8], develop a context-aware system based on an ML framework to enhance service selection for individuals with impairments, conducted in two stages: assessing service adequacy based on user goals and profiles, and refining selection considering quality-of-service variables and specific needs. This offers insights into ML's application in adaptive service selection, particularly in UAV-based scenarios.

Tsipi et *et al.* [9], introduce a study that aims to enhance cooperative aerial and device-to-device (D2D) networks in 5G and future wireless systems with non-orthogonal multiple access (NOMA). They employ Deep Neural Networks (DNN) for UAV service selection, particularly focusing on positioning in urban settings.

B. Service Selection based Multi-Criteria Decision Making

This section offers a succinct review of studies that employ MCDM for service selection processes. By utilizing a diverse array of techniques and conceptual frameworks, MCDM is instrumental in addressing a broad spectrum of challenges. This overview underscores the versatility and effectiveness of MCDM methods in facilitating informed and nuanced decision-making processes in the context of service selection, demonstrating their critical role in optimizing outcomes and enhancing operational efficiency.

Gandhi *et al.* [10] introduce an efficient ML for consumers to make informed choices about cloud services. It employs Multiple Criteria Decision Making (MCDM) to evaluate parameters like service performance, cost, reliability, and user requirements, facilitating recommendations for suitable cloud options. authors [11]proposed a methodology for determining the appropriate cloud service by integrating the AHP weighing method with the TOPSIS method, focusing on the quality of services (QoS) as a significant factor for service selection and user satisfaction in cloud computing. The study in [12] focused on selecting the most suitable Unmanned UAVs for defense applications using MCDM methods, specifically integrating the Analytic Hierarchy Process AHP and the TOPSIS. It outlines the importance of UAVs in military and civilian contexts, details the methodology for evaluating and ranking UAV alternatives based on various performance criteria, and concludes with selecting the best UAV alternative. The study's findings, validated through sensitivity analysis, aim to assist decision-makers in the defense sector in making informed UAV selections. Hamurcuet *et al.* [12] presents a study on the selection of UAVs using Fuzzy MCDM analysis. It addresses the complex decision-making process involved in UAV selection, considering various factors like payload capacity, maximum speed, endurance, altitude, avionics systems, price, economic life, and maximum range. The methodology employs standard fuzzy set techniques to enable decision-makers to use linguistic terms, enhancing the decision-making process's practicality and interpretability.

III. Methodology

This section presents the proposed design for selecting UAV services using MCDM-deep learning, enabling users to choose UAV service providers based on multiple criteria.

A. Overall framework

The overall structure of our proposed concept is depicted in Fig.2. When a user requests a service from UAV-based providers, we employ Multi-Criteria Decision-Making (MCDM) techniques to identify the most efficient UAV service by evaluating factors such as delay, packet loss, throughput, and energy efficiency.

This deep learning-driven approach aims to predict the optimal UAV service provider for each user request, framing the decision-making process as a categorical prediction task.

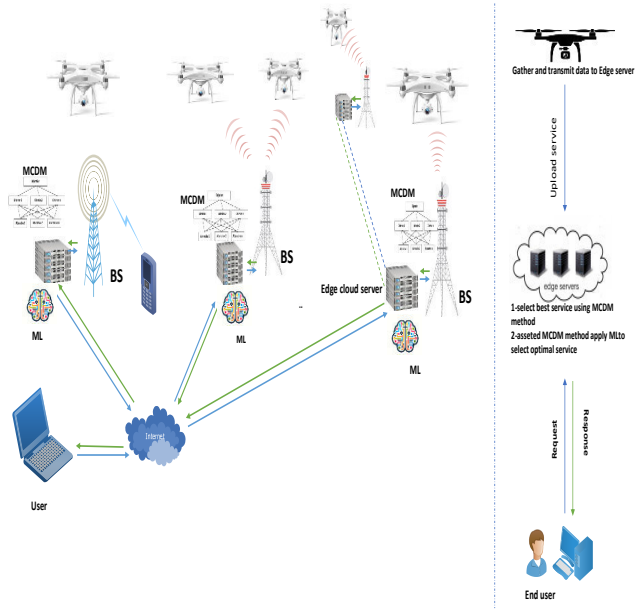


Fig. 2. System Model

Heres the rephrased version with "machine learning" replaced by "deep learning":

—

B. Design Components

UAV Service Providers: UAVs provide various services to remote users in mobile environments, including delivery, communication relays, data collection, healthcare support, and infrastructure inspections. To enhance their limited computational capabilities, these UAVs rely on edge computing.

Edge Computing Servers: Deep learning and Multi-Criteria Decision-Making (MCDM) algorithms are integrated within edge computing servers to meet the stringent requirements of low latency and high computational efficiency in

service selection applications. Accounting for the computational constraints of UAVs and end-user devices is critical to enhancing response times and ensuring the optimal selection of services and UAVs for completing tasks.

User: The process begins with users submitting service requests to UAV-based service providers via an application or user interface. These requests can range across a variety of services, such as data collection, surveillance, delivery, and communication relays, depending on the application's purpose. Typically, these requests include detailed specifications and requirements.

IV. Performance Evaluation

This section outlines a concise summary of the system model developed in the initial phase. Subsequently, we discuss the dataset chosen for evaluation, followed by an overview of the performance metrics utilized. Finally, we present a comparative analysis of the various classifiers under consideration.

A. Proposed Model

Our service selection approach utilizes the TOPSIS method, where higher scores represent better alignment with the desired characteristics determined by weighted criteria. This approach offers a comprehensive framework for evaluating and selecting the most suitable service UAVs by balancing multiple performance metrics, incorporating both benefits and costs associated with each UAV's performance features. Criteria such as delay, packet loss, throughput, and residual energy are considered in the selection process.

The key steps for identifying the optimal service using TOPSIS are outlined below [13].

Define the Criteria and Alternatives: Delay, packet loss, throughput, and residual energy are considered key criteria in the decision-making process, with UAVs serving as the available alternatives.

Create the Decision Matrix: Construct a decision matrix with alternatives represented in rows j and criteria in columns i. by inputting values r_{ij} that indicate how well each alternative satisfies each criterion.

Normalize the Decision Matrix To make all the criterion scales equivalent and comparable, we normalize the decision matrix. A common method is the square root of the sum of squares method.

For a value r_{ij} in the matrix, the normalized value x_{ij} is calculated as:

$$x_{ij} = \frac{r_{ij}}{\sqrt{\sum_{k=1}^{t} r_{kj}^2}}$$

where t is the number of UAVs.

Identify the Optimal and Pessimistic Solutions: Determine the optimal and worst-case scenarios for each criterion. The optimal solution A_j^* increases benefits and minimizes costs, whereas the negative-ideal solution A_j^- accomplishes the opposite.

Distance Calculation

- The Euclidean distance from each alternative to the ideal (D_i^*) and negative-ideal (D_i^-) solutions are calculated as:

$$D_i^* = \sqrt{\sum_{j=1}^{n} w_j (x_{ij} - A_j^*)^2}$$

$$D_i^- = \sqrt{\sum_{j=1}^{n} w_j (x_{ij} - A_j^-)^2}$$

- where:
  - w_j is the weight of the j^{th} metric.
  - x_{ij} is the value of the j^{th} metric for the i^{th} alternative.
  - A_j^* is the value of the j^{th} metric in Positive Ideal Solution
  - A_j^- is the value of the j^{th} metric in Negative Ideal Solution.
  - n is the total number of metrics.

- **TOPSIS Score** The overall score for each alternative is calculated as:

$$\text{TOPSIS Score}_i = \frac{D_i^-}{D_i^* + D_i^-}$$

Ranking Alternatives (UAVs, in this context) are ranked based on their TOPSIS scores, where higher scores indicate better performance. These scores range from 0 to 1, with values closer to 1 representing UAV service providers that are nearer to the ideal solution and, therefore, more desirable.

B. Dataset description

TABLE I presents a synthetic dataset created by gathering various Quality of Service (QoS) metrics, including throughput, end-to-end latency, packet loss ratio, and residual energy. This dataset is utilized to evaluate, analyze, and select the most efficient UAV services.

The approach integrates parameters linked to QoS requirements, with two additional columns added to the dataset: one for computing the TOPSIS score based on QoS metrics and another for assigning labels categorized as "good," "acceptable," or "ugly." The dataset is divided into two parts, with eighty used for training and twenty reserved for performance evaluation and validation.

To identify suitable UAVs, well-established machine learning algorithms are applied, including Naive Bayes, Support Vector Machine (SVM), Random Forest, and Decision Tree, each configured with specific hyperparameters.

C. Performance metrics

Performance measures in machine learning are crucial for assessing and contrasting the efficiency of models. Common measurements consist of::

- **Accuracy:** The percentage of true outcomes (true positives and true negatives) across all instances analyzed.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}}$$

TABLE I
DATASET SNAPSHOT

UAV ID	Delay	Throughput	Packet Loss	Residual Energy	TOPSIS Score	**Label**
uav243	0.035480375	0.170560013	0.234953558	0.455321561	0.94688933	good
uav266	0.035495611	0.173068987	0.234014272	0.449809255	0.940813011	good
uav958	0.023402112	0.176412665	0.233716203	0.45780353	0.931130942	good
uav269	0.027363339	0.156877037	0.233967458	0.444200239	0.913747892	good
...	...	...	...	...	....	.....
uav142	0.014534155	0.069295486	0.232782227	0.376362677	0.719784624	acceptable
uav428	0.027256444	0.17638152	0.233136388	0.328186786	0.719362124	acceptable
uav3	0.010661707	0.084059237	0.2328577	0.366410276	0.71904528	acceptable
uav422	0.007408826	0.017942547	0.230851247	0.430471109	0.715827948	acceptable
...	...	...	...	...	....	.....
uav235	0.019307369	0.000758012	0.233002381	0.014693878	0.037509537	ugly
uav947	0.009424711	0.001031794	0.233676312	0.015712657	0.031937793	ugly
uav937	0.016278053	0.002582798	0.231467119	0.013185648	0.031837863	ugly
uav618	0.010021295	0.009398879	0.233172693	0.008905042	0.026547942	ugly

- **Precision:**The ratio of true positives to the sum of true and false positives, which reflects the model's accuracy in predicting positive classes.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **Recall:** The metric quantifies the model's capability to identify positive instances by dividing the number of true positives by the sum of true positives and false negatives.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- **F1 Score:** Achieving a balance between precision and recall through the harmonic mean, particularly in cases of irregular class distributions.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- Where:
  -True positives (TP) are those that accurately forecast positive observations.
  -True negatives (TN) are observations that were predicted to be negative.
  -False positives (FP) mean that positive observations are predicted inaccurately.
  False negatives (FN) are predictions of negative observations that are inaccurate.
  Each metric provides different insights into the model's performance, emphasizing the importance of selecting appropriate metrics based on the specific objectives and challenges of the machine learning task.

D. Experimental Setup and Implementation

In this paper, we employed an MLP as the foundational deep learning algorithm, consisting of three hidden layers with 64, 64, and 32 neurons, respectively, utilizing the ReLU activation function. The model was trained over 100 epochs with a batch size of 64, using the Adam optimizer and categorical cross-entropy loss. The data was split into 60% for training, 20% for validation, and 20% for testing. A comparative analysis was conducted against traditional machine learning algorithms, including Naïve Bayes and SVM, using their basic hyperparameter configurations.

E. Comparative Performance Analysis of Classification Algorithms for UAV QoS Assessment

Table II presents a comprehensive comparison of performance metrics, including execution time (E-time), for the various classification algorithms evaluated in this study. Accuracy is emphasized as a fundamental measure for evaluating the overall performance of classifiers, providing a broad indication of their ability to correctly categorize UAVs into predefined quality categories: good, acceptable, and poor. These categorizations are determined based on UAV Quality of Service (QoS) parameters.

Both Table II and figure 3 depict the comparative performance of the evaluated algorithms, highlighting critical

TABLE II
THE OVERALL PERFORMANCE OF THE VARIOUS CLASSIFICATION PROCESSES..

Algo	Acc	Prec	Recall	F1-Score	E-time (S)
MLP	0.99	0.99	0.99	0.99	0.11
Naïve B	0.97	0.97	0.97	0.0.97	0.04
SVM	0.96	0.96	0.96	0.96	0.13

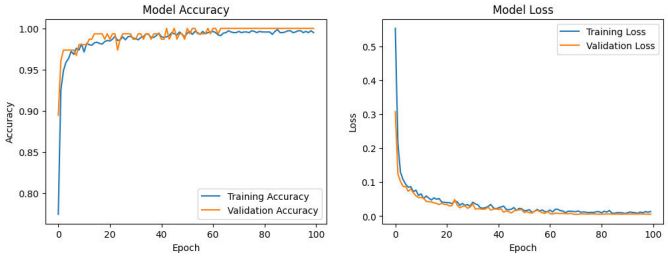


Fig. 4. Training and Validation Performance Analysis of the MLP Model

metrics. As demonstrated, the MLP algorithm outperforms its counterparts across all metrics, achieving near-perfect results with a minimal execution time of 0.11 seconds. The Naïve Bayes algorithm follows closely, showcasing strong performance while achieving the lowest execution time of 0.04 seconds. In contrast, the SVM exhibits comparatively lower performance across all measures, revealing its limitations in this particular context. Figure 4 illustrates that the model is robust and well-suited for the given classification task, demonstrating efficient optimization of both hyperparameters and architecture. The findings underscore the superior effectiveness of the MLP algorithm in accurately classifying UAV QoS categories. Its exceptional precision and recall make it a strong candidate for real-world applications where robust classification performance and computational efficiency are critical. Moreover, the MLP algorithm demonstrates a significant advantage in handling large datasets compared to traditional machine learning approaches, underscoring its scalability and suitability for complex networks with extensive data.

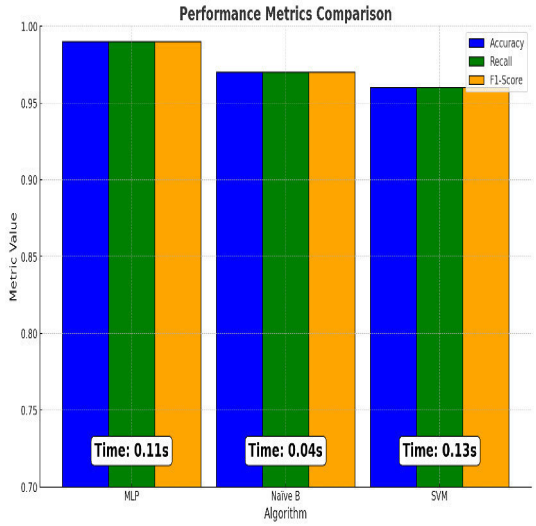


Fig. 3. Comparison between the different classifiers

V. CONCLUSIONS AND FUTURE WORKS

This paper emphasizes the effectiveness of integrating deep learning with Multi-Criteria Decision-Making (MCDM) methods, specifically TOPSIS, to select optimal UAV services based on metrics such as delay, packet loss, throughput, and residual energy. The study aims to enhance the decision-making process for UAV service delivery, ultimately improving user satisfaction. Experimental results show that incorporating MCDM labels into deep learning models for training and UAV service prediction achieves superior performance in terms of accuracy, error rate, precision, recall, and F-score metrics.

As future work, the focus will shift toward enhancing security measures to address cybersecurity threats, particularly phishing attacks [14]. This research will explore the development of robust strategies and technologies to strengthen systems against malicious activities and safeguard sensitive information.

REFERENCES

[1] Y. Sahraoui, C. A. Kerrache, M. Amadeo, A. M. Vegni, A. Korichi, J. Nebhen, and M. Imran, "A cooperative crowdsensing system based on flying and ground vehicles to control respiratory viral disease outbreaks," *Ad Hoc Networks*, vol. 124, p. 102699, 2022.

[2] A. M. Hadjkouider, C. A. Kerrache, A. Korichi, Y. Sahraoui, and C. T. Calafate, "Stackelberg game approach for service selection in uav networks," *Sensors*, vol. 23, no. 9, p. 4220, 2023.

[3] G. P. Bhole and T. Deshmukh, "Multi-criteria decision making (mcdm) methods and its applications," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 6, no. 5, pp. 899–915, 2018.

[4] A. M. Farid and M. Mouhoub, "A multi-objective approach for unmanned aerial vehicle mapping," in *2023 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2023, pp. 257–264.

[5] Y.-E. Kang and Y.-H. Jung, "Machine learning-based air-to-ground channel model selection method for uav communications using digital surface model data," *Sensors*, vol. 22, no. 23, p. 9234, 2022.

[6] A. S. Khan, G. Chen, Y. Rahulamathavan, G. Zheng, B. Assadhan, and S. Lambotharan, "Trusted uav network coverage using blockchain, machine learning, and auction mechanisms," *IEEE Access*, vol. 8, pp. 118 219–118 234, 2020.

[7] G. Manogaran, J. Ngangmeni, J. Stewart, D. B. Rawat, and T. N. Nguyen, "Deep learning-based concurrent resource allocation for enhancing service response in secure 6g network-in-box users using iiot," *IEEE Internet of Things Journal*, 2021.

[8] A. Namoun, A. A. Abi Sen, A. Tufail, A. Alshanqiti, W. Nawaz, and O. BenRhouma, "A two-phase machine learning framework for context-aware service selection to empower people with disabilities," *Sensors*, vol. 22, no. 14, p. 5142, 2022.

[9] L. Tsipi, M. Karavolos, P. S. Bithas, and D. Vouyioukas, "Machine learning-based methods for enhancement of uav-noma and d2d cooperative networks," *Sensors*, vol. 23, no. 6, p. 3014, 2023.

[10] U. Gandhi, A. Bothera, N. Garg, Neeraj, and I. Gupta, "A machine learning-based approach for efficient cloud service selection," in *International Advanced Computing Conference*. Springer, 2021, pp. 626–632.

[11] R. R. Kumar and C. Kumar, "A multi criteria decision making method for cloud service selection and ranking," *International Journal of Ambient Computing and Intelligence (IJACI)*, vol. 9, no. 3, pp. 1–14, 2018.

[12] M. Hamurcu and T. Eren, "Selection of unmanned aerial vehicles by using multicriteria decision-making for defence," *Journal of Mathematics*, vol. 2020, pp. 1–11, 2020.

[13] S. Chakraborty, "Topsis and modified topsis: A comparative analysis," *Decision Analytics Journal*, vol. 2, p. 100021, 2022.

[14] Y. Sahraoui, C. A. Kerrache, A. Korichi, A. M. Vegni, and M. Amadeo, "Learnphi: A real-time learning model for early prediction of phishing attacks in iov," in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2022, pp. 252–255.

3bit Quantized On-device Recommendation System

1st Ramzi Khantouchi
Computer Science and Applied Mathematics Laboratory
Chadli Bendjedid University
El Tarf 36000, Algeria
r.khantouchi@univ-eltarf.dz

2nd Ibtissem Gasmi
Computer Science and Applied Mathematics Laboratory
Chadli Bendjedid University
El Tarf 36000, Algeria
gasmi-ibtissem@univ-eltarf.dz

3rd Abderaouf Bahi
Computer Science and Applied Mathematics Laboratory
Chadli Bendjedid University
El Tarf 36000, Algeria
a.bahi@univ-eltarf.dz

***Abstract*—Recommender systems are integral to modern e-commerce, yet their deployment on resource-constrained devices remains a challenge due to the high computational and memory demands of deep learning models. This paper presents a novel 3-bit fixed-point quantized Multi-Layer Perceptron (MLP) recommendation system, specifically designed for on-device usage. Leveraging a training-aware quantization technique, this approach reduces model parameters to 3-bit precision, significantly decreasing memory and computational overhead while maintaining performance. The proposed system is evaluated using the Amazon Beauty dataset, demonstrating competitive accuracy across metrics such as precision, recall, and F1-score, with a substantial reduction in resource requirements. These results underscore the potential of the quantized MLP model for enabling real-time recommendations on edge devices such as smartphones and IoT systems, paving the way for efficient, privacy-preserving recommender systems.**

***Index Terms*—On Device Recommender System, Quantization, Recommendation System, E-commerce**

I. Introduction

Recommender systems have gained considerable attention due to their importance in many sectors such as e-commerce [1]. Today's recommender systems benefit from huge deep neural networks such as large language models (llms) [2] to capture user preferences to recommend items that meets the user's taste. Despite their capabilities to enhance recommendations quality, those models suffer from several issues such as high computational complexity that leads to significant power consumption and their huge size which requires significant memory consumption. They also face privacy concerns because they are fueled by massive user behavior data that sent through the internet. Recently, on-device machine learning [3] has attracted the attention of the research community because it enables models to run on limited resource-constrained devices such as mobile phones. Since the principle of on-device machine learning is well-aligned with the need for low-cost and privacy-preserving recommender systems, there has been a growing interest towards on-device recommendation. Authors in [4] Proposed a method for learning flexible item embeddings, enabling recommendation models to be adapted to arbitrary device-specific memory constraints without the need for retraining. Authors in [5] proposed LLRec, a lightweight model for next Point-of-Interest (POI) recommendation, leveraging tensor-train factorization to achieve efficiency and scalability. Authors in [6] Proposes utilizing the user's real-time intentions to generate on-device recommendations by employing a sequence- and context-aware algorithm to encode user intentions. Prediction is then performed using a neighborhood search method combined with a sequence-matching algorithm. Authors in [7] introduces an expert selection strategy and refines the ranking-based sampling approach to facilitate knowledge transfer. Authors in [8] employs model pruning and embedding sparsification techniques to enable model training on mobile devices and fine-tunes the model using local user data, achieving comparable performance in sequential recommendation while reducing parameters by a factor of 10. Authors in [9] represents items and users using a set of binary vectors paired with a sparse weight vector, and introduces an integer weight approximation scheme to enhance the method's computational efficiency.

In this paper, we proposed 3-bit fixed-point quantized MLP recommendation model which can be used on edge devices such as smartphones. We used a training-aware quantization technique proposed by [10] to quantize the model parameters into 3-bit fixed-point precision to make it suitable for limited resource-constrained devices and maintains its accuracy.

The main contributions of this study include the following:

- Propose a 3-bit fixed-point precision quantized MLP recommender system.
- Employ a training-aware quantization technique to quantize the recommendation model.
- Evaluate the performance of the proposed model using an ecommerce benchmark dataset.
- Use multiple evaluation metrics such as precision, recall, F1 score, MAE, RMSE and MSE to asses the performance of the proposed approach.

The code source of this work will be available at https://github.com/ramzikhantouchi/3-bit_quantized_On-device_recommender_sysetm after paper acceptance.

II. Approach

We propose a recommendation system based on a Multi-Layer Perceptron (MLP) architecture that balances high performance with computational and memory efficiency. Recognizing the limitations of deploying neural networks on resource-constrained devices such as IoT systems (e.g., smartwatches, drones, or smart home devices), we adopt a training-aware quantization technique inspired by the work of [10]. This technique significantly reduces the computational complexity and energy consumption of the model, making it practical for low-resource environments. The quantization-aware training method proposed by [10] focuses on reducing the precision of the model's weights and biases, converting them from a 32-bit floating-point representation to an 8-bit fixed-point format. This change reduces the memory footprint, accelerates computations, and lowers energy requirements, making the model suitable for deployment on devices with constrained resources.

In a neural network, matrix–vector multiplication operations dominate the computational cost. The weights W and biases b are traditionally stored in high-precision formats (16-bit or 32-bit), resulting in significant energy and memory costs for Multiply–Accumulate (MAC) operations during inference. By quantizing these parameters to fixed-point integers, the memory requirements are reduced, and the efficiency of MAC operations is enhanced. The technique involves approximating floating-point weights and biases as integer values scaled by a scalar factor. To optimize hardware efficiency, the scalar factor is restricted to powers of two. This restriction ensures that division operations, which are computationally expensive, are replaced by simple bit-shifting operations, reducing computational overhead while maintaining numerical stability. It applies quantization iteratively, layer by layer, to minimize the noise introduced during rounding. Starting with the first layer, its weights and biases are quantized and fixed, and the model is retrained to update the parameters of subsequent layers. Retraining ensures that errors caused by quantization in earlier layers are corrected by later layers. This iterative process continues until all layers are quantized, maintaining a balance between performance and efficiency. The proposed MLP model incorporates the ReLU (Equation (1)) activation function within its hidden layers to introduce non-linearity and enhance the network's ability to model complex relationships. The output layer employs a sigmoid (Equation (2)) activation function to generate probabilities indicating whether a user is likely to prefer an item. We use Adam optimzizer and the binary cross-entropy loss function to train our model. By leveraging the quantization-aware training technique from [10], the proposed system achieves substantial improvements in efficiency. Specifically, memory consumption is reduced by more than a factor of 10, and computational complexity is decreased by a factor of 16. These improvements enable real-time recommendation tasks on resource-constrained devices without significant loss of accuracy.

$$ReLU(x) = \max(0, x) \tag{1}$$

$$Sigmoid(x) = \frac{1}{1 + e^{-}x} \tag{2}$$

III. Experiments

A. Dataset

Amazon Beauty is an online reviews and ratings dataset from Amazon reviews which are widely recognized benchmarks for assessing recommendation algorithms. This dataset contains a wide range of beauty and personal care products, including skincare, haircare, and cosmetics items, and includes detailed user-generated content such as reviews, star ratings, and feedback about product efficacy.

B. Evalutation Metrics

To evaluate the performance of recommendation systems, standard metrics such as Recall (Equation (3)), Precision (Equation (4)), and F1 Score (Equation (5)) are frequently utilized to quantify the relevance and accuracy of the recommended items. Beyond these ranking-based measures, error metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) are employed to gauge the accuracy of the predicted relevance scores. MSE (Equation (6)) quantifies the average squared deviation between predicted and actual relevance values, with a greater emphasis on larger errors. RMSE (Equation (7)) provides a more interpretable measure by taking the square root of MSE, facilitating model comparisons. Meanwhile, MAE (Equation (8)) evaluates the average absolute deviation between predictions and ground truth, offering a straightforward representation of prediction error. Together, these metrics enable a comprehensive assessment of the recommendation model's effectiveness in ranking items.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{7}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{8}$$

TP, TN, FP, FN, y_i, $\hat{y}_i$ and N represent True Positive, True Negative, False Positive, and False Negative, the actual relevance score, the predicted relevance score and the total number of observations respectively.

C. Results

The performance comparison between the MLP model and its quantized version is presented in Figures 1 and 2, and summarized in Table I. The results show that the quantized version of the recommender system demonstrates competitive performance while offering substantial advantages in terms of memory and computational efficiency.

In terms of prediction accuracy, both models exhibit similar results across the metrics. The Precision of the quantized MLP (0.7776) slightly surpasses that of the original MLP (0.7706), while the Recall of the quantized model (0.9785) is marginally lower than that of the MLP model (0.9996). The F1-Score of the quantized MLP (0.8666) remains nearly identical to the MLP model (0.8702), indicating that the quantization process does not significantly compromise the overall accuracy of the model.

For error metrics, the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) of the quantized model are slightly higher (0.2321 and 0.4818, respectively) compared to the MLP model (0.2296 and 0.4791, respectively). Similarly, the Mean Absolute Error (MAE) of the quantized model (0.2321) is marginally higher than that of the MLP model (0.2296). However, these differences are minimal and do not significantly affect the overall recommendation quality.

Despite these minor variations in performance, the quantized recommender system excels in memory and computational efficiency. The quantized model achieves more than a 10x reduction in memory consumption compared to the original MLP model, enabling it to run efficiently on devices with limited resources such as mobile phones. Moreover, the quantized version reduces computational complexity by a factor of 16, resulting in faster inference times and lower energy consumption. This makes the quantized MLP an ideal choice for resource-constrained environments, where efficiency is critical without sacrificing significant performance in recommendation tasks.

These findings demonstrate that the quantized recommender system provides a memory- and energy-efficient solution for real-time recommendation tasks, offering a practical trade-off between model accuracy and resource usage.

TABLE I
COMPARISON OF MLP AND QUANTIZED MLP RECOMMENDERS

Metric	MLP recommender	Quantized MLP recommender
Precision	0.7706	0.7776
Recall	0.9996	0.9785
F1-Score	0.8702	0.8666
MSE	0.2296	0.2321
RMSE	0.4791	0.4818
MAE	0.2296	0.2321

IV. CONCLUSION

This work introduces a 3-bit fixed-point quantized MLP recommendation system, tailored for deployment on resource-constrained devices. By employing the quantization-aware training technique proposed by [10], the system effectively reduces memory consumption by more and 10x and computational complexity by a factor of 16. Experimental results using the Amazon Beauty dataset confirm that the quantized model delivers performance comparable to its full-precision counterpart, with minimal loss in metrics such as precision, recall, and F1-score. These findings highlight the practicality of deploying efficient recommendation systems on edge devices without sacrificing recommendation quality.

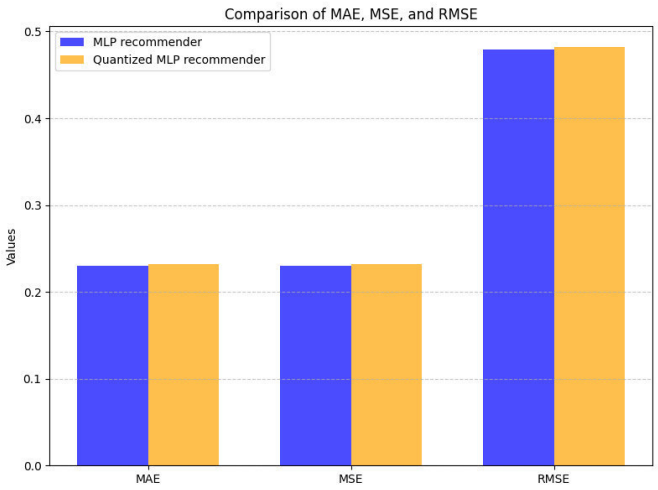


Fig. 1. Comparison of MAE, MSE, and RMSE between MLP and Quantized MLP recommenders.

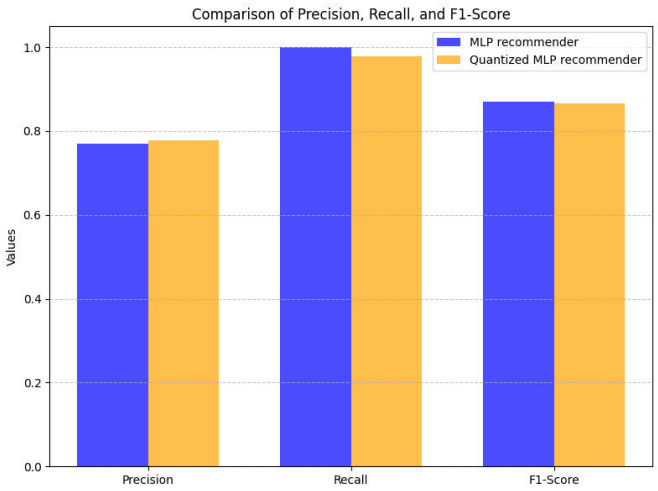


Fig. 2. Comparison of Precision, Recall, and F1-Score between MLP and Quantized MLP recommenders.

REFERENCES

[1] G. Zhou, X. Zhu, C. Song, Y. Fan, H. Zhu, X. Ma, Y. Yan, J. Jin, H. Li, and K. Gai, "Deep interest network for click-through rate prediction," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 1059–1068.

[2] R. Khantouchi, I. Gasmi, and A. Bahi, "Enhancing recommendation quality in sparse scenarios using large language models (llms)," 2024.

FT/Boumerdes/NCASEE-24-Conference

[3] S. Dhar, J. Guo, J. Liu, S. Tripathi, U. Kurup, and M. Shah, "A survey of on-device machine learning: An algorithms and learning theory perspective," *ACM Transactions on Internet of Things*, vol. 2, no. 3, pp. 1–49, 2021.

[4] T. Chen, H. Yin, Y. Zheng, Z. Huang, Y. Wang, and M. Wang, "Learning elastic embeddings for customizing on-device recommenders," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 138–147.

[5] Q. Wang, H. Yin, T. Chen, Z. Huang, H. Wang, Y. Zhao, and N. Q. Viet Hung, "Next point-of-interest recommendation on resource-constrained mobile devices," in *Proceedings of the Web conference 2020*, 2020, pp. 906–916.

[6] B. M. Changmai, D. Nagaraju, D. P. Mohanty, K. Singh, K. Bansal, and S. Moharana, "On-device user intent prediction for context and sequence aware recommendation," *arXiv preprint arXiv:1909.12756*, 2019.

[7] S. Kang, J. Hwang, W. Kweon, and H. Yu, "De-rrd: A knowledge distillation framework for recommender system," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 605–614.

[8] J. Han, Y. Ma, Q. Mei, and X. Liu, "Deeprec: On-device deep learning for privacy-preserving sequential recommendation in mobile commerce," in *Proceedings of the Web Conference 2021*, 2021, pp. 900–911.

[9] C. Liu, T. Lu, X. Wang, Z. Cheng, J. Sun, and S. C. Hoi, "Compositional coding for collaborative filtering," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2019, pp. 145–154.

[10] R. Khantouchi, I. Gasmi, and M. A. Ferrag, "Eye-net: A low-complexity distributed denial of service attack-detection system based on multilayer perceptron," *Journal of Sensor and Actuator Networks*, vol. 13, no. 4, p. 45, 2024.

AlexNet-Based Feature Extraction Combined with a KNN Classifier for Facial Expression Recognition

Hamza Roubhi
ETA Laboratory, University
Mohamed El Bachir El Ibrahimi of
Bordj Bou Arreridj, Algeria
hamza.roubhi@univ-bba.dz

Abdenour Hacine-Gharbi
LMSE Laboratory, University
Mohamed El Bachir El Ibrahimi of
Bordj Bou Arreridj, Algeria
abdenour.hacinegharbi@univ-bba.dz

Fateh Ghazali
LMSE Laboratory, University
Mohamed El Bachir El Ibrahimi of
Bordj Bou Arreridj, Algeria
laboauto@hotmail.fr

***Abstract*— Facial Expression Recognition (FER) involves identifying an individual's emotional state and plays a crucial role in areas such as human-computer interaction, healthcare, and marketing. Numerous studies have employed hand-crafted feature extraction techniques, such as the Discrete Wavelet Transform (DWT). With advancements in deep learning, particularly Convolutional Neural Networks (CNNs), researchers have used features derived from the convolutional or Fully Connected (FC) layers of CNNs trained for visual recognition tasks in other applications. In this work, a pre-trained deep CNN, AlexNet, is employed for feature extraction instead of traditional hand-crafted features, followed by a K-Nearest Neighbors (KNN) classifier using various distance metrics, including Euclidean, Cityblock, Correlation, and Cosine. The experiments conducted on the benchmark CK+ dataset demonstrate the effectiveness of the proposed system, achieving a recognition rate of 96.32% using the FC7 layer with the KNN classifier at Cosine distance metric.**

Keywords*— *Facial Expression Recognition, Feature Extraction, KNN, AlexNet, Deep Learning.

I. INTRODUCTION

Facial expression is an essential means of communication through which people convey their feelings. It is crucial for recognizing and analyzing an individual's emotional states. Identifying emotional states via facial expressions is gaining significance owing to its many applications, including human-computer interaction [1], criminal investigation [2], and virtual reality [3].

Feature extraction is an essential phase in the image classification process. The extracted features can significantly impact the accuracy of classification. Feature extraction can be performed using either handcrafted techniques or deep learning approaches based on CNN [4].

In [5], the authors presented a new feature distribution by combining shape and texture features, analyzing facial regions, and identifying critical regions for facial expression recognition. Shape and texture features considered include Local Phase Quantization (LPQ), Local Binary Pattern (LBP), and Histogram of Oriented Gradients (HOG). Multiclass Support Vector Machine (MSVM) is used for one-on-one classification. CK+, KDEF, and JAFFE benchmark facial expression datasets are used for the proposed study. This study achieves a recognition rate of 94.2% on CK+ and 93.7% on KDEF.

Qin et al. In [6] proposed a Gabor wavelet transform integrated with a convolutional neural network. Initially, they conducted preprocessing on the expression images, then extracted key frames, and used the Gabor wavelet transform. Finally, the designed 2-channel CNN for classification, the experiment was conducted on CK+ dataset and attained an accuracy of 96.81%.

Bendjillali et al. In [7] suggested a Facial Expression Recognition system using discrete wavelet transform (DWT) and convolutional neural networks (CNN). The Viola–Jones technique was used for face detection, followed by the application of contrast limited adaptive histogram equalization (CLAHE) for image improvement. Next, the facial features were extracted using DWT to train the CNN network. The model achieved an accuracy of 96.46% using CK+ dataset and 98.43 % using JAFFE dataset.

Using deep neural networks for feature extraction, particularly AlexNet, is regarded as one of the simplest and fastest methods, as it requires just a single pass through the data. A robust and real-time facial expression recognition model can be achieved by integrating the AlexNet pre-trained feature extractor with a straightforward classifier like KNN.

The main contribution of this proposed work lies in the development of an FER system using a pre-trained AlexNet model with different fully connected layers as a feature extractor, combined with a KNN classifier employing various distance metrics.

The remainder of the paper is structured as follows: Section II describes the related work, Section III introduces our proposed FER system, Section IV presents the experimental results along with a discussion, and Section V concludes the paper.

II. RELATED WORK

In [8], the authors proposed a FER framework to enhance system performance, consisting of three stages. The first stage involves feature extraction using three descriptors derived from the Discrete Wavelet Transform (DWT). In the second stage, feature selection is performed using a wrapper approach, followed by classification using an SVM classifier in the final stage. This framework achieved an accuracy of 87.76% using the CK+ dataset and 89.66% using the JAFFE dataset.

In contrast, we aim to use deep neural networks for feature extraction instead of hand-crafted methods like DWT. Specifically, AlexNet, known for its simplicity and efficiency due to requiring only a single pass through the data, This technique is chosen for feature extraction by combining the AlexNet pre-trained feature extractor with a straightforward classifier such as K-NN, a robust and real-time facial expression recognition model can be developed.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

III. PROPOSED FER SYSTEM

The design of a FER system involves a training phase to model various emotion categories and a testing phase to classify test images and evaluate the system's performance. Both phases include a feature extraction process that converts images into feature vectors, which are then input into the K-NN classifier.

Figure 1 presents an overview of the proposed FER system, consisting of a series of blocks. In the following sections, we will describe the underlying principles of our proposed FER system and provide a detailed explanation of each block.

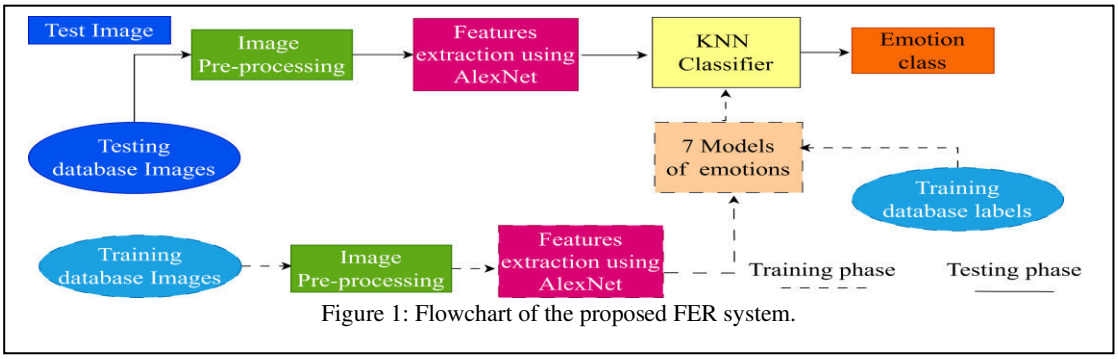


Figure 1: Flowchart of the proposed FER system.

A. *Image Preprocessing:*

Prior feature extraction phase, all face images in the CK+ dataset are pre-processed, with the facial region being detected using the Haar cascade classifier introduced by Viola-Jones [9]. Subsequently, these latter images have been transformed into a format compatible with the AlexNet architecture and resized to dimensions of 227 × 227 pixels.

B. *Feature extraction using AlexNet*

AlexNet is a type of convolutional neural network developed by Alex Krizhevsky et al [10], it has been trained using the ImageNet dataset and has a total of eight layers containing 62.3 million learnable parameters. The model architecture consists of five convolutional layers with max pooling, followed by three fully connected layers and two dropout layers. The ReLU activation function is applied to all layers except the final output layer, which uses the Softmax activation function, as illustrated in Figure 2 [11].

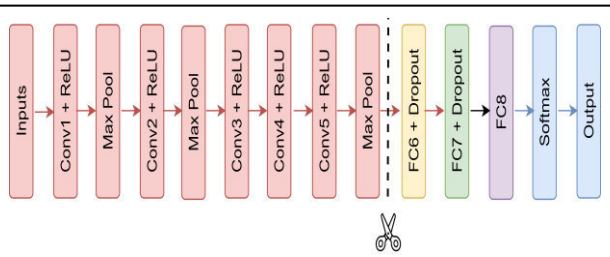


Figure 2: Feature Extraction Using the AlexNet Pretrained Model

In this study, AlexNet was used for feature extraction phase. By leveraging the transfer learning approach, the need to re-train AlexNet was eliminated within the context of the FER task, resulting in significant time savings. The fully connected layers of the AlexNet model, specifically Fc6, Fc7, and Fc8, were employed for feature extraction. Additionally, these three fully connected layers were combined. The number of features in the output vector for each fully connected layer is presented in Table 1.

TABLE1: NUMBER OF FEATURE FOR FULLY CONNECTED LAYERS IN ALEXNET

Fully Connected Layers	Number of Features
FC6	4096
FC7	4096
FC8	1000
Combined Layers	9192

C. *KNN classifier*

K-NN [12] is a non-parametric pattern classification technique that assigns a class to data points based on their proximity to the 'k' closest neighbors within the training dataset. Instead of making assumptions about data structure, it uses features and labels. Classifying a new point involves looking at the categories of its 'k' neighbor points, generally quantified using Euclidean distance and assigning the most frequent category. The value of 'k' determines how many neighboring points influence the classification [13].

The basic K-NN classifier procedure is outlined as follows:

- Calculate the distance between the unknown test feature vectors $A\ (A_1, A_2, \dots A_d)$ and all the known training feature vectors $B\ (B_1, B_2, \dots B_d)$ using the four following metric distances:

$$distance(A,B) = \begin{cases} \sqrt{\sum_{i=1}^{d} (A_i - B_i)^2} & for\ Euclidean \\ \sum_{i=1}^{d} |A_i - B_i| & for\ Cityblock \\ 1 - \frac{\sum_{i=1}^{d} A_i B_i}{\sqrt{\sum_{i=1}^{d} A_i^2}\sqrt{\sum_{i=1}^{d} B_i^2}} & for\ Cosine \\ \frac{1}{2}\left(1 - \frac{\sum_{i=1}^{d} (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^{d} (A_i - \bar{A})^2}\sqrt{\sum_{i=1}^{d} (B_i - \bar{B})^2}}\right) & for\ Correlation \end{cases} \quad (1)$$

where d denotes the number of features, $\bar{A}$ and $\bar{B}$ are the means of A_i and B_i respectively.

- Select the shortest k distances from the unknown test feature vectors.

- Determine the most common class label among these k neighbors using a majority voting approach.

IV. Experimental Results and Discussion

The contribution of the proposed work lies in the development of an FER system using a pre-trained AlexNet model as a feature extractor combined with a KNN classifier. The main contributions of this paper are summarized as follows:

- Feature extraction using the pre-trained AlexNet model with different fully connected layers.
- Implementation of a KNN classifier with different distance metrics.
- Evaluation and analysis of different fully connected layers of the AlexNet model with the KNN classifier.

The performance metrics for each emotion were evaluated in terms of the Recognition Rate (RR), as defined below:

$$RR\,(\%) = \frac{\text{number of correctly recognized images}}{\text{number of test images}} \times 100 \qquad (2)$$

A. CK+ Dataset description

Extended Cohn–Kanade (CK+) [14], it is among the most widely used datasets in FER systems. The dataset comprises 981 images extracted from 327 video sequences involving 118 participants. Each sequence includes 10 to 60 frames, capturing the transition from a neutral to an extreme facial expression. The images are categorized into seven basic expressions and have a resolution of 48 × 48 pixels, as detailed in Table 2.

The dataset was divided into training and testing sets, with 70% for training and 30% used for testing the model's accuracy, following the standard split ratio. The emotions are denoted by the following abbreviations: Anger (AN), Disgust (DI), Fear (FE), Happiness (HA), Contempt (CO), Sadness (SA) and Surprise (SU).

Table 2. Number of images for each expression in the CK+ dataset.

Dataset	AN	DI	FE	HA	CO	SA	SU	∑
CK +	135	177	75	207	54	84	249	981

B. Optimal Configuration of KNN Classifier Using AlexNet Features

The aim of this experiment is to determine the optimal configuration of the KNN classifier by selecting the best value of the k nearest neighbors number and the most effective distance metric, using features extracted from different fully connected layers of the AlexNet model.

1) KNN Optimization with AlexNet FC6 Features

Table 3 presents the recognition rates for the different values of k varying from 1 to 10 and for the different distances (Euclidean, Cityblock, Correlation and Cosine) using the FC6 AlexNet features.

Table 3: Performance of KNN Classifier Using Different Distance Metrics with AlexNet FC6 Features

K	Euclidean	Cityblock	Correlation	Cosine
1	94.31	93.97	94.31	94.31
2	79.93	79.93	80.93	80.93
3	73.91	74.58	71.57	71.90
4	67.22	67.55	63.54	64.21
5	66.22	66.88	63.21	63.54
6	67.55	67.89	66.55	67.89
7	72.90	72.57	69.89	69.56
8	73.24	73.91	71.57	71.57
9	74.91	75.25	71.90	71.90
10	76.58	77.59	73.24	71.57

The FC6 layer of AlexNet provides a vector with 4096 features, which was used for the KNN optimization analysis. Here are the summarized points:

The highest recognition rate (94.31%) is obtained at k=1 with all distances except Cityblock.

Recognition rates decrease significantly as K increases from 1 to 5, dropping to approximately 66%. This suggests that increasing K reduces the influence of the nearest neighbor.

After k=5, recognition rates begin to improve slightly, indicating potential stabilization as more neighbors are considered (e.g., 76.58% at K=10for Euclidean).

2) KNN Optimization with AlexNet FC7 Features

The FC7 layer of AlexNet provides the same number of features as the previous layer FC6 with a vector of 4096 features.

Table 4 presents the recognition rates for the different values of k varying from 1 to 10 and for the different distances (Euclidean, Cityblock, Correlation and Cosine) using the FC7 AlexNet features.

Table 4: Performance of KNN Classifier Using Different Distance Metrics with AlexNet FC7 Features

K	Euclidean	Cityblock	Correlation	Cosine
1	95.32	95.31	95.31	**96.32**
2	79.59	79.59	79.93	79.93
3	73.91	73.91	71.90	71.23
4	63.54	63.87	65.55	65.21
5	62.20	62.54	62.20	61.20
6	63.21	63.21	65.88	65.21
7	67.89	66.88	69.23	68.56
8	69.56	68.56	70.56	70.23
9	69.56	69.89	70.56	70.23
10	69.23	68.89	68.89	71.90

From this table we conclude that:

The highest recognition rate is observed when K=1, with the Cosine distance metric achieving the highest value of 96.32% showing that it works better with the feature vector that the AlexNet FC7 layer generates.

Correlation and Cosine tend to yield better results than Euclidean and Cityblock as K increases, which could indicate that these metrics capture higher-level dependencies better in this feature vector.

3) KNN Optimization with AlexNet FC8 Features

Table 5 presents the performance of a KNN classifier across various values of K using AlexNet FC8 with a vector of 1000 features and different distance metrics: Euclidean, Cityblock, Correlation, and Cosine. The following observations can be made:

For k=1, the classifier achieves the highest accuracy (95.65%) consistently across all metrics, except for Cityblock, which yields an accuracy of 95.31%. As K increases, the accuracy generally decreases, indicating the potential drawback of considering too many neighbors in the KNN classifier. The Correlation and Cosine metrics gives relatively the same performance across different values of K.

TABLE 5: PERFORMANCE OF KNN CLASSIFIER USING DIFFERENT DISTANCE METRICS WITH ALEXNET FC8 FEATURES

K	Euclidean	Cityblock	Correlation	Cosine
1	95.65	95.31	95.65	95.65
2	80.93	80.93	81.60	81.60
3	75.25	74.58	74.91	74.91
4	66.88	65.55	70.23	70.23
5	62.87	63.87	67.22	67.22
6	62.54	63.21	68.56	68.56
7	62.21	65.88	67.22	67.22
8	65.88	67.22	67.89	67.89
9	66.22	65.88	65.55	65.55
10	65.21	66.22	66.22	66.22

4) KNN Optimization with Combined AlexNet FC6, FC7, and FC8 Features

Table 6 outlines the performance of the KNN classifier using combined AlexNet features from layers FC6, FC7, and FC8 with 9192 features with four different distance metrics: Euclidean, Cityblock, Correlation, and Cosine. This approach aims to leverage the complementary information captured by these fully connected layers, which may improve the effectiveness of the feature set and enhance system performance.

TABLE 6: PERFORMANCE OF KNN CLASSIFIER USING DIFFERENT DISTANCE METRICS WITH COMBINED ALEXNET FC6, FC7, AND FC8 FEATURES

K	Euclidean	Cityblock	Correlation	Cosine
1	94.31	94.31	94.31	94.31
2	79.59	79.93	80.60	80.60
3	74.58	73.91	71.57	72.57
4	65.88	64.88	64.21	64.88
5	66.55	66.55	64.21	63.87
6	69.23	69.56	65.55	65.55
7	72.57	70.56	70.23	69.89
8	73.57	72.57	70.90	71.57
9	75.58	72.57	70.23	70.23
10	77.59	74.24	70.56	70.56

From Table 5 we can conclude that at K=1, the KNN classifier achieves the highest accuracy (94.31%) across all distance metrics. This consistency suggests that when combining different FC layers, making the choice of distance metric less impactful.

C. Comparison of Different Extracted Features

Figure 3 presents the recognition rates of the KNN classifier for the different distances (Euclidean, Cityblock, Correlation and Cosine) using various Fully connected layers (FC6, FC7, FC8 and their combined).

We can observe from Figure 3 that the FC7 layer combined with the KNN using Cosine distance outperforms the other layers with different distance metrics, achieving a recognition rate of 96.32%. This demonstrates that FC7 may contain more relevant features for the FER task. On the other hand, FC6 provides the lowest recognition rate of 93.97% when using the Cityblock distance. FC8 yields a good result with 95.31%, despite having fewer feature vectors compared to the other layers. No improvement was observed when combining the layers, with a consistent recognition rate of 94.31%, which may be attributed to the high dimensionality of the feature vectors.

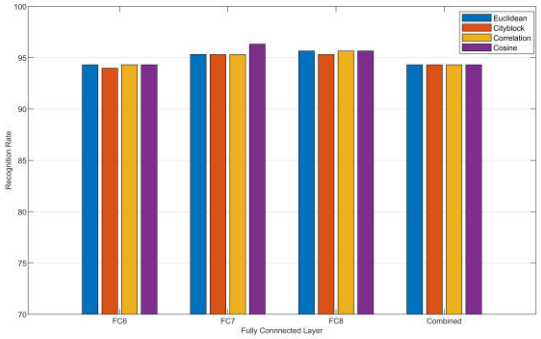


Figure 3:Comparison of Different Extracted Features

V. CONCLUSION

In this study, we used the pre-trained AlexNet model as a feature extractor for a FER system. Features were extracted from the fully connected layers (FC6, FC7, and FC8), as well as a combination of these layers, and were classified using the KNN classifier with various distance metrics, including Euclidean, Cityblock, Correlation, and Cosine. The results indicate that the features from the FC7 layer combined with the KNN classifier using the Cosine distance metric achieved the highest recognition rate of 96.32%. This was followed by features from the FC8 layer, which achieved 95.65% accuracy with all KNN distance metrics except for Cityblock, which yielded 95.32%. In contrast, combining features from all three FC layers did not improve the system's performance. For future work, we plan to explore and compare other pre-trained models, such as InceptionV3, combined with classifiers like Support Vector Machine (SVM).

REFERENCES

[1] A. R. Dalabehera, S. Bebortta, A. K. Singh, and D. Senapati, "Real-time Criticality Evaluation Through Vision-based Human-centric Emotion, Activity and Object Interactions," in 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU), Mar. 2024, pp. 1–6. doi: 10.1109/IC-CGU58078.2024.10530702.

[2] D. Zhu, Y. Fu, X. Zhao, X. Wang, and H. Yi, “Facial Emotion Recognition Using a Novel Fusion of Convolutional Neural Network and Local Binary Pattern in Crime Investigation,” Computational Intelligence and Neuroscience, vol. 2022, no. 1, p. 2249417, 2022, doi: 10.1155/2022/2249417.
[3] Z. Zou and S. Ergan, “Towards emotionally intelligent buildings: A Convolutional neural network based approach to classify human emotional experience in virtual built environments,” Advanced Engineering Informatics, vol. 55, p. 101868, Jan. 2023, doi: 10.1016/j.aei.2022.101868.
[4] E. Tsalera, A. Papadakis, M. Samarakou, and I. Voyiatzis, “Feature Extraction with Handcrafted Methods and Convolutional Neural Networks for Facial Emotion Recognition,” Applied Sciences, vol. 12, no. 17, Art. no. 17, Jan. 2022, doi: 10.3390/app12178455.
[5] N. Kumar H N, A. S. Kumar, G. Prasad M S, and M. A. Shah, “Automatic facial expression recognition combining texture and shape features from prominent facial regions,” IET Image Processing, vol. 17, no. 4, pp. 1111–1125, 2023, doi: 10.1049/ipr2.12700.
[6] S. Qin, Z. Zhu, Y. Zou, and X. Wang, “Facial expression recognition based on Gabor wavelet transform and 2-channel CNN,” International Journal of Wavelets, Multiresolution and Information Processing, Nov. 2019, doi: 10.1142/S0219691320500034.
[7] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, “Improved Facial Expression Recognition Based on DWT Feature for Deep CNN,” Electronics, vol. 8, no. 3, Art. no. 3, Mar. 2019, doi: 10.3390/electronics8030324.
[8] F. Z. Boukhobza, A. Hacine Gharbi, and K. Rouabah, “A New Facial Expression Recognition Algorithm Based on DWT Feature Extraction and Selection,” IAJIT, vol. 21, no. 4, 2024, doi: 10.34028/iajit/21/4/6.
[9] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Dec. 2001, p. I–I. doi: 10.1109/CVPR.2001.990517.
[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2012. Accessed: Nov. 10, 2024. [Online]. Available: https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
[11] S. Anber, W. Alsaggaf, and W. Shalash, “A Hybrid Driver Fatigue and Distraction Detection Model Using AlexNet Based on Facial Features,” Electronics, vol. 11, no. 2, Art. no. 2, Jan. 2022, doi: 10.3390/electronics11020285.
[12] E. Fix and J. L. Hodges, “Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties,” International Statistical Review / Revue Internationale de Statistique, vol. 57, no. 3, pp. 238–247, 1989, doi: 10.2307/1403797.
[13] H. Roubhi, A. H. Gharbi, K. Rouabah, and P. Ravier, “Mutual Information-based Feature Selection Strategy for Speech Emotion Recognition using Machine Learning Algorithms Combined with the Voting Rules Method,” Engineering, Technology & Applied Science Research, vol. 15, no. 1, Art. no. 1, Feb. 2025, doi: 10.48084/etasr.9066.
[14] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression,” in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, Jun. 2010, pp. 94–101. doi: 10.1109/CVPRW.2010.5543262.

Comparison of three image fusion methods: applied to ALSAT-2A images

Samia MASSOUT [a,b]
[a]Department of Telecommunication, Faculty of electricla engineering
Houari Boumedienne University of Science and Technology
Bab Ezzouar, Algeria
[b]Department of Electrical Systems Engineering, Faculty of Technology
Boumerdes, Algeria
af_samia@yahoo.fr

Abstract— **In remote sensing, image fusion consists in combining a high spatial resolution panchromatic image with a low resolution multispectral image in order to synthesize a high resolution multispectral image. The objective of this article is to present and compare three methods of image fusion. These methods are based on the IHS transform (Intensity, Hue and Saturation), and the tradeoff parameter t, the 'à trous' algorithm, the combination of IHS transform and the 'à trous' wavelet transform. These methods are applied to the panchromatic and the multispectral images acquired by the Algerian satellite ALSAT-2A. The resolution ratio between the two images is four. The products of the image fusion are evaluated qualitatively and quantitatively.**

Keywords—image fusion, IHS transform, 'à trous' algorithm, spatial resolution, spectral resolution

I. INTRODUCTION

The earth observation optical satellite provide two types of data, namely the panchromatic (PAN) images and the multispectral (MS) images. The PAN image is characterized by a high spatial resolution and a low resolution spectral, whereas the MS image is characterized by a high spectral resolution and a low spatial resolution. The process of image fusion consists in combining the two types of images to synthesize a new high spatial resolution multispectral image.

Different methods of image fusion have been developed in the literature [1], [2] and [3]. According to Vivone et al [4], the methods of image fusion are grouped into two class: methods based on component substitution such as IHS (Intensity, Hue and Saturation) transform [5] and principal components analysis (PCA) [6], and techniques based on multiresolution analysis such as wavelet transform [7], curvelet transform [8] and [9] and contourlet transform [10]. Indeed, the principle of image fusion methods based on multiresolution analysis is to decompose the image in order to separate high frequencies form low frequencies. The recomposition process enables to inject the high frequencies into the multispectral (MS) image. The use of these methods in the field of image fusion has demonstrated their capabilities in improving the spatial information and in preserving the spectral one.

The objective of this article is to present and compare three methods of image fusion. These methods are based on the IHS transform and the tradeoff parameter t, the 'à trous' algorithm, and the combination of IHS transform and the 'à trous' algorithm,

This article is organized as follow. The level of image fusion is given in Section II. In Section III, we first introduce three methods of image fusion: image fusion method based on IHS transform and the tradeoff parameter t, image fusion method based on 'à trous' algorithm, and image fusion method based on combination of IHS transform and 'à trous'. In Section IV, we present the dadaset used to validate the results. Finally, in Section V, we compare the fusion results of three methods.

II. LEVEL OF IMAGE FUSION

Image fusion can be performed at three levels: pixel level, feature level and decision level. The pixel level fusion determines the value of each pixel based on a set of pixels from the source images [11] and [12]. The feature level fusion extracts salient features from each source image and performs the integration based on extracted features.

Figure illustrated three levels of fusion methods, it necessary that the source images be accurately superimposed. Therefore, both images have to be coregistered and resampled to achieve pixel alignment.

A. *Pixel Level Fusion*

Pixel level fusion generates a fused image in which information content associated with each pixel is determined from a set of pixel is determined from a set of pixels in source images.

Fusion at this level can be performed either in spatial or in frequency domain. However, pixel level fusion may conduct to contrast reduction.

B. *Feature level fusion*

Feature level fusion requires the extraction of salient features which are depending on their environment such as pixel intensities, edges or textures. These similar features form the input images are fused. This fusion level can be used as a means of creating additional composite features. The fused image can also be used for classification or detection.

C. *Decision level fusion*

Decision level is a higher level of fusion. Input images are processed individually for information extraction. The

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

obtained information is then combined applying decision rules to reinforce common interpretation.

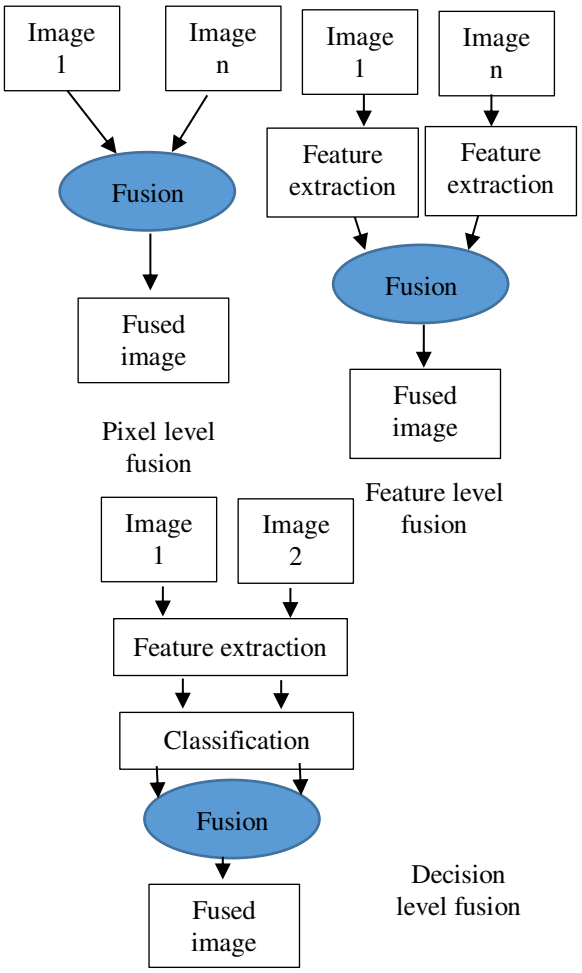


Figure 1. Block diagram of level of image fusion.

III. IMAGE FUSION METHODS

The three image fusion methods are described in the next sections.

A. *Image fusion based on IHS and tradeoff parameter t*

The IHS transform is widely used in image fusion to exploit the complementary nature of MS images. The HIS fusion method converts a color image from the red, green, and bleu (RGB) space into the IHS color space. The intensity band (I) is replaced by a high resolution panchromatic image and then transformed back into the original RGB space together with the previous hue bans and the saturation band, resulting in an IHS fused image.

The image fusion based on IHS transform and tradeoff parameter [13] is expressed as follows:

$$\begin{bmatrix} R^* \\ G^* \\ B^* \\ NIR^* \end{bmatrix} = \begin{bmatrix} R + (I_{new} - I) \\ G + (I_{new} - I) \\ B + (I_{new} - I) \\ NIR + (I_{new} - I) \end{bmatrix}$$

$$\begin{bmatrix} R^* \\ G^* \\ B^* \\ NIR^* \end{bmatrix} = \begin{bmatrix} PAN - \frac{(PAN-I)}{t} + (R - I) \\ PAN - \frac{(PAN-I)}{t} + (G - I) \\ PAN - \frac{(PAN-I)}{t} + (B - I) \\ PAN - \frac{(PAN-I)}{t} + (NIR - I) \end{bmatrix} \quad (1)$$

Where:

R^*, G^*, B^* and NIR^* are the fused bands for red, green, blue and near-infrared bande respectively.

This method is very fast and easy to implement because all the MS images can be fused at one time.

B. *Image fusion base on 'à trous' algorithm*

The wavelet transform provides a framework to decompose images into a number of new images, each of them with a decreasing degree of resolution, and to separate the spatial detail information of the image between two successive resolution degrees [7].

In the 'à trous' algorithm, the image decomposition can be represented with a pyramid as in Mallat with a parallelepiped. The basis of the parallelepiped is the original image, A_2^j at a resolution 2^j, with C columns and R rows. Each level of the parallelepiped is an approximation to the original image. When climbing up through the resolution levels, the successive approximation images have a coarser spatial resolution but the same number of pixels as the original image (figure 2). If a dyadic decomposition approach is applied, the resolution of the approximation image at the *N*th level is 2^{j-N}.

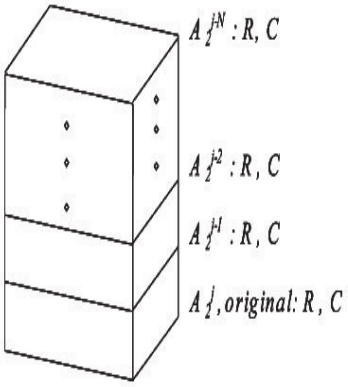


figure 2. Parallepiped representation of the 'à trous' wavelet transform.

The central idea of all image fusion methods is based on multiresoultion analysis and discrete wavelet transform (DWT) is to extract from the panchromatic image the spatial detail that is not present in the multispectral image in order to insert it later in the latter. The detail information of panchromatic image can be extracted using the 'à trous' discrete wavelet transform algorithm [12]. The spatial detail is collected in the wavelet coefficient images or wavelet planes and it could be injected into the multispectral image.

For the practical implementation of the 'à trous' algorithm, the following low pass filter (h) is used :

$$h = \frac{1}{256}\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \quad (2)$$

The steps for merging multispectral and panchromatic images using this method are :

1- Generate new panchromatic images, whose histograms match those of each band of the multispectral image.

2- Perform the second level wavelet transform only on the panchromatic images.

3- Add the wavelet planes of the panchromatic decomposition to each band of the multispectral dataset.

The image fusion method based on combination of HIS transform and 'à trous' algorithm is presented in the next section.

C. Image fusion based on IHS and'à trous' algorithm

The steps of image fusion multispectral and panchromatic images using the combination of the IHS transform and 'à trous' algorithm are the following [14], [15] and [17]:

- Apply the IHS transform to the RGB composition of the multispectral image. This transformation separates the spatial information of the multispectral image into the intensity component.

- Generate a new panchromatic image, whose histogram matches the histogram of the intensity component.

- Decompose only the new panchromatic image, using the 'à trous' algorithm, and obtain the first and second wavelet planes that pick the spatial detail of this image not present in the multispectral one.

- Add these details to the intensity component.

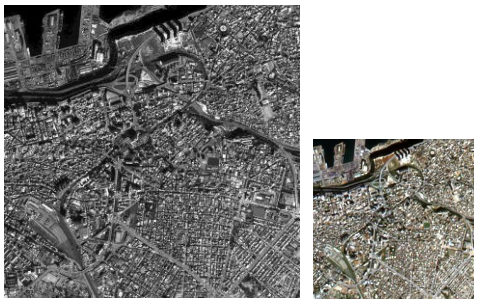
- insert the spatial information of the panchromatic image into the multispectral one through the inverse IHS transform.

The disadvantage of the fusion methods based on the IHS transform is that they can only be applied to three band RGB components.

The IHS transform and tradeoff parameter t method, 'à trous' algorithm, and the combination of the IHS transform and 'à trous' algorithm have been used to fusion ALSAT-2A panchromatic and multispectral images. These images are shown in the next section.

IV. IMAGES TEST

In this paper, the data set is images of the region of Algiers (Algeria) from the satellite ALSAT-2A (Algerian Satellite). It is composed of a panchromatic image and a multispectral image with four bands (bleu= B, green =G, red =R and near infrared = NIR).



(a) (b)

Figure 3. ALSAT-2A images. (a) panchromatic image, (b) multispectral image.

The three image fusion methods are described in the next sections. The panchromatic image size is 1024 × 1024 pixels and the multispectral image size is 256 × 256 pixels. Panchromatic and multispectral images have a spatial resolution equal respectively to 2.5 m and 10 m. The resolution ration between the two images is equal to 4. The panchromatic image and the color composition of the red, green and blue bands are given in figure 3.

Table 1. Characteristic of satellite data used in the study

Satellite Data	Band	Spatial Resolution	Size of images (pixels)
MS data	Infrared	10 m	256 × 256
	Red	10 m	256 × 256
	Green	10 m	256 × 256
	Bleu	10 m	256 × 256
PAN data	PAN	2.5 m	1024 × 1024

The application of the image fusion on the data set allows synthesizing multispectral image with resolution of 2.5 m. the results of the image fusion are presented in the next section.

V. RESULTS AND ANALYSIS

In this section, we present and compared the results obtained by the three methods. The comparison is based on the visual interpretation and the statistical assessment [18], [19] and [20]. Two types of statistical parameters are used to quantitatively estimate the products of image fusion.

The first parameter category includes the correlation coefficient [18], the quality index Q [21] and the ERGAS (Erreur Relative Globale Adimensionnelle de synthese) parameter [18]. These three parameters allow estimating the spectral quality of the fused images. The correlation coefficient (CC) is calculated globally for each spectral band. The image quality index Q is calculated locally with a 32 × 32 window. The ERGAS parameter is estimated globally for all multispectral bands.

- Correlation coefficient (CC)

The CC between the original image and the fused image is defined as:

$$CC = \frac{\sum_{mn}(A_{mn}-\bar{A})(B_{mn}-B)}{\sqrt{(\sum_{mn}(A_{mn}-\bar{A})^2)(\sum_{mn}(B_{mn}-\bar{B})^2)}} \quad (5)$$

Where:

$\bar{A}$ is the mean value of the original image,

$\bar{B}$ is the mean value of the fused image.

The result of this equation shows similarity in the small structures between the original image and the fused image.

- ERGAS

The ERGAS index for the fusion images is expressed as follow

$$ERGAS = 100\frac{h}{l}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{RMSE^2(B_i)}{M_i^2}} \quad (6)$$

Where:

h is the resolution of the high spatial resolution image,

l is the resolution of the low spatial resolution,

M_i is the mean radiance of each spectral band involved in the fusion.

N is the number of spectral bands.

The RMSE is the root mean square error computed in the following expression :

$$RMSE^2(B_i) = bias(B_i) + SD^2(B_i) \quad (7)$$

- Image quality index Q

The image quality index Q proposed by Wang and Bovik [21].

$$Q = \frac{4\sigma_{AB}\bar{A}\bar{B}}{(\sigma_A^2+\sigma_B^2)[(\bar{A})^2+(\bar{B})^2]} \quad (8)$$

Where:

σ_A^2 and σ_B^2 are the variances of A and B,

σ_{AB} is the covariance between A and B.

The Q index estimates the difference between two images as a combination of three different factors: loss of correlation, luminance distortion and contrast distortion. As image quality is often space dependent, Wang and Bovik recommend to calculate the Q index using a sliding window method. In this work, sliding windows with a size of 32 × 32 is used. As the Q index can only be applied to monochromatic image, the average value (Q_{avg}) is used as a quality index for multispectral images. The higher Q_{avg}) value the higher the spectral and radiometric quality of the fused images.

Parameters of the second category are : the entropy and spatial correlation coefficient (CC_s) [7]. These two parameters are calculated in order to estimate the spatial information of the fused images.

- Entropy

The entropy parameter measures the spatial information quality contained in an image. Shannon was the first person to introduce entropy to quantify the information.

The entropy of the image can be evaluated as :

$$Entropy = -\sum_{i=0}^{255} p_i \, log_2(p_i) \quad (9)$$

Entropy can directly reflect the average information content of an image. The maximum value of entropy can be produced when each gray level of the whole rang has the same frequency. If entropy of fused image is higher than parent image.

- spatial correlation coefficient (CC_s)

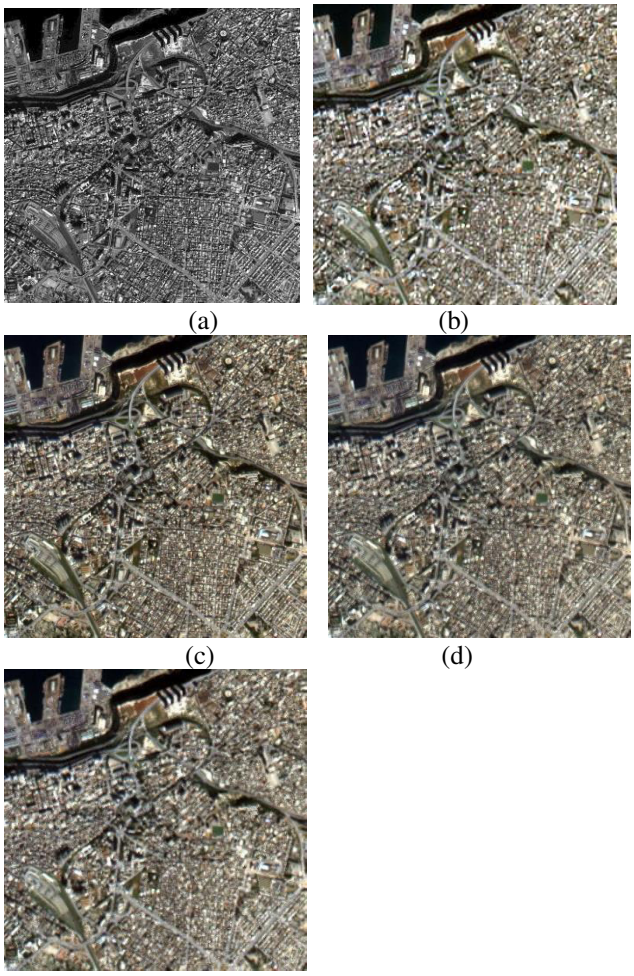
The spatial correlation coefficient is applied between the high frequencies of fused and panchromatic images. The high spatial frequencies are extracted using a Laplacian filter (H) as follows:

$$H = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (10)$$

The high correlation coefficient between the fused filtered image and the panchromatic filtered image (CC_s) indicate that most of the spatial information of the panchromatic image was incorporated during the fusion process. The CC_s has the same definition as the CC.

A. Qualitative assessment

Figure 4 (a) and 2(b) represent respectively the panchromatic. Figures 4 (c) – (e) represent the fused multispectral images (2.5 m) using on the IHS transform and the tradeoff parameter t, the 'à trous' algorithm, the combination of IHS transform and the 'à trous' algorithm respectively.



(a) (b) (c) (d) (e)

Figure 4. Fused ALSAT-2A images : (a) panchromatic image, (b) original multispectral image. Fused MS image using : (c) the IHS transform and the tradeoff parameter t, (d) the 'à trous' algorithm, (e) the combination of IHS transform and the 'à trous' algorithm.

The qualitative assessment of the fusion product is based on the visual interpretation. Comparing the results of the image fusion with the original images, we notice the appearance of spatial detail in the fused images using : the IHS transform and the tradeoff parameter t, the 'à trous' algorithm, and the combination of the IHS and the 'à trous' algorithme. However, the fused result obtained from the IHS transform and the tradeoff parameter t is visually better compared to the other results in terms of conservation of spectral information.

B. Quantitaive assessment

The results of the spatial and spectral assessment of the fusion products are presented in Table 1 and Table 2. The best values of the indices obtained from the statistical evaluation are given in bold.

From Table 2, we see that the values of the CC_s parameter calculated for the IHS transform and tradeoff parameter t method is able to improve the spatial resolution of the ALSAT-2A satellite.

From Table 3, we notice that the values of CC, Q index and ERGAS show the IHS transform and tradeoff parameter t is able to preserve the spectral information of the MS image.

Table 2. Spatial evaluation of image fusion products

		CC_s	Entropy
IHS transform and parameter t	B	**99.662 %**	4.260
	G	**99.233 %**	4.568
	R	**98.895 %**	4.733
	NIR	**98.714 %**	4.815
'à trous' algorithm	B	95.680 %	**4.289**
	G	95.487 %	**4.568**
	R	95.528 %	**4.711**
	NIR	95.320 %	**4.822**
IHS and 'à trous' algorithm	B	91.565 %	4.156
	G	91.534 %	4.480
	R	91.681 %	4.634
	NIR	/	/

Table 3. Spectral evaluation of image fusion products

		CC	Q	ERGAS
IHS transform and parameter t	B	**92.399 %**	**0.563**	**1.854**
	G	**95.659 %**	**0.436**	
	R	**95.990 %**	**0.785**	
	NIR	**96.178 %**	**0.784**	
'à trous' Algorithm	B	92.197 %	0.589	2.148
	G	93.446 %	0.734	
	R	93.686 %	0.785	
	NIR	94.670 %	0.791	
IHS and 'à trous' algorithm	B	92.657 %	0.723	2.094
	G	93.291 %	0.837	
	R	93.674 %	0.872	
	NIR	/	/	

From the qualitative and quantitative evaluations, we are able to see that the IHS transform and tradeoff parameter t is better compared to other methods.

VI. Conculsion

A presentation and a comparison of three methods of panchromatic and multispectral image fusion is presented in this article. These methods are based on the IHS transform and tradeoff parameter t, the 'à trous' algorithm, and the combination of the IHS transform and the 'à trous' algorithm. These methods are applied on the ALSAT-2A images. The fused products were evaluated qualitatively and quantitatively. The assessment and the comparison of the fused images have shown that the IHS transform and tradeoff parameter t is able to produce multispectral image with high spatial resolution available with better conservation of the spectral information.

References

[1] W. Carper, T. Lillesand, and R. Kiefer, ''The use of intensity –Hue-saturation transformations for merging SPOT panchromatic and multispectral image data,'' *Photogramm. Eng. Remote Sens*, No.4, Vol. 506, pp.459 – 467, 1990.

[2] K. Amolins, Y. Zhang, and P. Dars, ''Wavelets based image fusion techniques –an introduction, review and comparison'', *J. Photogramm. Remote Sen.*, No. 4, Vol. 62, pp. 249 – 263, 2007.

[3] V.P.Shah, N H. Younan, and R. L. King, ''Pan-sharpening via the contourlet transform,'' in Proc. *IEEE Int. Geo-science Remote Sens. Symp. IGARSS 2007*, pp. 310 – 313, 2007.

[4] G. Vivone et al., ''A critical comparison among pansharpening algorithms'', *IEEE Trans. Geosci. Remote Sens*. No. 5, Vol. 53, pp. 2565 – 2586, 2015.

[5] T-M. Tu, S-C. Su, H-C. Chyun, P. S. Huang, ''A new look at IHS-like image fusion methods,'' *Inf. Fusion* . No. 3, Vol. 2, pp. 177 – 186, 2001.

[6] Gonzalez-Audicana M., Saleta J. L., Catalan R. G., and Garcia R., ''Fusion of multispectral and panchromatic Images using improved IHS and PCA mergers based on wavelet decomposition'', *IEEE Trans. Geos. Remote Sens.*, No. 6, Vol. 42, pp. 1291 – 1299, 2004.

[7] J. Zhou, D. L. Civco, and J. A. Silander, ''A wavelet transform method to merge Landsat TM and SPOT panchromatic data'', *Int. J. Remote Sens*. No. 4, Vol. 19, pp. 743 – 757, 1998.

[8] F. Nencini, A. Garzelli, S. Baronti, L. Alparone, ''*Remote sensing image fusion using the curvelet transform*'', *Inf. Fusion*, Vol. 2, No. 8, pp. 143–156, 2007.

[9] B. Zhong, C. Zhang, M.W. Liao, H.S. Cai, ''Spectral preservation fusion for remote sensing images using focus measure operators based on fast discrete curvelet transform and hyperspherical color space'' *J. Appl. Remote Sens.*, No. 3, Vol. 12, 035017, 2018.

[10] M. Mangolini, ''Apport de la fusion d'images satellitaires multicapteurs au niveau pixel en télédétection et photo-interprétation'', *Thèse de Doctorat en Sciences de Ingénieur, Université de Nice – Sophia Antipolis,* Nice, France, 174 p, 1994.

[11] R. Aguilar-Ponce, J. L. Tecpanecatl – Xihuitl, A. Kumar and M. Bayoumi, ''Pixel level image fusion based on linear Algebra'', *IEEE Int. Symposi on Circuit and Systems*, pp. 268 – 2661, 2007.

[12] S. Bhagyamma, A. L. Choodarathnakara, T. K. Ranjitha, A. P. Ramya, K. M. Niranjan, ''Performance analysis of image fusion techniques to improve quality of satellite data'', *Int. J. Scientif. Resea. Scien. Eng. Tech*, No. 3, Vol. 3, pp. 314 – 326, 2017.

[13] M. Choi, ''A new Intensity-Hue-Saturation fusion approach to image fusion with a tradeoff parameter '', *IEEE Trans. Geos.Remote Sens*., No. 6, Vol 44, pp. 1672 – 1682, 2006.

[14] M Gonzalez-Audicana., X Otazu., O Fors., and A Seco., ''Comparison between Mallat's and the 'à trous' discrete wavelet transform based algorithms for the fusion of

multispectral and panchromatic images'', *Int. J. Remote Sens.*, No. 3, Vol. 26, pp. 595 – 614, 2005.

[15] L. Wald, ''Data Fusion: Definition and Architectures fusion of images of spatial resolution,'' *Les Presses, Ecole des Mines de Paris*, 2002.

[16] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, ''A global quality measurement of pan-sharpened multispectral imagery,'' *Remote Sens. Lett* , No. 4, Vol. 1, pp. 313 – 317, 2004.

[17] G. Hong and Y. Zhang, ''Comparison and improvement of wavelet based image fusion'', *Int. J. Remote Sens.*, Vol. 29, No. 3, pp. 673 – 691, 2008.

[18] T. Ranchin and L. Wald, ''Fusion of high spatial and spectral resolution images : the ARSIS concept and its implementation'' , *Photogramm. Eng. Remote Sens.*, Vol. 66, pp. 49 – 61, 2000.

[19] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti and L. Wald, ''Image fusion – the ARSIS concept and some successful implementation schemes'' , *ISPRS J. Photogramm. Remote Sens.*, Vol. 58, pp. 4- 18, 2003.

[20] S. T. Budhewar, ''Wavelet and curvelet transform based image fusion algorihm'', *Int. J. of comput. Science and Inf. Tech.*, Vol. 5, No. 3, pp. 3703 – 3707.

[21] Z. Wang, and Bovik A. C., ''A universal image qualiy index'', *IEEE Signal Process. Lett.*, No. 3, Vol. 9, pp. 81 – 84, 2002.

An Enhanced Artificial gorilla troops optimizer with Chaotic map for LEDs Placement in Indoor VLC Systems

1st Abdelbaki Benayad
University of M'Hamed Bougara Boumerdes
LIMOSE Laboratory
Avenue of Independence, 35000 Boumerdes, Algeria
a.benayad@univ-boumerdes.dz

2nd Yassine Meraihi
University of M'Hamed Bougara Boumerdes
LIST Laboratory
Avenue of Independence, 35000 Boumerdes, Algeria
y.meraihi@univ-boumerdes.dz

3rd Amel Boustil
University of M'Hamed Bougara Boumerdes
LIMOSE Laboratory
Avenue of Independence, 35000 Boumerdes, Algeria
a.boustil@univ-boumerdes.dz

***Abstract*—Visible Light Communication (VLC) is a new technology that has attracted lately much interest from researchers and academics. It allows communication between users using photo-detectors (PDs) as receivers and light emitting diodes (LEDs) as transmitters. The deployment of LEDs in indoor VLC Systems is an important issue that affects the coverage of the network. In this article, we propose an improved version of Artificial gorilla troops optimizer (GTO) , named (EGTO), to resolve the LEDs placement problem in indoor visible light communication (VLC) systems. The EGTO is based on the integration of chaotic map concept into the standard GTO to improve its optimization performance. By taking into account the user throughput and coverage metrics while employing several produced instances and evaluating results against some meta-heuristics, the usefulness of EGTO was confirmed. The meta-heuristics that we used in the comparison are GTO, (MRFO) Manta Ray Foraging Optimizer, (CHIO) Herd immunity coronavirus optimizer, (MPA) Marine Predator Algorithm,(BA) Bat Algorithm and (PSO) Particle Swarm Optimizer. The results showed that EGTO is more effective in finding optimal LEDs positions.**

***Index Terms*—Artificial gorilla troops optimizer, LED placement problem, Visible Light Communications, Chaotic map.**

I. Introduction

Currently, Light Communication (LC) technology is gaining more interest as a result of the massive bandwidth available in the uncontrolled optical spectrum [1]. Visible Light Communication (VLC) is among of such technology that can be considered as an alternative or complement to radio communications [2]. It can be used for many indoor applications such as localization [3], [4], networking at home [5], and communication in locations like hospitals and airplane cabins where radio frequency (RF) radiation is restricted [6]. This technology is based on LEDs, which are regarded like a sustainable and energy-efficient source of light [7]. LEDs can both send data and provide illumination. thus, achieving dual functionality. The placement of LEDs in indoor VLC systems is an important issue that affects the performance of the network. It is known to be an NP-hard problem [8]–[11]. Therefore, meta-heuristics are intelligent methods that can provide efficient solutions to solve it. Some studies based on meta-heuristics have been proposed for solving the issue of LEDs placement in indoor VLC communication systems. In this context, Rui et al. [10] utilized the PSO algorithm for solving the LED deployment problem in an indoor VLC communication system. The effect of changing the LEDs and user numbers were discussed. The results revealed that PSO gives good results compared to random models by obtaining the best-LEDs deployment and minimizing the average outage area rate.

In the work of Chaochuan et al. [11], an enhanced Cuckoo Search Algorithm, called VLP-IACS, was proposed for solving the 3D indoor LEDs positioning problem in an indoor VLC system. Simulations were carried out in a room size of 5 m × 5 m × 6 m and the findings indicated that the VLP-IACS algorithm demonstrated strong performance.

Benayad et al. [12] tackled the challenge of positioning LEDs in indoor VLC systems by proposing an Hybrid Coronavirus Herd Immunity Optimizer, known as (ICHIO-FA), within the confines of an indoor room measuring 10 m × 10m × 3 m, . Simulation outcomes indicated that (ICHIO-FA) outperformed other algorithms such as classical WOA, MRFO, MPA, BA, and PSO by achieving optimal layouts for LEDs.

In their study, Wen et al. [9] tackled the issue of LED placement in indoor VLC systems by proposing a Modified Artificial Fish Swarm Algorithm (MAFSA). Through experi-

mentation, a positioning error of only 4.05 mm was achieved.

Kumawat et al. [8] used the Wale Optimization Algorithm for solving the LED panel placement Problem in the indoor VLC system. WOA was performed in a room with dimensions 5 m × 5 m × 3 m, having 4 to 6 LED panels considering the received power and SNR parameters. The findings that the WOA algorithm performed better than the PSO algorithm.

Artificial gorilla troops optimizer (GTO) is a novel meta-heuristic algorithm inspired by gorilla troops' social intelligence in nature, introduced by Abdollahzadeh and Mirjalili [13] in 2021. This meta-heuristic has some issues such as poor exploration and local optimum. To improve its solutions, GTO was mixed with chaotic map concept.

Benayad et al. [14], Yancang Li et al. [15], Hangqi Ding et al. [16], Gaganpreet Kaur et al. [17], Dharmbir Prasad et al. [18], and Diego Oliva et al. [19], for example, combined some of chaotic maps with WOA to balance between exploitation and exploration, and take it away from local optima.

In this work, we proposed an enhanced version of GTO, called EGTO, for solving the LED deployment issue in an indoor VLC communication system. The proposed EGTO integrate chaotic map concept to improve the optimization performance of GTO. The performance of EGTO was assessed considering the user coverage and throughput parameters. in a conference room with dimensions 10 m × 10 m × 4 m. The rest of this work is arranged as follows. In section II, the formulation of the LEDs placement problem in indoor VLC communication systems is presented. Section III describes the structure of the GTO algorithm, chaotic map. Section IV describes the EGTO model for solving the LEDs placement in indoor VLC systems. In section V, The simulation's findings are shown and discussed. Section VI presents the final conclusion of the study.

II. LEDs PLACEMENT PROBLEM FORMULATION

A. System model

In this work, we consider a typical indoor VLC system (V) in a room of dimension, $X \times Y \times Z$ as illustrated in Figure 1. This room is equipped with N Light Emitting Diodes (LEDs) installed on the ceiling to serve as optical access points (transmitters) and M receiving users inside this room. We further define d_{ij} the vertical distance between the i^{th} LED and the j^{th} receiver. Therefore, let:

- L is the set of N LEDs: $L = \{L_1, L_2, ..., L_N\}$ should be positioned for installation on the ceiling at locations(x_i, y_i, z_i), where $i \in (1, 2, 3, ...N)$.
- U is the set of M Users $U = \{U_1, U_2, ..., U_M\}$, each user has a photo detector (PD) with him, used as wireless receiver. We assume that the users are randomly distributed in the $3D$ room. To say a LED covers the user U_j, if the value of the power received Pr from PD of the user U_j is positive, It should be noted that each user U_j can be associated with at most one LED (i.e., L_i). In other words, it is associated with the closest LED and the higher power (Pr).

B. Mathematical model

In our studies, we focus on the network's coverage and throughput as the two primary aspects of a VLC system that should be optimized.

- With the following Equations (1)(2). the network's coverage is described:

$$Cov(V) = \sum_{j=1}^{M} max_{i \in \{1,...N\}}(Cov_{U_j}^{L_i}) \tag{1}$$

  where $Cov_{U_j}^{L_i}$ demonstrates how each user U_j is covered by LED L_i. By (2), we can determine the value of coverage.

$$Cov_{U_j}^{L_i} = 1 \;\; if \; Pr_j > 0 \tag{2}$$

- The throughput of the network is represented as follows:

$$Tr(V) = \sum_{j=1}^{M} max_{i \in \{1,...N\}}(Tr_{U_j}^{L_i}) \tag{3}$$

  where with every user U_j covered by LED L_i has a throughput defined by $Tr_{U_j}^{L_i}$. It can be given as (4) [20]

$$Tr_{U_j}^{L_i} = B \times log_2(1 + SNR) \tag{4}$$

  where B and SNR is the bandwidth, the signal-to-noise-ratio respectively. SNR can be expressed as (5)

$$SNR = \frac{(R.Pr)^2}{\sigma_t^2} \tag{5}$$

  where R is the PD responsivity and σ_t^2 is the total noise variance. Pr is the power received by PD.

Therefore, Here is how the objective function can be expressed:

$$f = (\lambda)(\frac{Cov(V)}{M}) + (1 - \lambda)(\frac{Tr(V)}{M \times Tr_{max}}) \tag{6}$$

where $\lambda \in [0, 1]$

III. PRELIMINARIES

A. Artificial gorilla troops optimizer (GTO)

In this section, inspired by gorillas' group behaviors, we provided a new metaheuristic algorithm called GTO, where specific mathematical mechanisms are presented to explain the two phases of exploration and exploitation fully.

Three different operators have been used in the exploration phase: migration to an unknown place to increase GTO exploration. The second operator, a movement to the other gorillas, increases the balance between exploration and exploitation. The third operator in the exploration phase, that is, migration towards a known location, significantly increases the GTO capability to search for different optimization spaces. On the other hand, two operators are used in the exploitations phase, which significantly increases the search performance in exploitation.

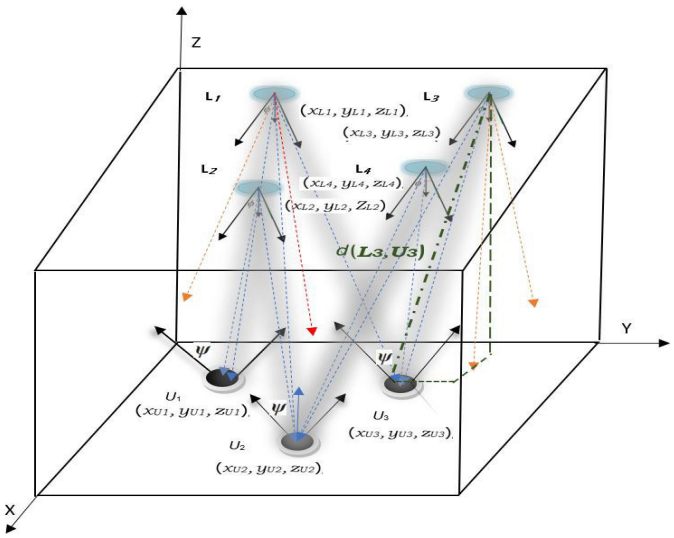


Fig. 1. System model of our VLC system

1) Exploration phase : We used three different mechanisms for the exploration phase, that is, migration to an unknown location, migration towards a known location, and moving to other gorillas. Each of these three mechanisms is selected according to a general procedure.

Equation (7) has been used to simulate the three mechanisms used in the exploration phase.

2) Exploitation phase : In the (GTO) algorithm's exploitation phase, two behaviors of Follow the Silverback and Competition for adult females are applied.

To select either follow the Silverback or Competition for adult females using the C value in Equation (8). If $C \geq W$, the follow the Silverback mechanism is selected, equation (13) is used to simulate this behavior, but if $C < W$, adult females' Competition is taken, equation (16) is used to simulate this behavior. W is a parameter to be set before the optimization operation.

$$GX(t+1) = \begin{cases} (UB - LB) \times n_1 + LB, & \text{if rand} < p, \\ (r_2 - C) \times X_r(t) + L \times H, & \text{if rand} \geq 0.5, \\ X(i) - L \times \big(L \times (X(t) - GX_r(t)) & \\ +r_3 \times (X(t) - GX(t))\big), & \text{if rand} < 0.5, \end{cases} \quad (7)$$

$$C = F \times \left(1 - \frac{It}{\text{MaxIt}}\right), \quad (8)$$

$$F = \cos(2 \times r_4) + 1, \quad (9)$$

$$L = C \times l, \quad (10)$$

$$H = Z \times X(t), \quad (11)$$

$$Z = [-C, C]. \quad (12)$$

$$GX(t+1) = L \times M \times \big(X(t) - X_{\text{silverback}}\big) + X(t), \quad (13)$$

$$M = \left(\frac{1}{N}\sum_{i=1}^{N} GX_i(t)\right)^{\frac{1}{g}}, \quad (14)$$

$$g = 2^L. \quad (15)$$

$$GX(i) = X_{\text{silverback}} - \big(X_{\text{silverback}} \times Q - X(t) \times Q\big) \times A, \quad (16)$$

$$Q = 2 \times r_5 - 1, \quad (17)$$

$$A = \beta \times E, \quad (18)$$

$$E = \begin{cases} N_1, & \text{if rand} \geq 0.5, \\ N_2, & \text{if rand} < 0.5. \end{cases} \quad (19)$$

B. Chaotic map concept

To improve also the searching capability and to increase the convergence rate, we used the Chaotic map concept, it's a deterministic approach to studying the behavior of dynamic and nonlinear systems. It has numerous crucial characteristics, including, non-converging, stochastic, ergodicity, bounded, regularity, non-repetitive, unpredictable, and non-periodic. These characteristics have been turned into a variety of chaotic maps, which are mathematical equations used to produce random parameters in meta-heuristics. Multiple chaotic maps have also been illustrated in the literature, including Sinusoidal map, Logistic map, Circle map, Gauss map, Piecewise map, Chebyshev map, Sine map, Piecewise map, Tent map, and Iterative map [21], [22]. With straightforward operations and dynamic randomization, the sine map is among the most basic and often used chaotic maps Eq.20 is the definition of the sine map equation.

$$SM_{k+1} = \frac{ac}{4} \sin(\pi SM_k), 0 \leq SM_O \leq 1 \quad (20)$$

Where SM_k is the value of Sine map at the *k-th* iteration. It is in [0, 1] and the control parameter ac between $0 < ac \leq 4$.

IV. The Proposed EGTO for Solving the LEDs Placement Problem

One of the main challenges with GTO is its slow convergence speed. In order to address this issue and improve overall performance, chaos theory were incorporated into the GTO optimization phase. These modifications allowed for greater control over the exploitation and exploration phases, ultimately enhancing the algorithm's global convergence speed. To speed up convergence rate and prevent being caught in local optima, chaotic maps are employed in the GTO algorithm. The pseudo-code of EGTO is given in Algorithm 1.

Regarding the **Chaos** method, the parameter p of Eq.(7) is modified using Sine chaotic map during iterations, and the new frequency equation is written as:

$$p_{i+1} = \frac{ac}{4} \sin(\pi p_i) \quad (21)$$

where p_i is the chaotic map value of parameter p, in the $i-th$ iteration, and it must fall among 0 and 1. ac=4 and p_0 = 0.7

Algorithm 1 Pseudocode of EGTO

Inputs: The population size N, maximum number of iterations T, and parameters β and p_0
Outputs: The location of Gorilla and its fitness value
—— Initialization
Initialize the random population X_i $(i = 1, 2, \ldots, N)$
Calculate the fitness values of Gorillas
while stopping condition is not met **do**
  Update C using Equation (8)
  Update L using Equation (10)
  —— Exploration phase
  for each Gorilla (X_i) **do**
    —— Chaotic integration
    Adjuster the parameter p using Equation (21)
    Update the location of Gorilla using Equation (7)
  end for
  Calculate the fitness values of Gorillas
  if GX is better than X **then**
    Replace them
  end if
  Set $X_{\text{silverback}}$ as the location of *silverback* (best location)
  —— Exploitation phase
  for each Gorilla (X_i) **do**
    if $|C| \geq 1$ **then**
      Update the location of Gorilla using Equation (13)
    else
      Update the location of Gorilla using Equation (16)
    end if
  end for
  Calculate the fitness values of Gorillas
  if new solutions are better than previous solutions **then**
    Replace them
  end if
  Set $X_{\text{silverback}}$ as the location of *silverback* (best location)
end while
return $X_{\text{BestGorilla}}$, bestFitness

V. Experimental results and discussions

This section will evaluate the effectiveness of the EGTO algorithm, to find the optimal LEDs placement and compare it with the original GTO, MRFO, CHIO, MPA, BA and PSO algorithms. Each algorithm is coded using Matlab. Runs all simulations in a Core i5-4310U 2.6 GHz-CPU RAM 12Go.A typical scheme of a room used for our simulations is shown in Figures 1.

The EGTO algorithm's performance was validated through mean throughput and coverage per user metrics using 16 scenarios. These scenarios consisted of various numbers of LEDs (ranging from 2 to 9) and users (ranging from 5 to 40). Each result in this chapter is the mean of 30 runs and were run for a total of 1000 iterations, while the simulation parameters used for the study can be found in Table I.

A. Effect of varying the number of LEDs

In this case, We'll discuss the results of eight scenarios which consists in the changing of the number of LEDs from 2 to 9 with a fixed number of users (30), on the coverage and throughput. The results are presented in Table I. It is shown that as the number of LEDs in the room increases, the throughput increases and the coverage rises until covering approximately all users. When there are more than 7 LEDs, this is realized. In reality, as the number of LEDs rises, users have a greater probability of being covered, raising the metrics for throughput and coverage. In the same conditions, EGTO gives usually better performance in terms of throughput and coverage.

B. Effect of varying the number of users

In this case, We'll discuss the results of the eight remaining scenarios, which consists in the changing of the number of users from 5 to 40 with a fixed number of LEDs (30) on the coverage and throughput. The results are shown in Table I. It is observed that as the number of users in the room increases, the coverage and the throughput decrease. We can easily see that in most cases, EGTO performs better than other algorithms in terms of coverage and throughput.

VI. Conclusion

In this paper, we propose an Enhance Artificial gorilla troops optimizer, including chaotic map concept, for solving the LEDs placement problem in indoor visible light communication (VLC) system in an indoor room with dimensions 10 m × 10 m × 4 m. In our system, we have considered N LEDs and M users. The efficiency of the proposed EGTO is assessed considering the user coverage and throughput metrics in comparison with other six algorithms. The simulation results showed that EGTO obtained optimal LEDs layouts compared to other algorithms.

References

[1] Svilen Dimitrov and Harald Haas. *Principles of LED light communications: towards networked Li-Fi*. Cambridge University Press, 2015.

[2] Ying-Dar Lin. Third quarter 2018 ieee communications surveys and tutorials. *IEEE Communications Surveys & Tutorials*, 20(3):1607–1615, 2018.

[3] Panagiotis Botsinis, Dimitrios Alanis, Simeng Feng, Zunaira Babar, Hung Viet Nguyen, Daryus Chandra, Soon Xin Ng, Rong Zhang, and Lajos Hanzo. Quantum-assisted indoor localization for uplink mm-wave and downlink visible light communication systems. *IEEE Access*, 5:23327–23351, 2017.

[4] Musa Furkan Keskin, Sinan Gezici, and Orhan Arikan. Direct and two-step positioning in visible light systems. *IEEE Transactions on Communications*, 66(1):239–254, 2017.

[5] Suseela Vappangi and Venkata Mani Vakamulla. Synchronization in visible light communication for smart cities. *IEEE Sensors Journal*, 18(5):1877–1886, 2017.

[6] D Tagliaferri and C Capsoni. High-speed wireless infrared uplink scheme for airplane passengers' communications. *Electronics Letters*, 53(13):887–888, 2017.

TABLE I
COVERAGE, MEAN THROUGHPUT, AND FITNESS UNDER VARIOUS VALUES OF LED'S AND USERS

number of LEDs	2	3	4	5	6	7	8	9
Coverage (%)								
EGTO	58.8	**76.22**	88	**95.44**	**98.2**	**100**	**100**	**100**
GTO	60.44	72.88	80.22	88.88	94.44	97.22	96.88	99.22
MRFO	51.11	60.77	69.11	79	80.55	84.44	87.33	85.22
CHIO	**62.4**	73.84	**89.1**	92.61	98	**100**	**100**	**100**
MPA	58.7	63.45	73.1	85	95.8	97	100	100
BA	59.22	75.66	87.11	94.55	97.33	**100**	**100**	**100**
PSO	57.1	71.44	83.33	90	95.4	99	**100**	**100**
Mean throughput per user (Mbps)								
EGTO	0.69	1.32	1.74	2.94	**4.23**	**6.1**	**7.14**	8.33
GTO	0.65	1.30	1.59	1.94	2.48	2.48	4.26	4.87
MRFO	0.60	1.27	1.57	1.63	2.20	2.28	2.30	2.35
CHIO	**0.71**	**1.34**	1.59	2.15	2.84	3.23	4.87	5.1
MPA	0.61	1.15	1.43	1.86	2.3	2.57	3.7	4.21
BA	0.52	1.28	1.68	**3.07**	4.03	5.89	7.05	**8.86**
PSO	0.62	1.21	**1.89**	2.63	3.8	4.2	4.33	4.83
Fitness								
EGTO	0.36	**0.42**	0.45	**0.51**	**0.57**	**0.58**	**0.60**	**0.64**
GTO	0.30	0.30	0.41	0.45	0.48	0.50	0.51	0.52
MRFO	0.26	0.31	0.35	0.40	0.41	0.43	0.44	0.44
CHIO	**0.39**	0.41	**0.47**	0.48	0.53	0.55	0.57	0.61
MPA	0.27	0.33	0.37	0.44	0.46	0.47	0.49	0.51
BA	0.31	0.39	0.46	0.49	0.52	0.53	0.54	0.55
PSO	0.28	0.34	0.36	0.41	0.45	0.49	0.52	0.53

number of clients	5	10	15	20	25	30	35	40
Coverage (%)								
EGTO	**100**	**100**	**100**	**97.14**	**90.2**	83.45	79.33	76.11
GTO	**100**	**100**	96.74	91.64	85.14	80.85	75.54	73.46
MRFO	**100**	95.47	87.11	85.16	80	79.11	72.28	72
CHIO	**100**	**100**	**100**	95	88.14	85.11	79.3	**78.44**
MPA	**100**	94.8	90	84	81.5	77.14	73.11	72
BA	**100**	**100**	98.37	91.12	88.73	**86.44**	**81.33**	77.64
PSO	**100**	**100**	98.14	93.44	87.5	83.52	78.44	77.21
Mean throughput per user (Mbps)								
EGTO	**18.3**	12.44	**10.84**	7.71	4.44	3.33	2.63	2.1
GTO	13.35	9.73	7.70	5.21	3.50	2.09	1.84	**2.62**
MRFO	6.50	5.26	4.27	3.31	2.64	1.68	1.48	1.41
CHIO	17.4	**13.11**	10.2	7.31	4.14	3.11	2.87	2.24
MPA	7.4	6.81	6.1	3. 62	2.91	2.14	1.81	1.44
BA	16.41	12.88	9.08	6.09	4.42	3.19	2.83	2.56
PSO	14.3	9.45	9.03	**8.32**	**5.74**	**3.51**	**2.92**	1.84
Fitness								
EGTO	**0.69**	**0.65**	**0.59**	0.52	0.48	0.49	0.46	0.46
GTO	0.58	0.56	0.54	0.50	0.46	0.45	0.45	0.43
MRFO	0.54	0.53	0.46	0.44	0.41	0.40	0.40	0.38
CHIO	0.67	0.61	0.53	**0.54**	0.51	**0.50**	0.48	0.46
MPA	0.55	0.51	0.50	0.48	0.46	0.42	0.41	0.40
BA	0.60	0.58	0.56	**0.54**	**0.52**	0.49	**0.49**	**0.48**
PSO	0.63	0.59	0.53	0.50	0.49	0.47	0.47	0.45

[7] Vaneet Aggarwal, Zhe Wang, Xiaodong Wang, and Muhammad Ismail. Energy scheduling for optical channels with energy harvesting devices. *IEEE Transactions on Green Communications and Networking*, 2(1):154–162, 2017.

[8] Ishwar Ram Kumawat, Satyasai Nanda, and Ravi Maddila. *Positioning LED Panel for Uniform Illuminance in Indoor VLC System Using Whale Optimization*, pages 131–139. 01 2018.

[9] Shangsheng Wen, Xiaoge Cai, Guan W.P., Jiajia Jiang, Bangdong Chen, and Mouxiao Huang. High-precision indoor three-dimensional positioning system based on visible light communication using modified artificial fish swarm algorithm. *Optical Engineering*, 57:1, 10 2018.

[10] Rui Guan, Jin-Yuan Wang, Yun-Peng Wen, Jun-Bo Wang, and Ming Chen. Pso-based led deployment optimization for visible light communications. In *2013 International Conference on Wireless Communications and Signal Processing*, pages 1–6, 2013.

[11] Chaochuan Jia, Yang Ting, Wang Chuanjiang, and Sun Mengli. High-accuracy 3d indoor visible light positioning method based on the improved adaptive cuckoo search algorithm. *Arabian Journal for Science and Engineering*, pages 1–20, 10 2021.

[12] Abdelbaki Benayad, Amel Boustil, Yassine Meraihi, Selma Yahia, Sylia Mekhmoukh Taleb, Amylia Ait Saadi, and Amar Ramdane-Cherif. Solving the leds placement problem in indoor vlc system using a hybrid coronavirus herd immunity optimizer. *Journal of Optics*, pages 1–32, 2024.

[13] Benyamin Abdollahzadeh, Farhad Soleimanian Gharehchopogh, and Seyedali Mirjalili. Artificial gorilla troops optimizer: a new nature-inspired metaheuristic algorithm for global optimization problems. *International Journal of Intelligent Systems*, 36(10):5887–5958, 2021.

[14] Abdelbaki Benayad, Amel Boustil, Yassine Meraihi, Seyedali Mirjalili, Selma Yahia, and Sylia Mekhmoukh Taleb. An enhanced whale optimization algorithm with opposition-based learning for leds placement in indoor vlc systems. In *Handbook of Whale Optimization Algorithm*, pages 279–289. Elsevier, 2024.

[15] Muxuan. Li, Yancang. Han and Qinglin Guo. Modified whale optimization algorithm based on tent chaotic mapping and its application in structural optimization. *KSCE Journal of Civil Engineering*, 24, 2020.

[16] Hangqi Ding, Zhiyong Wu, and Luchen Zhao. Whale optimization algorithm based on nonlinear convergence factor and chaotic inertial weight. *Concurrency and Computation: Practice and Experience*, 32(24):e5949, 2020.

[17] Gaganpreet Kaur and Sankalap Arora. Chaotic whale optimization algorithm. *Journal of Computational Design and Engineering*, 5(3):275–284, 01 2018.

[18] Dharmbir Prasad, Aparajita Mukherjee, Gauri Shankar, and Vivekananda Mukherjee. Application of chaotic whale optimisation algorithm for transient stability constrained optimal power flow. *IET Science, Measurement & Technology*, 11(8):1002–1013, 2017.

[19] Diego Oliva, Mohamed Abd El Aziz, and Aboul Ella Hassanien. Parameter estimation of photovoltaic cells using an improved chaotic whale optimization algorithm. *Applied Energy*, 200:141–154, 2017.

[20] Yusuf Said Eroglu, Alphan Sahin, Ismail Guvenc, Nezih Pala, and Murat Yuksel. Multi-element transmitter design and performance evaluation for visible light communication. In *2015 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2015.

[21] Yassine Meraihi, Dalila Acheli, and Amar Ramdane-Cherif. Qos multicast routing for wireless mesh network based on a modified binary bat algorithm. *Neural Computing and Applications*, 31(7):3057–3073, 2019.

[22] Shahrzad Saremi, Seyedali Mirjalili, and Andrew Lewis. Biogeography-based optimisation with chaos. *Neural Computing and Applications*, 25(5):1077–1097, 2014.

Facial Expression Recognition for Mobile Robotics Control

1st Amine Rami BENNACER
Faculty of Electrical Engineering
Ecole Nationnale Polytechnique
Algeirs, Algeria
amine_rami.bennacer@g.enp.edu.dz

2nd Yacine HATEM
Faculty of Electrical Engineering
Ecole Nationnale Polytechnique
Algeirs, Algeria
yacine.hatem@g.enp.edu.dz

3[3d] Isma AKLI
Division Productic and Robotic
Centre de Developpement des Technologies Avancées
Algiers, Algeria
iakli@cdta.dz

***Abstract*—Artificial vision (AV) has significantly advanced over the past decade in robotics. This article presents an artificial vision system capable of detecting the facial expressions (emotions) of an individual using a camera, while controlling a mobile robot in both virtual and real environments. The vision-based artificial intelligence (AI) model detects the facial expressions of a human operator, and the mobile robot performs corresponding actions, such as moving straight or turning left/right. This application demands high accuracy and rapid processing for emotion detection. A pre-trained deep learning model, YoloV5, is employed for this purpose. The approach is validated in simulation and implemented on a real robot prototype. The software architecture is tested using ROS(Robots Operating System) framework, while the hardware consists of a mecanum-wheeled mobile robot controlled by an Arduino board.**

I. INTRODUCTION

In this article, we propose using a camera to detect human facial expressions in order to guide the behavior of a mobile robot. This approach requires the integration of Artificial Vision (AV) and Artificial Intelligence (AI), two disciplines closely related to computer science. AV leverages AI techniques to interpret video data. The advancement of AI—defined as "an engineered system that generates outputs such as content, forecasts, recommendations, or decisions for a given set of human-defined objectives" (ISO/IEC 22989:2022)—has significantly contributed to the progress of robotics, enabling computer systems to acquire capabilities similar to human intelligence. Similarly, the development of AV tools—defined as "the ability of a functional unit to succeed in processing and interpreting data representing images or video" (ISO/IEC 22989:2022)—has made it possible to apply effective and robust image recognition methods. These advancements enable robots to perceive their environment through visual sensors and determine the appropriate actions to take. In this article, we focus on AV applications for human emotion detection in robotics. This AI-driven approach is particularly relevant to service robotics, with applications in serious games. The concept can, however, be extended to other fields. We utilize the YOLO (You Only Look Once) tool [1] for facial expression detection, and we also employ the ROS framework to control the movements of a simulated mobile robot. Finally, we implement this solution on a real mecanum-wheeled robot, controlled by an Arduino prototyping board.

This paper is organized as follows: Section II provides an overview of related work in connection with the presented topic. Section III describes the architecture of the AV system. Section IV explains the motion generation process applied to both the simulated and real robots. Finally, Section V presents the results.

II. Related Work

Emotion recognition is an expanding area of research in robotics with various implications, particularly in Human-Robot Interaction (HRI), which remains one of the most important research areas. Generally, humans interacting with robots can express different types of emotions through facial expressions. The robot must perceive these emotions to improve interaction with humans.

Pino et al. [2] propose the use of a humanoid robot, Nao, to interact with elderly people. Software is implemented to train the robot to detect faces and smiles, while measuring the frequency and duration of the subjects' gaze. Melinte et al. [3] also explore Human-Robot Interaction with the NAO robot, employing deep convolutional neural networks (CNNs). They design two optimized CNNs for face recognition (FR) and facial expression recognition (FER). Rani et al. [4] propose emotion-sensitive human-robot cooperation frameworks, enabling the robot to modify its behavior based on emotional perception. Wearable biofeedback sensors detect and assess the anxiety level of a human collaborator, allowing the robot to respond to both implicit and explicit communication and adjust its actions accordingly. Liu et al. [5] present emotion recognition based on single-modal and multimodal cues for robotics, where robots perceive human emotions and respond using facial expressions, speech, gestures, etc. Marjorie et al. [6] investigate the use of emotional agents in decision-making processes for mobile robots, utilizing a fuzzy logic model to capture the inherent uncertainty of emotions. The agent makes decisions based on a combination of emotions generated by various states. Rajendran et al. [7] propose a voice emotion classification method, where the intensity level of emotions is calculated using machine learning models, and emotions are classified into four categories. This solution is specifically designed for assistive robots. Weng et al. [8] develop a visual-perception system for a dual-arm mobile robot, using a cognitive system to detect operator engagement and intentions. The intention model infers the operator's intention based on their emotional state.

Our contribution to this project involves implementing our model (YOLO) on ROS with a camera and creating a ROS node that sends messages containing the detected emotion. Based on these messages, the robot performs predefined actions.

III. Global Architecture of the AV System for Robotics

The following diagram provides a general representation of the different modules required to develop an AV system for robotics.

The image is captured by the camera and transmitted to the image processing module. The main attributes of the captured scene are then communicated to the AI model. Image recognition is performed, and the relevant data of interest is detected. The decision-making module processes this information and determines the appropriate behavior for the robot based on the output from the AI model. The corresponding command is then sent to the real or simulated robot.

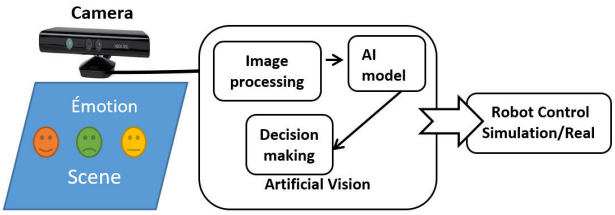


Fig. 1: General architecture of the AV system

The control of the mobile system, both in simulation and experimentally, requires the design of a ROS architecture Fig. 2 to address the presented issue. ROS is a robotic middleware that enables the design of a software system as a set of nodes that communicate with each other via topics. The real-time transfer of relevant information between nodes is performed based on the publish/subscribe paradigm.

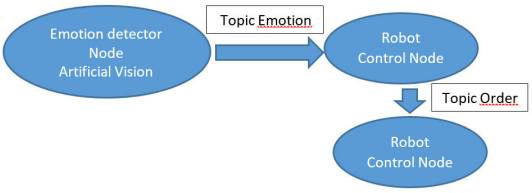


Fig. 2: ROS Software Architecture

There are three nodes communicating via topics. The first topic transmits the facial expression of

the human operator (topic: ”emotion”). The second topic transmits the movement command for the mobile robot, based on the information conveyed through the ”emotion” topic. Each detected emotion corresponds to a specific movement command.

Regarding the nodes, the emotion detector node is the artificial vision module, which uses an AI model to detect the perceived emotion of the human operator. The second node receives information from the ”emotion” topic and makes the decision regarding the command to be sent to the robot controller. The third node sends the command to the robot, which depends on the robot's control system (e.g., angular and linear velocity, wheel speed, etc.).

1) The AI YOLO Model: “You Only Look Once” (YOLO) is an object detection system [1]. YOLO is a real-time object detection algorithm that uses AI to localize and classify multiple objects in an image. YOLO is known for its speed and accuracy in real-time object detection. It can detect a variety of attributes in both images and videos. Depending on the application, YOLO can be trained using specific datasets. YOLO models are pre-trained on the COCO database, with the option to re-train for more specialized applications. In our work, we use a re-trained model available at (https://github.com/George-Ogden/emotion) Fig. 3, which detects faces and recognizes the corresponding emotion. This model is based on YOLOv5.

The following figure illustrates how the model works 3: The AI model receives a video as an

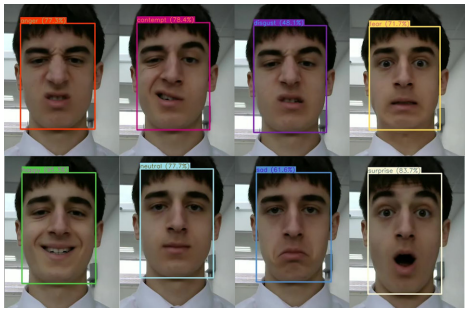


Fig. 3: Emotion Detection with YOLO v5 [9]

input source and performs real-time classification, providing an instant response to emotions detected on faces in the video. This model offers an effective solution for live emotion analysis, enabling a wide range of applications, including the one presented in this work. In this paper , we consider eight kinds of emotions as presented in Fig. 3

IV. Mobile robot motion generation

A. Simulation System

The mobile robot is simulated using the ROS Gazebo simulator. In the context of this work, any 3D model of a mobile robot is valid.

We propose using the TurtleBot simulation package to validate the communication between the robot controller and the emotion detection node. The mobile robot is differential drive. The system topic used is cmd_vel, and the control involves commanding both the angular and linear velocities.

B. experimental System

We experimented the proposed solution on a real mecanum wheeled mobile robot. This kind of robots is composed of four actuated wheels allowing moving in different directions regarding to the velocity and the sense of rotation of each actuator. We integrate an Arduino board into the mechanical



Fig. 4: Mecanum Wheeled mobile robot

system to send commands to the DC actuators of the four wheels. Communication between the robot and the PC is facilitated by a Bluetooth module (HC5). Two L298N H-bridges are required, as each one controls the movement of two actuators. Two batteries are needed: one dedicated to the Arduino board and the other to the H-bridges.

The robot's architecture requires us to utilize the rosserial library. This library facilitates communication between ROS nodes and the Arduino board,

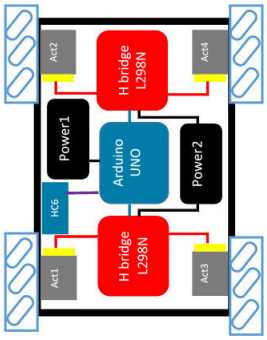


Fig. 5: electrical representation of the mobile robot

with the latter acting as a subscriber. Additionally, we used a Bluetooth HC-05 module to enable wireless communication, allowing for remote control and monitoring of the robot's operation.

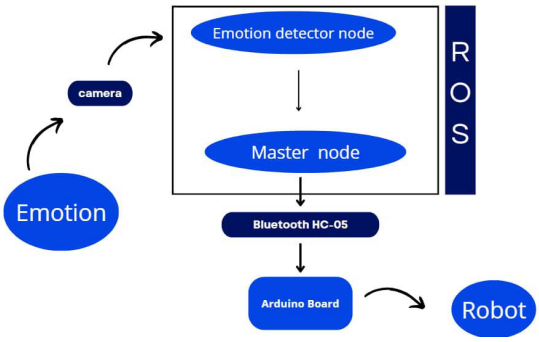


Fig. 6: hard/soft architecture for a emotion based mechnum wheeled mobile robot contol

V. Results

In this section we present some simulation and experimental results. Each emotion generates the movement of a simulated mobile robot and the goal is to reach the destination point. We used a model based on YOLO V5 to detect human emotions. As long as the system does not detect the emotions required to move the robot forward, it remains static. The operator transmits a succession of facial expressions in the form of emotion. Our proposed network takes an input as a black and white image of size 448 x 448. The results of the application look like this Fig. 7 :

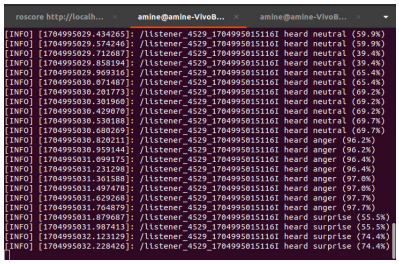


Fig. 7: emotion node transmitting the detected emotion in real time

the following figure shows the simulation results, where a mobile robot moves regarding the perceived facial expression Fig. 8 :

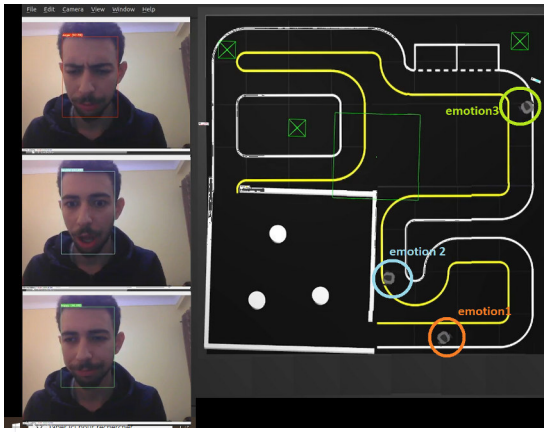


Fig. 8: simulation of the robot emotion controlled

To integrate robot motion with facial emotion recognition in a real-time application, we implemented an open-loop control system.

In this setup, we created a publisher node in ROS using Python. Based on the detected emotion, this node sends appropriate commands to the robot's actuators via the Bluetooth HC-05 module and the Arduino, enabling it to move in specific ways. Since our robot is a mecanum-wheel platform capable of performing advanced movements, the control was designed as follows:

Happy: The command instructs the robot to move forward. All actuators receive the same control signal in the positive direction.

Anger: The command instructs the robot to move backward. In this case, all actuators receive the same

control signal in the negative direction.

Neutral: The robot rotates on the spot. To achieve this, the right-side actuators receive a positive control signal, while the left-side actuators receive a negative one.

Sad: The robot performs a lateral (sideways) movement. For this, the diagonal actuators are driven in the positive direction, while the anti-diagonal actuators are driven in the negative direction.

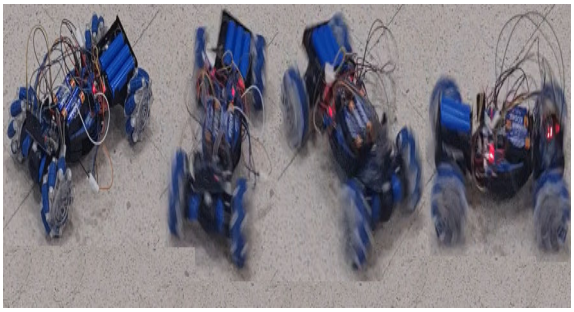


Fig. 9: circular on spot movement of the mobile robot

VI. Conclusion

In this article, we presented an AI and artificial vision application dedicated to mobile robotics. We showed that the YOLO tool is a very effective and easy-to-use. We showed an illustrative application highlighting the need to use AI tools. One of the perspectives of this work is an implementation for other kinds of robots and to adapt the proposed solutions to new applications.

References

[1] J. Redmon, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[2] O. Pino, G. Palestra, R. Trevino, and B. De Carolis, "The humanoid robot nao as trainer in a memory program for elderly people with mild cognitive impairment," *International Journal of Social Robotics*, vol. 12, pp. 21–33, 2020.

[3] D. O. Melinte and L. Vladareanu, "Facial expressions recognition for human–robot interaction using deep convolutional neural networks with rectified adam optimizer," *Sensors*, vol. 20, no. 8, p. 2393, 2020.

[4] P. Rani and N. Sarkar, "Emotion-sensitive robots-a new paradigm for human-robot interaction," in *4th IEEE/RAS International Conference on Humanoid Robots, 2004.*, vol. 1, pp. 149–167, IEEE, 2004.

[5] Z.-T. Liu, F.-F. Pan, M. Wu, W.-H. Cao, L.-F. Chen, J.-P. Xu, R. Zhang, and M.-T. Zhou, "A multimodal emotional communication based humans-robots interaction system," in *2016 35th Chinese Control Conference (CCC)*, pp. 6363–6368, IEEE, 2016.

[6] M. S. El-Nasr and M. Skubic, "A fuzzy emotional agent for decision-making in a mobile robot," in *1998 IEEE international conference on fuzzy systems proceedings. IEEE world congress on computational intelligence (Cat. No. 98CH36228)*, vol. 1, pp. 135–140, IEEE, 1998.

[7] H. Rajendran, H. R. T. Bandara, D. Chandima, and A. B. P. Jayasekara, "Voice response based emotion intensity classification for assistive robots," in *2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS)*, pp. 1–6, IEEE, 2023.

[8] W.-T. Weng, H.-P. Huang, Y.-L. Zhao, and C.-Y. Lin, "Development of a visual perception system on a dual-arm mobile robot for human-robot interaction," *Sensors*, vol. 22, no. 23, p. 9545, 2022.

[9] (https://github.com/George-Ogden/emotion. Accessed: 2023-12-10.

Application of Support Vector Machine Regression with Radial Basis Function Kernel in Non-Linear Function Approximation Using AI Techniques.

Hemza Gueddar
Electrical Engineering Department, University of Jijel
Jijel, Algeria
hamza01gueddar@gmail.com

Sid-Ahmed Touil
Institute of Electrical and Electronic Engineering
University M'hamed Bougara.
Boumerdes, Algeria
s.touil@univ-boumerdes.dz

Nidhal Cherrat
Dept. of Electrical Systems Engineering
ISE University of Boumerdes
Laboratoire d'Automatique de Jijel-LAJ, University MSBY
Boumerdes, Algeria
n.cherrat@univ-boumerdes.dz

***Abstract*— This paper explores the use of Support Vector Machine (SVM) regression with a Radial Basis Function (RBF) kernel to model non-linear relationships in data. The implementation is conducted using the Gaussian kernel, where epsilon ϵ and box constraint C parameters are optimized to balance fitting accuracy and generalization. This approach, a mainstay in Artificial Intelligence (AI) for regression tasks, provides robust performance on non-linear datasets due to its strong capacity for non-linear approximation and its resistance to overfitting. The paper evaluates the performance of this model through error metrics and discusses its relevance to AI applications requiring high-precision non-linear function estimation.**

Keywords—*Support Vector Machine, SVM regression, Radial Basis Function, non-linear function, metamodeling, function approximation*

I. Introduction

Artificial Intelligence (AI) has demonstrated remarkable capabilities in modeling complex, non-linear relationships across various scientific and engineering domains. Traditional approaches like polynomial and linear regression often fall short when dealing with intricate non-linear data patterns. To address these limitations, machine learning methods, particularly Support Vector Machine (SVM) regression with a Radial Basis Function (RBF) kernel, have emerged as powerful alternatives [1]. SVM regression, known for its robustness in handling non-linear data, leverages the "kernel trick," mapping input data into a higher-dimensional feature space where linear relationships can be modeled effectively.

The RBF kernel is particularly suited for capturing localized, non-linear variations in data, making it ideal for applications requiring precise approximation and prediction. Key parameters in this method epsilon ϵ and box constraint C are essential for balancing fitting accuracy and model generalization. Epsilon defines a margin of error where predictions are not penalized, thus reducing sensitivity to noise, while C controls the penalty for errors beyond this margin, directly influencing model complexity[2] .

In this paper, we demonstrate SVM regression with an RBF kernel by constructing a metamodel for the non-linear function. This function, frequently used in signal processing, serves as an ideal test case due to its non-linear and oscillatory nature. We utilize training samples randomly selected from the interval $x \in [-3\pi, +3\pi]$, assessing the SVM's ability to approximate such a challenging function. This approach exemplifies the effectiveness of SVM with an RBF kernel for complex, non-linear function modeling and highlights its applicability in AI-driven predictive modeling.

II. The Non-Linear Function

The target function $y = \mathrm{sinc}(x)$ is commonly used in signal processing and various approximation tasks. This function is non-linear and oscillatory, making it suitable for testing the capability of SVM regression to model complex relationships [2] [3] .

II.1 SVM Regression with RBF Kernel

To train the SVM regression model, we begin by randomly selecting n samples (xi, yi), where the inputs xi are drawn from a uniform distribution $U([-3\pi, +3\pi])$ and the target values are defined as $yi = sinc(xi)$. The Support Vector Regression (SVR) algorithm employs the following decision function:

$$f(x) = \sum_{i=1}^{L} \beta_i k(x_i, x) + b \quad (1)$$

Where

$k(x_i, x) = \exp(-\gamma |x_i - x|^2)$ is the RBF kernel with hyperparameter;

β_i are the support vector coefficients, and b is the bias term;

The parameters β_i and b are optimized during training to minimize the error between predicted and actual values in the training set, while ensuring smoothness and generalization.

II.2 Kernel Function and Hyperparameter Selection

The Gaussian (RBF) kernel function used here is defined as:

$$K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2) \quad (2)$$

where γ determines the smoothness of the kernel. The flexibility of the RBF kernel helps capture intricate non-linear relationships by adjusting γ appropriately. For SVM regression, optimal values of ϵ and C must be chosen through techniques such as grid search or cross-validation to ensure the model is neither overfitted nor underfitted [4] [5] .

II.3 Training and Optimization

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

The training process in SVM aims to find a model $f(x)$ that minimizes a regularized risk function. This function is defined as:

$$R(M) = \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} L\epsilon(yi - f(xi)) \quad (3)$$

Where :

$|w|^2$: Represents the norm (or magnitude) of the weight vector w. This term is responsible for controlling the smoothness of the model, ensuring that the solution doesn't overfit to individual data points.
C: Regularization parameter that balances the trade-off between model complexity and prediction error.
$L\epsilon$: ϵ-epsilonϵ-insensitive loss function, which determines the penalty for prediction errors outside the ϵ\epsilonϵ-margin.

II.3.1 ϵ-Insensitive Loss Function:

The ϵ-insensitive loss function is defined as:

$$L\epsilon(z) = max(0, | z | - \epsilon) \quad (4)$$

Where :

z: Represents the prediction error $yi - f(xi)$
ϵ: Defines a margin of tolerance around the true output yi

The regularization term C controls the model's complexity. A higher C value allows fewer errors, which improves model fitting but increases sensitivity to noise [6] - [7]. We apply cross-validation on ϵ\epsilonϵ and C to identify values that maximize prediction accuracy while retaining generalization.

II.4 Dataset

To train the SVM regression model, we randomly select 70 samples (x_i, y_i) where $(x_i \sim U([-3\pi, +3\pi]))$ and $y_i = \text{sinc}(x_i)$. The use of random samples allows the model to capture the variability of the function over the entire interval. Additionally, a validation set of 1000 evenly spaced samples within the same interval is generated to test the accuracy of the metamodel.

III. Results and Discussion

The SVM model is trained on the 70-sample dataset, with the RBF kernel's γ parameter and regularization term C optimized using cross-validation. After training, the model is evaluated on the validation set of 1000 samples to measure prediction accuracy.

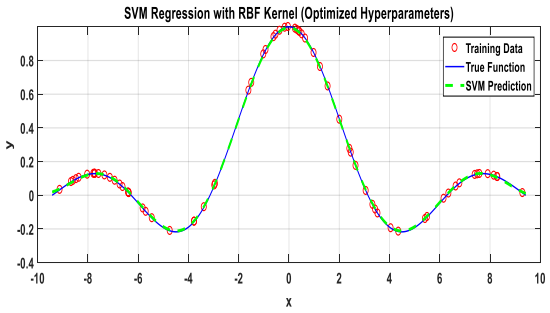


Fig.01 Non-linear function $\text{sinc}(x)$ predicted by SVM model regression.

The predicted values produced by the SVM model are visualized by plotting them against the actual function values in Figure (1), allowing for a direct comparison. –This graphical representation provides valuable insights into how well the model approximates the underlying function. Specifically, it highlights the model's effectiveness in capturing the oscillatory nature of the sinc function, which exhibits rapid fluctuations, especially in regions where the function approaches zero. These areas, often referred to as the "nulls", are crucial for assessing the model's precision in handling small values. Additionally, the visualization displays how accurately the model tracks the peaks of the function, which are characterized by sharp, localized increases. By examining these features, the plot not only demonstrates the SVM model's ability to capture intricate details of the sinc function but also offers a clear indication of its performance across both low and high amplitude regions. This helps to assess the model's robustness in modeling non-linear behavior with precision and consistency

III.1 Model Performance

After constructing the SVM regression model, we compare its predictions against the true $\text{sinc}(x)$ values in the validation set. Performance metrics include Mean Absolute Error (MAE), Root Mean Square Error (RMSE) , and the coefficient of determination R^2,which provides a quantitative measure of how well the model captures the data's variance. The results were as follows:

TABLE I. PERFORMANCE COMPARISON

Metric	***Value***
MAE	0.00010528
RMSE	0.00011419
R^2	1

These results demonstrate that the SVM model with an RBF kernel provides an accurate and robust fit for non-linear data.

III.2 Study Effect of Epsilon in SVM Regression

In Support Vector Regression (SVR), the parameter epsilon (ε) defines the width of the epsilon-insensitive zone, within which errors are not penalized. Adjusting ε has a significant impact on the sparsity of the SVM solution and the accuracy of the model:

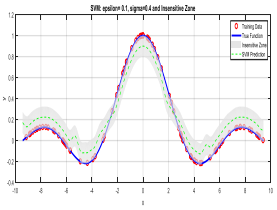


Fig.02 Non-linear function $\mathrm{sinc}(x)$ predicted by SVM model regression with epsilon=0.1

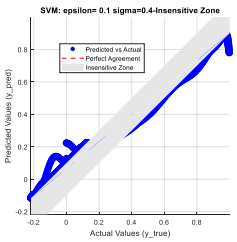


Fig.03 Scatter plot comparing the correlation between the actual values of y and the model prediction by SVM model

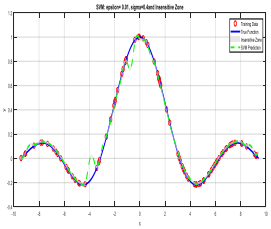


Fig.04 Non-linear function $\mathrm{sinc}(x)$ predicted by SVM model regression with epsilon= 0.01

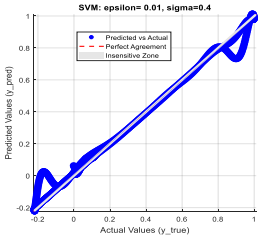


Fig.05 Scatter plot comparing the correlation between the actual values of y and the model prediction by SVM model

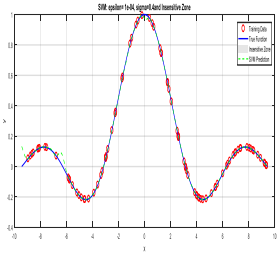


Fig.06 Non-linear function $\mathrm{sinc}(x)$ predicted by SVM model regression with epsilon =0.00001

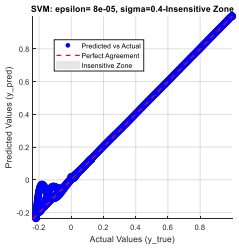


Fig.07 Scatter plot comparing the correlation between the actual values of y and the model prediction by SVM model

Increasing epsilon ε: As the epsilon-insensitive zone widens, the model tolerates larger deviations between predicted and actual values without considering them as errors. This leads to fewer support vectors being selected, making the model sparser. However, this increased tolerance may result in reduced prediction accuracy, as finer details in the data are ignored.

Decreasing epsilon ε: A smaller epsilon makes the model more sensitive to deviations, as it tries to minimize even small errors. This often improves accuracy but requires more support vectors, resulting in a less sparse solution and potentially higher computational costs.

Through experimentation, it is observed that reducing ε generally yields better predictive performance since the model captures finer patterns in the data. However, there is a trade-off between accuracy and model complexity that must be carefully balanced depending on the application.

IV. Conclusion

In this study, we constructed a reliable Support Vector Machine (SVM) -based metamodel with a limited, randomly sampled dataset, demonstrating that even with modest data resources, SVM regression with an RBF kernel can achieve accurate approximations of complex functions. This methodology is broadly applicable to any domain where precise non-linear function modeling is needed, underscoring its versatility and effectiveness in AI-driven predictive modeling.

SVM regression with an RBF kernel has proven to be a powerful method for approximating non-linear functions, as demonstrated by the robust performance observed in this study. Its ability to capture complex, non-linear relationships with high precision and minimal overfitting makes it a valuable tool for AI applications across a variety of fields. Although hyperparameter tuning presents a challenge, particularly with parameters like ϵ\epsilonϵ and C, cross-validation techniques have effectively identified optimal values, ensuring a well-balanced trade-off between model accuracy and generalization.

V. Future Work

Future studies may extend this work by experimenting with different kernel functions, adding noise to the dataset to test robustness, or using a larger sample size to further improve accuracy.

References

[1] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273-297, Sept. 1995.

[2] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.

[3] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Stat. Comput.*, vol. 14, no. 3, pp. 199-222, Aug. 2004.

[4] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. New York, NY, USA: Springer-Verlag, 1999.

[5] L. Wang, *Support Vector Machines: Theory and Applications*. Berlin, Germany: Springer-Verlag, 2005.

[6] J. Weston and C. Watkins, “Multi-class support vector machines,” presented at the *ESANN 1999 - Eur. Symp. Artif. Neural Netw.*, Brussels, Belgium, Apr. 1999.

[7] MathWorks, “fitrsvm: Regression SVM model with Gaussian kernel,” MATLAB Documentation. [Online]. Available: https://www.mathworks.com. [Accessed: Oct. 13, 2024].

6G Networks: A Vision for the Future and Comparative Analysis with 5G

1st TEMERICHT IMANE
University of abderrahmane Mira Bejaia
Electrical Engineering Laboratory of bejaia
(LGEB)
imane.temericht@univ-bejaia.dz

2nd H. Bellahsene
University of abderrahmane Mira Bejaia
Electrical Engineering Laboratory
hocine.bellahsene@univ-bejaia.dz

3rd A.Taleb-Ahmed
Electrical Engineering Laboratory
IEMN DOAE Laboratory
abdelmalik.aaleb-ahmed@uphf.fr

***Abstract*—Wireless communication systems have become a major focus of research due to their critical role in various fields. As fifth-generation (5G) networks continue to be deployed, the research and development community has begun to focus on the next evolution of mobile communication systems, known as the sixth generation (6G). The planning and conceptualization of 6G aim to address the communication needs anticipated for 2030 and beyond. This paper presents a vision for future 6G wireless communication, highlighting potential applications, the communication requirements they entail, and the emerging technologies that could make them a reality. Additionally, it includes a theoretical performance comparison between 5G and 6G networks in a scenario where connected devices require high data rates.**

***Index Terms*—6G, 5G, wirless communications, applications, data rate, latency.**

I. Introduction

The evolution of mobile networks has progressed steadily, with a new generation emerging approximately every decade. While the fifth-generation (5G) networks are now being widely deployed and entering their consolidation phase, research efforts have already shifted towards the development of sixth-generation (6G) networks [2]. By 2030, 5G is expected to reach its technical limits, unable to fully support the demands of future applications requiring highly automated and intelligent networks.

6G networks are designed to address these challenges by surpassing 5G's capabilities. They aim to create a seamless integration of the human, physical, and digital worlds, forming a unified cyber-physical continuum. This vision includes enabling revolutionary applications such as immersive real-time holograms, autonomous vehicles, and universal connectivity. 6G is expected to deliver unprecedented features, including significantly higher bandwidth, ultra-low latency, and intelligent adaptability to ensure the efficient and effective management of advanced services [5].

In this paper, we outline a vision for 6G wireless communication, emphasizing its transformative potential. We discuss its key applications, the communication requirements these applications entail, and the emerging technologies necessary to realize them [1]. Additionally, we provide a theoretical performance comparison between 5G and 6G networks through simulations of two critical scenarios: real-time holograms, requiring massive data rates and low latency, and autonomous vehicles, demanding moderate data rates but ultra-low latency. The simulation results highlight the advantages of 6G. With its extended bandwidth and enhanced capabilities, 6G can sustain significantly higher data rates and much lower latency, even in scenarios with a large number of connected devices. For example, 6G can deliver approximately 1 Gbps per device in high-demand situations, compared to only 100 Mbps with 5G. This demonstrates that 6G is well-equipped to meet the demands of next-generation applications. Finally, we emphasize that this analysis represents just one of many possible approaches to exploring the capabilities of 6G. Further research into different scenarios and use cases will be essential to fully understand the versatility and transformative potential of this next-generation technology. Through its innovative features, 6G is set to become a cornerstone of future networks, driving advancements in digital transformation and improving quality of life globally.

II. 6G requirements

6G networks are expected to be more intelligent, capable, reliable, and power-efficient to satisfy all the expectations that can not be realized with 5G [3]. Most of the key performance in dicators KPIs for 6G networks that describes 6G requirements are as expected an evolution of the KPIs for 5G networks and they are represented as follows [6]:

- Spectrum efficiencies 5 to 10 times higher than those of 5G
- Peak data rate: New services such as extended reality and holographic communications will in 6G communi cations to bring batter user experience. These services will demand a higher data rate, So 6G network should be designed to support much higher data rate as Tbps level.
- User experienced data rate: To ensure that users can obtain high-quality experience anytime and any where,wireless system must deliver high data rate, thus

for individual users data rate is expect to increase to at least 10Gbps, and up to 100Gbps
- latency: To overcome the boundaries between the real environment and the computational space, extremely low latency (0.1-1ms) is required.
- High mobility: it is expected to enable high mobility up to 1000 km/h and over the air.
- Huge connection density: up to 107 devices/km
- Energy efficiency is defined as the number of bits which can be transmitted over 1 Joule. The energy consumption for 6G networ should be 10–100 times higher than in 5G.
- reliability: The reliability requirements of 6G network are expected to be use case specific.Some services such as industrial control and remote surgery requires a very high reliability. Therefore, the reliability of 6G network transmission should go to 99.9999%

The comprehensive network performance requirements of 6G versus 5G are given in Table 1:

key performance indicators	5G KPIs	6G KPIs
operating frequency	3GHz-3OOGHz	1 up to 1 THz
peak data rate	20Gbps	1Tbps
latency	1ms	10-100µs
mobility	500Km/h	up to 1000Km/h
reliability	99.9%	about 99.999%
coverage	about 70%	99%
energy efficiency	high	ultra-high
spectral efficiency	30 bps/Hz	100 bps/Hz

Table1: Possible capabilities of 6G in comparison with 5G

III. USAGE SCENARIOS OF 6G

Similar to 5G classifications, 6G use cases that have similar requirements can be categorized into groups, or specific use cases [5]. Some scenarios predicted for 6G networks are described as an improvement of the scenarios defined for 5G Networks

- Ultra mobile broadband (uMBB): The use cases under this scenario require much higher data rate than the one required in 5G eMBB (enhanced Mobile Broadband) usage scenario [5].
- Ultra-massive machine-type communication (uMTC): The use cases under this scenario require much more massive number of simultaneous connections per space than the one required in 5G massive machine-type com munication (mMTC) usage scenario [5].
- ultra high speed with low latency communications (uH-SLLC): The use cases under this scenario provides a higher speeds and lower latency than the one required in 5G ultra-reliable low-latency communications (URLLC) usage scenario.
- mobile broadband reliable low latency communication (MBRLLC): new service class proposed, it generalizes classical URLLC and eMBB services. Energy efficiency is central for MBRLLC, not only because of its impact on reliability and rate, but also because of the resource limited nature of 6G devices.
- Human-Centric Services (HCS): a new class of 6G services, network performance is determined by the physi ology of the human users and their actions

IV. 6G FUTUR APPLICATIONS

Network technology is advancing towards a more flexible and agile infrastructure, and it is likely that 6G, like its predecessors, will be shaped by today's emerging technologies. 6G will emerge from an unprecedented wave of innovative applications that are set to transform human society in the 2030s. Until now, communication systems have primarily transmitted data interacting with two human senses: hearing and sight. One of the key ambitions of 6G is to expand this capability by transmitting data associated with other senses, such as touch, smell, and taste. Below is a list of potential 6G applications:

A. Internet of everything IOE

IOE is an extended version of IOT (internet of things) that includs things, data, people and processes. the main concept of IOE is to integrate various sensing devices that can be related to everything. 6G is expected to become a key enabler for IOE, that will be helpful in improving the services related internet of things. this also can mean that, many new applications will emerge from the fusion of IOE and 6G communication [7].

B. Holographic communication

6G is expected to be a conversion point from the traditional video conferencing to a virtual in-person meeting [2]. For this holographic communication is needed for its capabilities to project realistic, full motion, real-time three dimensional (3D) images for distant people and objects with huge high level of realism rivaling of the physical presence. In fact, transferring three dimensional images along the voice is not sufficient to convey the in person presence, so the use of three dimensional video with stereo audio is needed. Holographic communication will make people's lives easier in many areas like minimizing business travels costs, distance learning and precision medicine.

C. Extended Reality

Extended Reality (XR) is an umbrella term that encompasses all immersive technologies that merge the physical and digital worlds. This includes Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR). In the context of 6G, XR is expected to reach new heights, offering ultra-realistic and seamless experiences through advancements in connectivity, latency, and computational power. XR applications in 6G

could revolutionize industries such as education, healthcare, entertainment, and manufacturing by enabling more interactive and immersive environments.

D. Healthcare

It is expected that 6G communication technology will revolutionize the healthcare completely and healthcare will fully depend on 6G [4]. Because innovations such as XR, holographic communications, tactile internet and intelligent robots in 6G, the robotic controlled surgery, remote surgical procedures and enhanced precision surgical capabilities will be realized. This will build a smart healthcare system.

E. Connected and Autonomous Vehicles

6G is expected to play a crucial role in the development and deployment of connected and autonomous vehicles (CAVs). By providing ultra-low latency, high-speed communication, and massive network capacity, 6G will enable vehicles to communicate with each other (V2V), infrastructure (V2I), and pedestrians (V2P) in real-time. This will enhance the safety, efficiency, and coordination of autonomous systems, allowing for smoother traffic flow, faster decision-making, and better accident avoidance. Additionally, 6G's capabilities will support advanced features such as real-time mapping, predictive maintenance, and vehicle-to-cloud integration, enabling fully autonomous vehicles to operate more safely and effectively, while also paving the way for smarter, more sustainable transportation systems.

F. collaborative robots

also called cobots, are directly collaborate with people by work side by side with humans. These cobots take over tedious, repetitive and risky tasks to maintain human worker's health and safety and automate the production lines. They also can offer various benefits inn industry such as reliability, safety and trust.



Fig. 1. emerging 6G applications

Figure [1] illustrates the key emerging applications of 6G. These applications push the boundaries of current networks and highlight the potential of 6G to transform various sectors such as industry, healthcare, and transportation.

V. ENABLING TECHNOLOGIES

The services and requirements of 6G presented in the previous sections indicate various challenges to the development of the future wireless system [7]. Some new technologies will be added and some 5G technologies will be improved in 6G. Hence this network system will be driven by many technologies, a few expected vital technologies for 6G are presented below:

A. Terahertz (THz) Communication

6G is the first wireless com munication system with ultra high speed communication. This is why the use of Terahertz radiation (THz) also called sub millimeter radiation is required to support features that demand ultra-high data rates (several Tbps) and extremely low latency [3].

B. Artificial Intelligence AI

intelligence is the fundamental characteristic of 6G autonomous networks 1. AI will play a central role in managing 6G networks, optimizing resource allocation, enabling real-time decision-making, and ensuring networks are self-healing and adaptive.

C. Intelligent Reflecting Surfaces (IRS)

Intelligent Reflecting Surfaces (IRS) are a key enabler for 6G networks, consisting of programmable surfaces that control and optimize wireless signals. IRS enhances coverage, energy efficiency, and network capacity by dynamically directing signals, especially in challenging environments like urban areas or indoors [3].

D. Advanced Edge Computing in 6G

Advanced Edge Computing in 6G brings computational power closer to users and devices, reducing latency and enabling real-time processing. It supports applications like autonomous vehicles, smart cities, and extended reality, while integrating AI for intelligent decision-making and energy efficiency. This ensures faster responses, enhanced security, and better handling of massive data demands.

VI. DEVICES AND CONNECTED MACHINES FOR 6G

By 2030, the number of connected devices is expected to reach 500 billion, approximately 59 times the projected world population of 8.5 billion [8]. Beyond smartphones and tablets, new human-machine interfaces, such as AR glasses, VR headsets, and holographic devices, will make accessing and controlling information more convenient. Connected machines, including vehicles, robots, drones, smart sensors, and industrial equipment, will become the primary users of 6G networks. To meet these demands, new technologies will need to be developed to connect hundreds of billions of machines effectively [2]

VII. SIMULATION RESULT

To showcase the potential advantages of 6G in addressing the challenges of demanding use cases, we conducted simulations comparing the theoretical performance of 5G and 6G networks based on two critical metrics: data rate per device, determined by the total bandwidth (1 GHz for 5G and 10 GHz for 6G) shared among an increasing number of connected devices, and latency, modeled as the transmission delay per device while considering the impact of network saturation. The main assumptions in the simulation included dedicated bandwidth and constant spectral efficiency, with the results analyzed for a range of devices from 10 to 100 across two critical use cases: real-time holograms, which require massive data rates (1 Tbps) and low latency for a seamless user experience, and autonomous vehicles, which demand moderate data rates (100 Mbps) but ultra-low latency for real-time responsiveness.

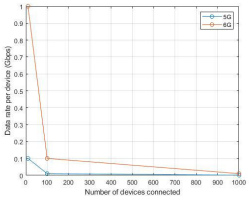


Fig. 2. Data rate per device as a function of the number of devices

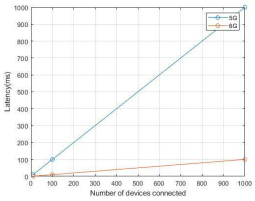


Fig. 3. Latency as a function of the number of devices

The results as presented in Fig. 2 and Fig. 3, clearly demonstrate that 6G outperforms 5G in both scenarios. Thanks to its massive bandwidth (10 GHz), 6G sustains superior data rates even with a high number of connected devices. For instance, with 100 devices, each achieves approximately 1 Gbps, compared to only 100 Mbps with 5G. Additionally, 6G achieves significantly reduced latency, enabling faster data processing, which is crucial for real-time critical decision-making in applications like autonomous vehicles. These findings underline 6G's ability to meet the demands of future applications where 5G reaches its technical limits, making it a pivotal technology for next-generation networks.

VIII. CONCLUSION

The tradition of deploying a new cellular generation approximately every decade will continue in the future, with 6G becoming a reality in the 2030s. Many new use cases can be expected, driven by societal demands and realized through emerging 6G technologies. In this paper, we presented a general vision of 6G, including its requirements, applications, and the enabling technologies that can bring these applications to life. We also provided simulation results that illustrate, in a simple and effective manner, the revolutionary potential of 6G to meet the demands of future applications while overcoming the technical limitations of 5G. Through this example, which compares use cases such as real-time holograms and autonomous vehicles, we demonstrated the significant advantages of 6G's extended bandwidth and reduced latency. It is important to note that this analysis represents just one approach among many possible ways to explore the capabilities of 6G. Other scenarios could be studied to deepen our understanding and confirm the versatility of this technology across various application domains. Thus, 6G emerges as a key technology to realize the vision of next-generation networks and support the digital transformation of the future.

REFERENCES

[1] M. Z. Chowdhury, Md. Shahjalal, S. Ahmed and Y. M. Jang "6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges,and Research Directions," IEEE Open J. Commun.Soc., vol. 1, pp.957-975,2020.

[2] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" Nat. Electron., vol. 3, no. 1, pp. 20–29, 2020

[3] S. Alraih, I. Shayea, M. Behjati, R.Nordin, N. F. Abdullah, A.Abu-Samah and D. Nandi, " Revolution or Evolution? Technical Requirements and Considerations towards 6G Mobile Communications," Sensors 2022, vol. 22(3), 762

[4] S. Nayak and R. Patgiri, " 6G Communication Technology: A Vision on Intelligent Healthcare, " IEEE internet of things journal, 2020.

[5] C. Yeh G. Do Jo, Y. Ko, H. Chung,"Perspectives on 6G wireless communications," ICT express, on line: January,05,2022. in press.

[6] H.VISWANATHAN,and P. E. MOGENSEN, "Communications in the 6G Era" IEEE access, vol. 8, pp. 57063- 57074, 2020.

[7] C. DE ALWIS, A. KALLA, Q. PHAM, P. KUMAR, k. DEV, W.HWANG, and M. LIYANAGE, "Survey on 6G Frontiers: Trends,Applications, Requirements, Technologies and Future Research" IEEEopen J. Commun. Soc., vol 2, 2021.

[8] Samsung researches"The Next Hyper Connected Experience for All, Samsung 6G vision", 2020.

Snake Optimizer for solving the UAV Base stations Placement Problem

1st Rachda El Taani
LIST Laboratory
University of M'Hamed Bougara Boumerdes
Avenue of Independence, 35000 Boumerdes, Algeria
r.eltaani@univ-boumerdes.dz

2nd Karim Baiche
LIST Laboratory
University of M'Hamed Bougara Boumerdes
Avenue of Independence, 35000 Boumerdes, Algeria
kbaiche@univ-boumerdes.dz

4th Yassin Meraihi
LIST Laboratory
University of M'Hamed Bougara Boumerdes
Avenue of Independence, 35000 Boumerdes, Algeria
y.meraihi@univ-boumerdes.dz

5th Sylia Mekhmoukh
LIST Laboratory
University of M'Hamed Bougara Boumerdes
Avenue of Independence, 35000 Boumerdes, Algeria
s.mekhmoukh@univ-boumerdes.dz

***Abstract*—Unmanned Aerial Vehicles (UAVs) acting as mobile base stations, or UAV-BSs, are becoming an increasingly viable solution for enhancing mobile communication, especially in scenarios like emergencies or network failures. UAVs can be quickly deployed to maintain connectivity when traditional terrestrial base stations are compromised or overloaded. This research focuses on optimizing the placement of UAV-BSs to maximize downlink coverage and minimize energy consumption, considering both user distribution and drone altitude. To tackle this complex optimization problem, we propose the use of the Snake Optimizer (SO), a recent metaheuristic inspired by the dynamic movement of snakes. The SO integrates both exploration and exploitation strategies to effectively search for optimal solutions. Through several test scenarios, our findings reveal that SO outperforms existing optimization algorithms, including the Marine Predators Algorithm (MPA), the Whale Optimization Algorithm (WOA), and the Grey Wolf Optimization (GWO), providing an efficient and reliable method for UAV placement in static environments.**

***Index Terms*—UAV base station, Snake Optimizer, Energy Efficient, Downlink coverage.**

I. Introduction

The landscape of global communication is evolving rapidly, driven by emerging technologies that promise to redefine connectivity. Among the most groundbreaking innovations are Unmanned Aerial Vehicles (UAVs) functioning as flying base stations (UAV-BSs). These adaptable aerial platforms are not only transforming mobile networks but also paving the way for a new era of global connectivity. With the rise of the Internet of Things (IoT), smart cities, and precision agriculture, the demand for flexible, high-performance communication networks has become more critical than ever. UAV-BSs offer a potential solution to overcome the connectivity challenges posed by traditional infrastructure, particularly in remote areas or during disaster recovery.

As IoT devices proliferate, everything from environmental sensors to agricultural tools is becoming interconnected. In this context, UAV-BSs play an increasingly essential role by enabling rapid deployment of communication networks. Whether it's for disaster response, providing connectivity to rural regions, or supporting high-bandwidth applications like real-time surveillance and IoT systems, UAV-BSs offer unprecedented flexibility in meeting the demands of modern networks [1], [2], [3]. However, despite their promise, integrating UAV-BSs into existing network infrastructures presents substantial challenges.

The growing demand for reliable, efficient wireless networks is exacerbated by the exponential rise in IoT devices. UAV-BSs are well-positioned to address gaps in coverage, but their deployment requires careful optimization. Each UAV operates with a limited energy supply, and the challenge lies in balancing coverage expansion with energy conservation. Although extending coverage often requires higher transmission power, doing so rapidly depletes battery life. Thus, the challenge becomes determining the optimal placement of UAVs to ensure maximum user coverage while minimizing energy consumption.

This UAV placement problem is recognized as NP-hard, meaning traditional optimization methods struggle to handle its complexity. As a result, researchers have increasingly turned to metaheuristic optimization techniques, which are effective for solving such high-dimensional, complex problems. Over the past decade, various approaches have been proposed, including the Grey Wolf Optimizer (GWO) for 3D placement [4], Particle Swarm Optimization (PSO) and Exhaustive Search (ES) for optimizing Quality of Service (QoS) [5], and Genetic Algorithms (GA) for 2D UAV positioning [6]. More recently, hybrid approaches such as combining the Whale Optimization Algorithm (WOA) with Simulated Annealing (SA) have been explored to improve the accuracy of UAV placement [7].

In recent years, numerous metaheuristic algorithms inspired by the behavior of viruses in nature have been developed. One such algorithm, the Snake Optimizer [8], has demonstrated its

effectiveness across various application domains.

The Snake Optimizer (SO), inspired by the foraging behavior of snakes, offers a fresh and powerful solution to this challenge. The dynamic and adaptive nature of snake movement allows for efficient exploration and exploitation of the search space, making it well-suited to solve complex problems like UAV placement. By balancing these two strategies, SO can identify optimal UAV positions that maximize coverage and minimize energy consumption.

In this paper, we apply the Snake Optimizer to solve the UAV placement problem in a static environment. The approach harnesses the algorithm's adaptive nature to find the optimal UAV locations, achieving a balance between coverage and energy efficiency.

The remainder of this paper is organized into five sections: Section II presents the formulation of the UAV placement problem, including both the system and mathematical models. Section III offers an overview of the Snake Optimizer. The application of the SO to the UAV placement problem is thoroughly explained in Section IV. Section V outlines the simulation results, and Section VI provides the conclusion of the study.

II. UAV-BSs Placement Problem Formulation

A. System Model

In this study, the UAV placement problem is modeled in an urban setting, with the area represented as a three-dimensional space measuring $W \times L \times H$ in meters. The coordinates of the area are illustrated in Fig. 1. The model includes:

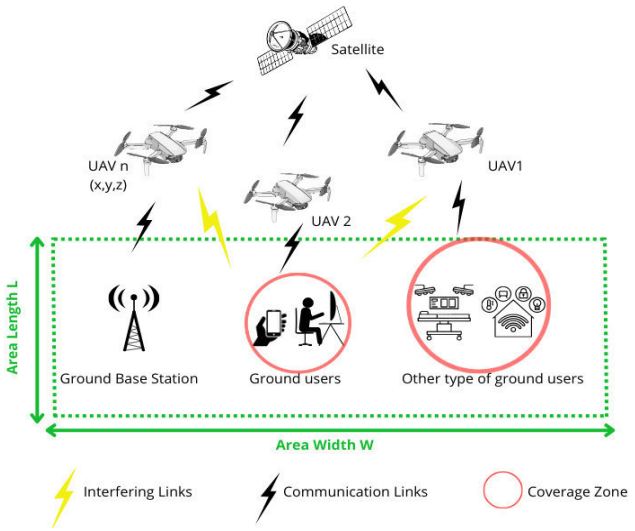


Fig. 1. System model

- A set of N UAVs, denoted as D, where each UAV D_j is positioned at coordinates (x_j, y_j, h_j), with h_j representing the drone's altitude, as shown in Fig. 1. The altitude is determined based on the coverage radius and visibility angle required for optimal coverage and energy efficiency. Each UAV is equipped with a radio interface to facilitate communication with ground users and other UAVs.
- A set of M users, denoted as $U = \{U_1, U_2, ..., U_M\}$, is randomly distributed across a two-dimensional area. To ensure optimal coverage, each user is paired with the nearest drone. A drone D_j can cover a user U_i if the user lies within the drone's coverage area R_j.

The altitude of each UAV is adjusted to achieve a balance between coverage and energy consumption. The energy consumption of a drone is directly proportional to its altitude, and is given by the formula:

$$E_{D_j} = M_{D_j} \cdot g \cdot h_j \tag{1}$$

where M_{D_j} represents the total weight of the drone, and g is the acceleration due to gravity.

B. Mathematical Model

The primary goal of this optimization is to enhance the coverage quality for ground users while simultaneously minimizing the energy consumption of the UAVs. As previously noted, increasing the altitude of a UAV boosts its coverage radius but also results in higher energy consumption. Therefore, the challenge lies in striking the right balance between these two factors for optimal UAV placement. The mathematical formulation of this problem can be described as follows:

The objective is to maximize the overall coverage quality across all users, as expressed in equation (2):

$$\max f(\gamma) = \sum_{i=1}^{n} \max_{j \in \{1,\dots,m\}} (\gamma_{U_i}^{D_j}) \tag{2}$$

Here, $\gamma_{U_i}^{D_j}$ represents the coverage quality for user U_i provided by drone D_j, which is determined by the distance between the user and the drone. This relationship is quantified in equation (3):

$$\gamma_{U_i D_j} = \max\left(0, \max(R) - d(U_i, D_j)\right) \tag{3}$$

where $\max(R)$ is the maximum coverage radius that depends on the drone's altitude and the visibility angle, while $d(U_i, D_j)$ is the Euclidean distance between the user U_i and the drone D_j.

To regulate the UAV's altitude, the UAV must operate within a predefined height range, as shown in equation (4):

$$h_{\min} \leq h_j \leq h_{\max} \tag{4}$$

Moreover, the placement of each UAV must ensure that it covers its assigned users. This constraint is enforced in equation (5), where the distance between any user U_i and a UAV D_j must not exceed the maximum coverage radius $R_{\max}$:

$$d(U_i, D_j) \leq R_{\max}, \quad \forall i \in \{1, \dots, M\}, \forall j \in \{1, \dots, N\} \tag{5}$$

Finally, the UAVs must be placed within the boundaries of the deployment area, which is restricted to the dimensions of the urban environment as described in equation (6):

$$0 \leq x_j \leq W \text{ and } 0 \leq y_j \leq L \tag{6}$$

Thus, the mathematical model incorporates the key parameters influencing both coverage quality and energy efficiency, ensuring that the placement of UAVs satisfies all the required constraints.

III. Overview of the Snake Optimizer (SO)

The Snake Optimizer (SO) is a metaheuristic algorithm inspired by the natural behaviors of snakes, particularly their mating, competitive, and survival strategies [8]. The algorithm emulates these dynamics using environmental factors such as temperature and food availability to guide the search process. By balancing exploration (broadly searching the space) and exploitation (refining solutions near promising regions), SO achieves robust performance on complex optimization problems. Below is an overview of its methodology.

A. Biological Inspiration

Snakes exhibit complex behaviors influenced by environmental factors. Mating typically occurs in colder regions with sufficient food availability. Males compete through fights to attract females, and females select the most suitable mates. This results in the generation of offspring (eggs), symbolizing diversity in the search space. When conditions are unfavorable, such as the absence of food or unsuitable temperatures, snakes prioritize survival, focusing on searching for food. These behaviors inspire the phases and mechanisms of the SO algorithm, which adapts dynamically to the optimization landscape.

B. Initialization

SO begins by initializing a random population of solutions (representing snakes) within the search space, defined by the problem's bounds. Each solution is expressed as:

$$X_i = X_{\text{min}} + r \cdot (X_{\text{max}} - X_{\text{min}}) \tag{7}$$

where X_i is the position of the i-th individual, r is a random number between 0 and 1, and $X_{\text{min}}, X_{\text{max}}$ are the lower and upper bounds, respectively.

C. Population Division

The population is divided into two groups: males and females, ensuring equal representation for mating and fighting phases. The number of individuals in each group is calculated as:

$$N_m = \left\lfloor \frac{N}{2} \right\rfloor, \quad N_f = N - N_m \tag{8}$$

where N is the total number of individuals, N_m is the number of males, and N_f is the number of females.

D. Environmental Factors

Two key factors dynamically influence the algorithm's behavior: temperature ($Temp$) and food quantity (Q).

1) Temperature: Temperature determines the likelihood of mating and fighting phases. It decreases over iterations, mimicking environmental cooling:

$$Temp = \exp\left(-\frac{t}{T}\right) \tag{9}$$

where t is the current iteration and T is the maximum number of iterations.

2) Food Quantity: Food quantity guides the transition between exploration and exploitation:

$$Q = c_1 \cdot \exp\left(\frac{t - T}{T}\right) \tag{10}$$

where c_1 is a constant, typically set to 0.5.

E. Phases of SO

The Snake Optimizer alternates between three main phases based on environmental conditions defined by food quantity (Q) and temperature ($Temp$). Both **male** and **female** groups are involved in the search process, with their unique behaviors incorporated into each phase:

1) Exploration Phase ($Q < 0.25$): When food availability is low, the snakes (both males and females) explore the search space by moving randomly to locate potential solutions.

Male Group:

$$X_{i,m}(t+1) = X_{\text{rand},m} \pm c_2 \cdot A_m \cdot ((X_{\text{max}} - X_{\text{min}}) \cdot r + X_{\text{min}}) \tag{11}$$

where:

$$A_m = \exp\left(-\frac{f_{\text{rand},m}}{f_{i,m}}\right) \tag{12}$$

Female Group:

$$X_{i,f}(t+1) = X_{\text{rand},f} \pm c_2 \cdot A_f \cdot ((X_{\text{max}} - X_{\text{min}}) \cdot r + X_{\text{min}}) \tag{13}$$

where:

$$A_f = \exp\left(-\frac{f_{\text{rand},f}}{f_{i,f}}\right) \tag{14}$$

Both groups independently search for food by leveraging their abilities to explore the space broadly.

2) Exploitation Phase ($Q > 0.6$): In this phase, snakes refine their search around promising solutions (exploitation), with their behavior depending on the temperature.

Hot Environment ($Temp > 0.6$):

$$X_{i,j}(t+1) = X_{\text{food}} \pm c_3 \cdot Temp \cdot r \cdot (X_{\text{food}} - X_{i,j}(t)) \tag{15}$$

Cold Environment ($Temp < 0.6$): Snakes engage in either **fight mode** or **mating mode**.

Fight Mode:

$$X_{i,m}(t+1) = X_{i,m}(t) + c_3 \cdot F_m \cdot r \cdot (Q \cdot X_{\text{best},f} - X_{i,m}(t)) \tag{16}$$

$$X_{i,f}(t+1) = X_{i,f}(t) + c_3 \cdot F_f \cdot r \cdot (Q \cdot X_{\text{best},m} - X_{i,f}(t)) \tag{17}$$

Mating Mode:

$$X_{i,m}(t+1) = X_{i,m}(t) + c_3 \cdot M_m \cdot r \cdot (Q \cdot X_{i,f}(t) - X_{i,m}(t)) \quad (18)$$

$$X_{i,f}(t+1) = X_{i,f}(t) + c_3 \cdot M_f \cdot r \cdot (Q \cdot X_{i,m}(t) - X_{i,f}(t)) \quad (19)$$

3) Egg Hatching and Diversity Preservation: After mating, new individuals are introduced by replacing the worst-performing snakes:

$$X_{\text{worst},m} = X_{\min} + r \cdot (X_{\max} - X_{\min}) \quad (20)$$

$$X_{\text{worst},f} = X_{\min} + r \cdot (X_{\max} - X_{\min}) \quad (21)$$

This ensures diversity and prevents premature convergence.

F. Termination

The algorithm terminates when a predefined stopping criterion is met, such as reaching the maximum number of iterations or achieving a satisfactory fitness value.

IV. Snake Optimizer for Solving UAV-BSs Placement

The Snake Optimizer (SO) is applied to optimize the placement of UAV-based base stations (UAV-BSs) in wireless communication networks. This algorithm is particularly suited for addressing such problems due to its effective balance between exploration and exploitation. The SO emulates natural snake behaviors, including adaptive movement and dynamic interactions, to navigate the solution space efficiently.

The algorithm initializes with a population of candidate UAV positions, divided into male and female groups, which then undergo exploration and exploitation phases. During exploration, snakes (UAV positions) search the solution space broadly, ensuring diverse coverage of potential configurations. During exploitation, the algorithm refines promising solutions, guided by the best-performing individuals within the population. This dual-phase mechanism enables the SO to identify optimal placements that maximize user coverage while minimizing energy consumption.

The flow chart of the Snake Optimizer is provided in figure 2.

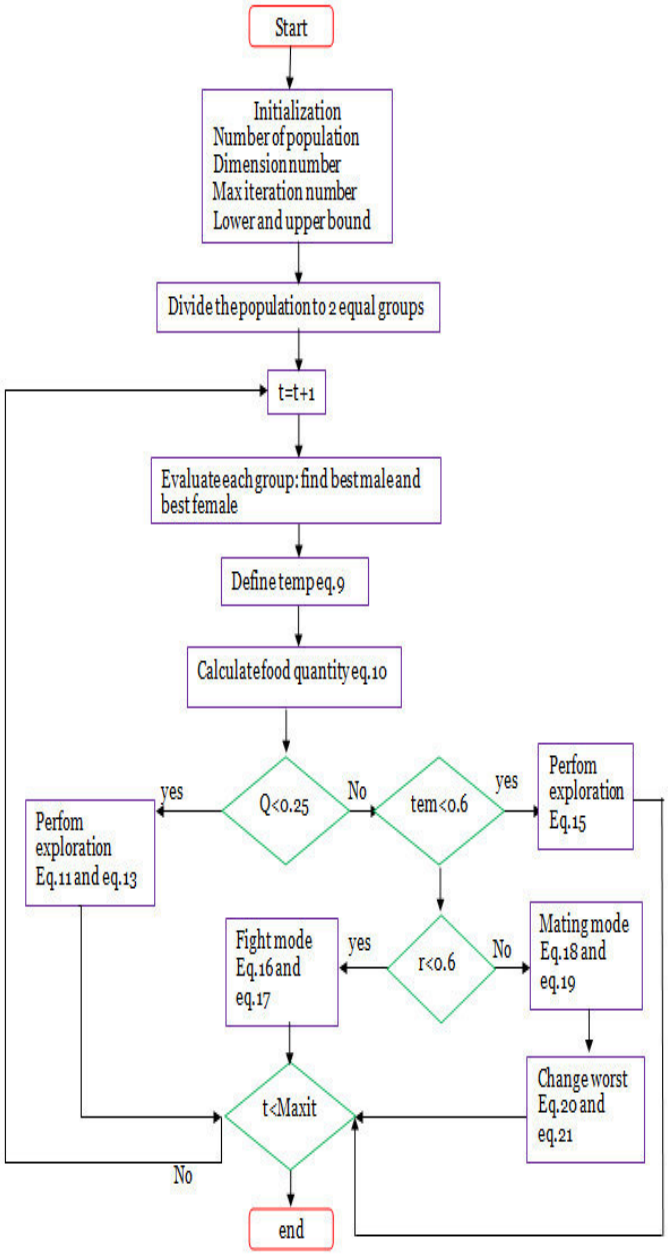


Fig. 2. Flow chart of the Snake Optimizer

V. Simulation Results

The following section evaluates the effectiveness of the Snake Optimizer (SO) in solving the optimal UAV placement problem. A comparative analysis is performed between SO and several well-known optimization algorithms, including Marine Preadtor Algorithm (MPA), Whale Optimization Algorithm (WOA), and Grey Wolf Optimizer (GWO). All algorithms are implemented using Matlab software. The performance of the proposed method is assessed in terms of the mean coverage radius and overall coverage quality. The simulations are conducted in a static environment across 16 different scenarios. The deployment area is set to a $100m \times 100m$ grid, with 5 to 40 users randomly distributed in each instance. The number of UAVs varies between 1 and 8, each having a maximum coverage radius of 25 meters and a 45-degree field of view. The coverage radius is assumed to be equal to the UAV altitude. A total of 1000 iterations is performed for each configuration. The remaining simulation parameters are summarized in Table I.

A. Impact of changing the number of UAVs

This section explores how the coverage performance is affected by varying the number of UAVs, ranging from 1 to 8, while keeping the number of users fixed at 30. The results are presented in Table II. Figure 3 illustrates the coverage outcomes, while Figure 4 shows the average coverage radius for different UAV quantities.

TABLE I
PARAMETERS VALUES CONSIDERED IN OUR SIMULATIONS

Parameter	Value
N° of UAVs m	[1 8]
N° of users n	[5 40]
Width W	100 m
Length L	100 m
Max coverage radius R	25 m
Population size N	30
N° of iteration t_{max}	1000
SO	
Threshold 1 T_{hr1}	0.25
Thershold 2 T_{hr2}	0.6
Constant C_1	0.5
Constant C_2	0.05
Constant C_3	2
MPA	
Fish Aggregating Devices $FADs$	0.2
Constant P	0.5
GWO	
Control parameter a_{min}	0
Control parameter a_{max}	2
WOA	
Control parameter a_{min}	0
Control parameter a_{max}	2

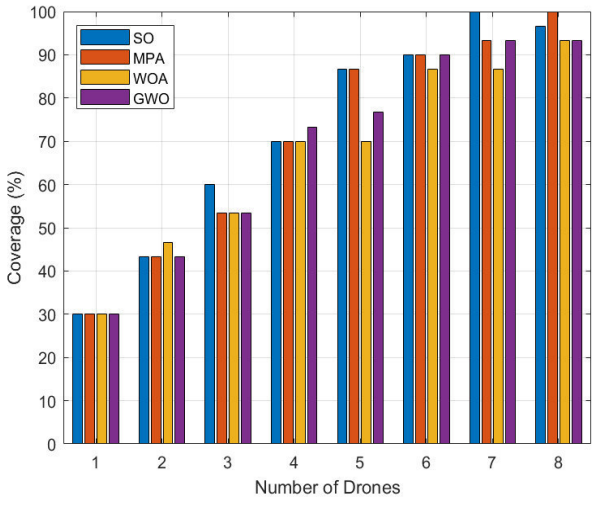


Fig. 3. Coverage under various numbers of UAVs

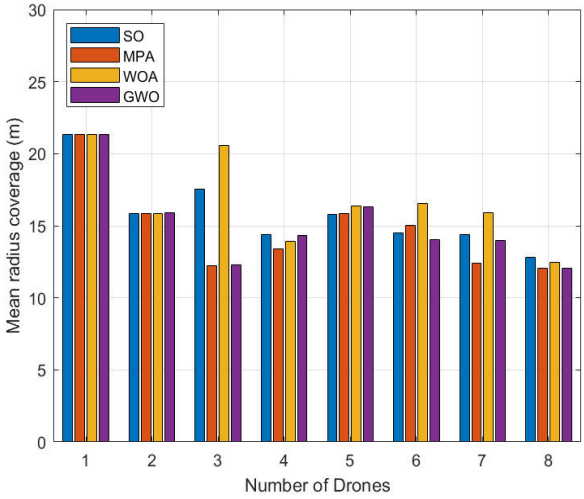


Fig. 4. Mean radius coverage under various numbers of UAVs

TABLE II
COVERAGE AND MEAN RADIUS COVERAGE UNDER VARIOUS NUMBERS OF UAVS

n	1	2	3	4	5	6	7	8
Coverage (%)								
SO	30%	43.3%	60%	70%	86.66%	90%	100%	96.66%
MPA	30%	43.3%	53.33%	70 %	86.66%	90%	93.33%	100%
WOA	30%	46.66%	53.33%	70%	70%	86.66%	86.66%	93.3%
GWO	30%	43.33%	53.33%	73.33%	76.66%	90%	93.33%	93.33%
Mean coverage radius(m)								
SO	21.32	15.87	17.53	14.38	15.80	14.53	14.41	12.81
MPA	21.32	15.87	12.23	13.38	15.82	15.05	12.42	12.04
WOA	21.32	15.85	20.59	13.94	16.38	16.57	15.89	12.44
GWO	21.33	15.88	12.28	14.33	16.34	14.01	13.98	12.06

The findings presented in Table II and Figure 3 reveal that increasing the number of UAVs enhances coverage, eventually reaching near-total user coverage once the UAV count exceeds 8. However, this increase in UAVs leads to a decrease in the average coverage radius, which corresponds to lower energy consumption per UAV. The deployment of additional UAVs helps to cover a larger area and balance the workload, resulting in reduced energy usage for each individual drone. In terms of both coverage and energy efficiency, the Snake Optimizer (SO) outperforms the MPA, WOA, and GWO algorithms in most scenarios.

B. Effect of varying the number of Users

This section explores the effect of increasing the number of users, ranging from 5 to 40, while keeping the number of UAVs fixed at 5. The analysis focuses on coverage and average coverage radius, with results summarized in Table III, Figure 5, and Figure 6.

As shown in Figure 5 and Table III, the coverage metric decreases significantly as the number of users increases. This decline occurs because the available UAVs are insufficient to cover the additional users. To accommodate the increased demand, the coverage radius of each UAV must be expanded,

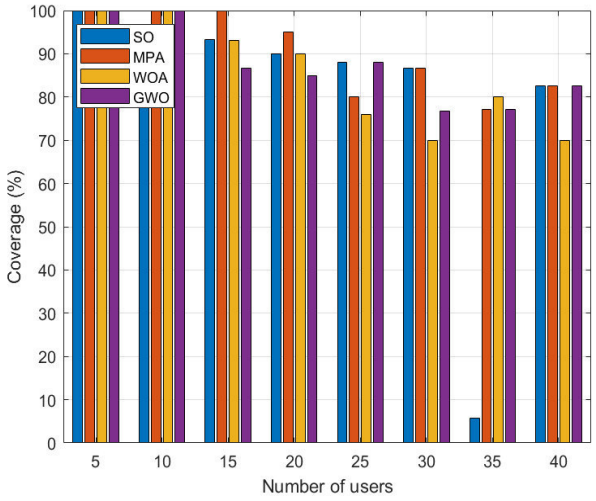


Fig. 5. Coverage under various numbers of users

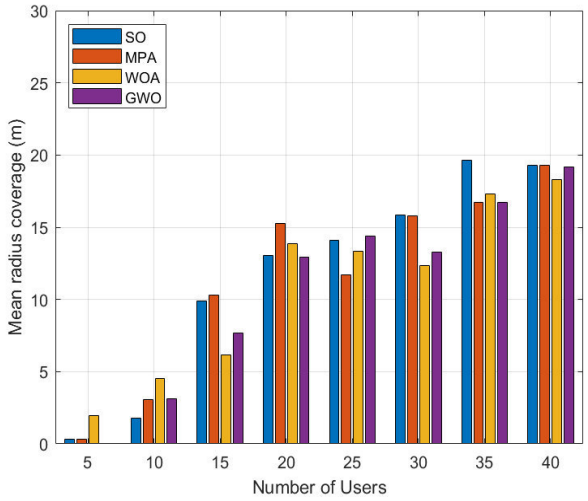


Fig. 6. Mean radius coverage under various numbers of users

TABLE III
COVERAGE AND MEAN RADIUS COVERAGE UNDER VARIOUS NUMBERS OF USERS

m	5	10	15	20	25	30	35	40
Coverage (%)								
SO	100%	90%	93.33%	90 %	88%	86.66%	85.71%	82.5%
MPA	100%	100 %	100%	95 %	80 %	86.66 %	77.14 %	82.5%
WOA	100%	100%	93%	90 %	76 %	70 %	80%	70%
GWO	100 %	100 %	86.66 %	85 %	88 %	76.66 %	77.14 %	82.5%
Mean coverage radius(m)								
SO	0.32	1.80	9.88	13.03	14.10	15.82	19.64	19.28
MPA	0.34	3.05	10.28	15.29	11.73	15.81	16.72	19.28
WOA	1.98	4.55	6.14	13.87	13.32	12.33	17.31	18.29
GWO	0.03	3.10	7.66	12.94	14.37	13.26	16.73	19.20

as depicted in Figure 6. The results also indicate that the Snake Optimizer (SO) consistently outperforms the MPA, WOA, and GWO algorithms in terms of coverage.

VI. CONCLUSION

This study introduced the Snake Optimizer (SO), a recent algorithm inspired by the behavior of snakes, to optimize the placement of UAV-based base stations (UAV-BSs) in wireless communication networks. The SO effectively balances exploration and exploitation, enabling efficient convergence towards optimal solutions. By leveraging the unique search mechanism of SO, which mimics the movement of snakes in the search space, this algorithm enhances diversity and avoids premature convergence, making it particularly suited for complex optimization tasks like UAV-BS placement. A comparative analysis was conducted, evaluating the performance of SO against other well-known algorithms, including MPA, WOA, and GWO. The simulation results demonstrated that SO outperforms these algorithms in terms of both coverage and energy efficiency. Its ability to balance coverage maximization with energy consumption minimization, combined with its robustness in diverse scenarios, makes SO a powerful tool for UAV placement problems in wireless communication networks.

REFERENCES

[1] Achilles D Boursianis, Maria S Papadopoulou, Panagiotis Diamantoulakis, Aglaia Liopa-Tsakalidi, Pantelis Barouchas, George Salahas, George Karagiannidis, Shaohua Wan, and Sotirios K Goudos. Internet of things (iot) and agricultural unmanned aerial vehicles (uavs) in smart farming: A comprehensive review. *Internet of Things*, 18:100187, 2022.

[2] Saptarshi Ghosh, Krishnachur Ghosh, Sayan Karamakar, Shubham Prasad, Nilava Debabhuti, Prolay Sharma, Bipan Tudu, Nabarun Bhattacharyya, and Rajib Bandyopadhyay. Development of an iot based robust architecture for environmental monitoring using uav. In *2019 IEEE 16th India Council International Conference (INDICON)*, pages 1–4. IEEE, 2019.

[3] Waleed Ejaz, Muhammad Awais Azam, Salman Saadat, Farkhund Iqbal, and Abdul Hanan. Unmanned aerial vehicles enabled iot platform for disaster management. *Energies*, 12(14):2706, 2019.

[4] Mohamed Amine Ouamri, Marius-Emil Oteşteanu, Gordana Barb, and Cedric Gueguen. Coverage analysis and efficient placement of drone-bss in 5g networks. *Engineering Proceedings*, 14(1):18, 2022.

[5] Khaled F Hayajneh, Khaled Bani-Hani, Hazim Shakhatreh, Muhammad Anan, and Ahmad Sawalmeh. 3d deployment of unmanned aerial vehicle-base station assisting ground-base station. *Wireless Communications and Mobile Computing*, 2021, 2021.

[6] Xukai Zhong, Yiming Huo, Xiaodai Dong, and Zhonghua Liang. Qos-compliant 3-d deployment optimization strategy for uav base stations. *IEEE Systems Journal*, 15(2):1795–1803, 2020.

[7] Sylia Mekhmoukh Taleb, Yassine Meraihi, Selma Yahia, Amar Ramdane-Cherif, Asma Benmessaoud Gabis, and Dalila Acheli. Hybrid whale optimization algorithm with simulated annealing for the uav placement problem. In *EAI International Conference on Computational Intelligence and Communications*, pages 77–88. Springer, 2022.

[8] Fatma A Hashim and Abdelazim G Hussien. Snake optimizer: A novel meta-heuristic optimization algorithm. *Knowledge-Based Systems*, 242:108320, 2022.

Energy Harvesting Study in V-VLC Systems Using Multi-hop and Commercial Headlights

Souad Refas
LIST Laboratory
University of M'Hamed Bougara
Boumerdes, Algeria
s.refas@univ-boumerdes.dz

Sidali Faical Rahmoune
LIST Laboratory
University of M'Hamed Bougara
Boumerdes, Algeria

Ihab Megdoud
LIST Laboratory
University of M'Hamed Bougara
Boumerdes, Algeria

Yassine Meraihi
LIST Laboratory
University of M'Hamed Bougara
Boumerdes, Algeria
y.meraihi@univ-boumerdes.dz

***Abstract*—This paper investigates the multi-hop relay and energy harvesting capabilities of a vehicle-to-vehicle (V2V) visible light communication (VLC) system, where commercial headlights (HLs) are used as wireless transmitters and a single photodetector (PD) serves as the receiver. We first derive a closed-form expression for the harvested energy, utilizing multi-hop relays and a realistic ray-tracing channel model that accounts for the asymmetrical beam pattern of commercial car headlights. We then analyze how various transceivers and system parameters impact the overall performance of the V2V VLC system. Our results show that the proposed V2V VLC system can harvest 2.9 J of energy while maintaining a bit error rate (BER) of 10^{-6}„ making it a promising solution for low-power wireless communication in V2V VLC scenarios.**

Furthermore, our findings highlight that employing intermediate vehicles as relay nodes can boost the harvested energy to 6 J, facilitating high data rate vehicular applications. Finely, this paper offers valuable insights into the potential of V2V VLC systems for energy-efficient communication, leveraging multi-hop relays in vehicular networks.

***Index Terms*—Energy harvesting, vehicular communications, visible light communication, multi-hop, V2V, headlights.**

I. Introduction

Intelligent Transportation Systems (ITS) play a crucial role in smart cities, enabling efficient communication between vehicles and traffic management systems through a range of connected services and applications [1], [2]. These services and applications facilitate the seamless integration of various modes of transport, provide real-time traffic updates, and enhance safety mechanisms [3]. In ITS, vehicular networks, which include vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and infrastructure-to-vehicle (I2V) communications, collectively referred to as vehicle-to-everything (V2X), rely on a variety of wireless technologies [4].

In recent years, radio frequency (RF) technologies, such as Dedicated Short-Range Communication (DSRC) and Cellular Vehicle-to-Everything (C-V2X), have garnered significant interest in V2X research and standardization efforts [4]. However, with the anticipated growth of V2X systems, the scalability of RF-based systems is limited due to restricted radio spectrum and high congestion levels in areas with medium to high node density [5]. Consequently, an alternative or complementary technology to RF is needed. Visible Light Communication (VLC) has been proposed as a promising solution, having already seen increasing adoption in indoor applications [6].

Vehicular VLC (VVLC) has also attracted significant attention from researchers in recent years [7]. This technology enables wireless communication and connectivity between moving vehicles, utilizing headlights (HL) and taillights (TL) as wireless transmitters and photodetectors (PD) as wireless receivers [8]. While wireless access is a critical component of VVLC systems, the battery capacity of associated devices can limit their usage. This limitation has led to the development of a new concept: energy harvesting (EH) systems. EH systems capture and store energy from ambient environmental sources, such as solar energy and light emitted by VVLC transmitters, and convert it into electrical energy [9].

Integrating EH systems is an important step toward creating sustainable VVLC networks. These systems can provide a reliable power source for VVLC receivers and other electronic devices, reducing the need for frequent battery replacements and lowering the environmental impact of VVLC technology.

Energy harvesting techniques have gained significant research interest in the context of indoor VLC applications [10]–[13]. These techniques offer a promising solution to overcome the limitations of traditional power sources and enable the creation of self-sustaining VLC devices. For instance, in [10], the authors proposed an energy-efficient and secure hybrid VLC-RF system with integrated energy harvesting capabilities. Meanwhile, Ramos et al. [11] explored the use of visible light for wake-up communication and energy harvesting using a different type of solar panel. Additionally, Wang et al. [12] investigated medium access control (MAC) protocols

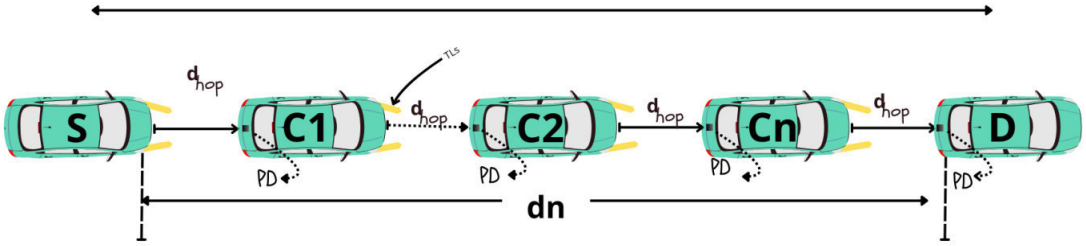


Fig. 1. V2V-VLC system with Headllights.

for indoor VLC systems with energy harvesting capabilities. They proposed an adaptive contention algorithm to address the fairness issue resulting from significant differences in energy harvesting rates among devices. The authors of [13] proposed an optimal design for a dual-hop VLC/RF communication system with energy harvesting. In their proposed system, a relay could harvest energy from the VLC link in the first hop and use the harvested energy to retransmit data to a mobile terminal via the RF link in the second hop. However, applications of energy harvesting in the context of vehicular communication have received less attention [14], [15]. Previous efforts have not considered the differences between indoor and outdoor light sources, such as taillights [16] and car headlights [17], which significantly impact channel modeling and can substantially affect energy harvesting analysis.

In this paper, we aim to bridge this gap by analyzing the communication range and energy harvesting potential of a V2V VLC system using commercial headlights as wireless transmitters and a single photodetector as a wireless receiver. Specifically, we derive an analytical expression for the harvested energy based on realistic channel modeling using ray tracing, accounting for the asymmetric beam pattern of commercial car headlights. Additionally, we examine the impact of system and transceiver parameters on the overall performance of the V2V VLC system.

The rest of this paper is organized as follows. In section II, we describe the V2V-VLC system and channel model. Section III illustrates the multi-hop and EH analysis. In section IV, we present and discuss the simulation results. Finally, section V concludes the paper.

II. V2V-VLC System and Channel Model with Headlights

A. System Model

As illustrated in Fig. 1, we consider a V2V-VLC system on a road where two cars, the source car (S) and the destination car (D), are separated by a distance of d_n, using intermediate vehicles acting as relay nodes (C1, ..., Cn). The system utilizes the headlights (HLs) of the source car as wireless VLC transmitters (TXs). Each HL has an electrical-to-optical conversion ratio of η and an electrical transmission power of P_t. The destination car is equipped with a single photodetector (PD) mounted at the front, serving as the wireless VLC receiver (RX), with a responsivity of R and an active area of A.

B. Channel Modeling Methodology

The channel modeling approach employed in this study is based on the non-sequential ray-tracing method, which was originally developed for indoor VLC channels [18]–[20] and later adapted for outdoor scenarios [21], [22]. The method involves constructing a 3D simulation platform with CAD models of the cars and roads in OpticStudio software. The surface coating material of the CAD object and the transceiver features, such as orientation, radiation pattern, aperture diameter, and field of view angle, are specified. Next, the non-sequential ray tracing model is executed, generating an output file that contains the power and path length of each ray that reaches the receiver. These statistics are then utilized in Matlab tools to obtain the channel impulse responses (CIRs). Suppose k is the total number of rays reaching the PD, P_i is the optical power, and τ_i is the propagation time of the i^{th} ray. Let δ denote the Dirac delta function. The CIR can be calculated as follows [18]

$$H(t) = \sum_{i=1}^{k} P_i \delta\left(t - \tau_i\right). \tag{1}$$

Based on this approach, the DC channel gains (in linear scale) of the forward V2V-VLC system (where the transmitter consists of the vehicle's headlights) are given respectively as

$$h_F = \left(\frac{D_r}{\alpha_f d}\right)^2 \exp\left(-cd\right), \tag{2}$$

Pl_0 represents the reference path loss at d_0 of 1 meter, and d is the separation distance between two vehicles. The values of α_b and β_b are 0.801 and 0.072, respectively, based on the Audi car model [23]. The correction coefficient α_f is 0.1585 for clear weather conditions [24]. Thus, c represents the extinction factor for a specific weather condition, and D_r is the size of the receiver's aperture.

III. Performance Analysis

Given that the communication distance at an acceptable bit error rate (BER) is critical for designing a multi-hop relay vehicular network scenario, this section will focus on analyzing the performance of the V2V-VLC system and the

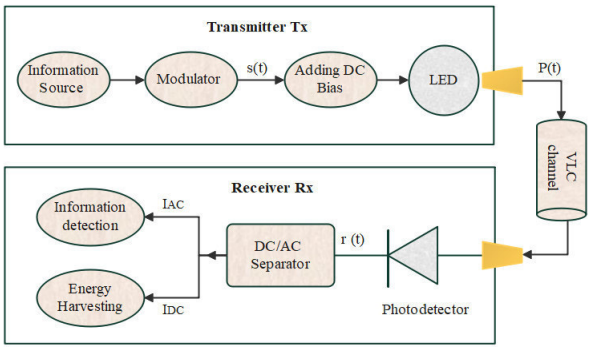


Fig. 2. Energy Harvesting strategy for V2V-VLC system.

relationship between the harvested energy in the forward direction that satisfies a BER threshold value.

A. Direct Link

In this case, there is a direct link between the source (S) and the destination, and the bit error rate (BER) (P_e) for the non-return-to-zero scheme with On-Off Keying (NRZ-OOK) modulation is given by [25]

$$P_e = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\sqrt{\gamma}}{2\sqrt{2}}\right), \tag{3}$$

where $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-t^2}dt$ is the complementary error function and γ is the signal-to-noise ratio (SNR), which is given by

$$\gamma = \frac{(\eta R h)^2 P_t}{N_0 B}, \tag{4}$$

where B is the bandwidth and N_0 is the noise power spectral density. By substituting from (4) in (3), we obtain the following equation:

$$P_e = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\sqrt{\frac{(\eta R h)^2 P_t}{N_0 B}}}{2\sqrt{2}}\right). \tag{5}$$

Then, from (5), the channel gain h can be expressed as

$$h = 2\sqrt{2}\left(\sqrt{\frac{N_0 B}{P_t \eta^2 R^2}}\right)\mathrm{erfc}^{-1}(2P_e). \tag{6}$$

B. Multi-hop Link

In this multi-hop vehicular network, in a platooning scenario where the cars are aligned with an identical distance between each consecutive pair of vehicles, and assuming that the transmitter and receiver parameters of each car are identical, the application of a decode-and-forward relay mechanism for the hops provides an upper bound on the end-to-end bit error rate (BER) after n hops, expressed by [8], [26].

$$P_e \leq \left[1 - (1 - P_{eH})^{n+1}\right]. \tag{7}$$

where $n \in \{M\}$ is the number of hops. P_{eH} is the BER occurring at each hop in this relay mechanism, which can be obtained for the considered platooning scenario by

$$P_{eH} \approx \frac{P_e}{n+1}. \tag{8}$$

C. Energy Harvesting Analysis

In this section, we analyze the energy harvesting capabilities of the HLs in the proposed V2V VLC system.

As shown in Fig. 2, the PD consists of both the Direct DC and Alternative AC signals. The DC component is blocked by a capacitor and passed through the energy harvesting branch to achieve energy harvesting. Therefore, the harvested energy is given by [27], [28]

$$EH = f_{EH} I_{DC} V_{oL}, \tag{9}$$

where f_{EH} is the fill factor, V_{oL} is the voltage across the light pole, and I_{DC} is the DC component of the output current. It can be given as follows [29], [30]

$$I_{DC} = I_{DC1} + I_{DC2}, \tag{10}$$

$$I_{DC1} = RhP_tM, \tag{11}$$

where I_{DC1} is the input bias current caused by the dedicated LED and I_{DC2} is input bias current caused by other light sources [29]. $M \in [I_{max}, I_{min}]$, with I_{min} and I_{max} respectively, are the min and the max input bias currents. Furthermore, In (8), V_{oL} can be written as

$$V_{oL} = V \ln\left(1 + \frac{I_{DC}}{I_d}\right), \tag{12}$$

where V is the thermal voltage and I_d is the dark saturation current of the PD. By replacing h in (10) using (6), we can write (12) as a function of threshold BER P_e^{th} as follows.

$$I_{DC1} = 2\sqrt{2}RP_tM\left(\sqrt{\frac{N_0 B}{P_t \eta^2 R^2}}\right)\mathrm{erfc}^{-1}(2P_e). \tag{13}$$

From (6-13), we obtain the total EH to achieve a threshold BER of P_e^{th} which is given by (14).

IV. Numerical Results and Discussions

In this section, we present numerical results of the achievable distance and harvested energy required to reach a specified BER threshold for the V2V-VLC system.

To conduct our simulation analysis, we assume $R = 0.54$ A/W, $\eta = 0.5$ W/A, A = 1 cm^2 N_0 = 10^{-22} W/Hz. We consider various electrical transmit power values, P_t = 20 dBm, 30 dBm, and 40 dBm, as well as different bandwidth values, B = 100 kHz, 1 MHz, and 5 MHz. Unless otherwise mentioned, we assume a fill factor f_{EH} = 0.75, I_{max} = 0.2 A, I_{min} = 0.15 A, I_d = 10^{-9} A . All simulation parameters are given in Table 1.

Figure 3 illustrates the effect of bandwidth B on the amount of harvested energy while maintaining the target bit error rate (BER) level. We assume a transmission power of P_t = 40 dBm. The observation shows that as the BER increases, the harvested energy decreases. This trend can be

$$EH \tilde{<} (M+1) f_{EH} V \left(2\sqrt{2} R P_t M_0 \sqrt{\frac{N_0 B (M+1)}{P_t \eta^2 R^2}} \operatorname{erfc}^{-1}\left(\frac{2P_e^{th}}{M+1}\right) + I_{DC2} \right) \ln \left(1 + \frac{2\sqrt{2} R P_t M_0 \sqrt{\frac{N_0 B (M+1)}{P_t \eta^2 R^2}} \operatorname{erfc}^{-1}\left(\frac{2P_e^{th}}{M+1}\right) + I_{DC2}}{I_d} \right). \tag{14}$$

TABLE I
SIMULATION PARAMETERS [23].

Parameter	Value
Electrical-to-Optical conversion factor (η)	0.5 (W/A)
Electrical transmission power (P_t)	20 dBm, 30 dBm, and 40 dBm
Responsivity (R)	0.54 (A/W)
Noise power density (N_0)	1×10^{-22} (A^2/Hz)
Bandwidth (B)	100 kHz, 1 MHz, and 5 MHz
Fill factor (f_{EH})	0.75 [29]
Min input current (I_{min})	0.15 A [31]
Max input current (I_{max})	0.2 A [31]
Thermal voltage (V)	12 V [31]
Dark saturation current (I_d)	10^{-9} A [29]

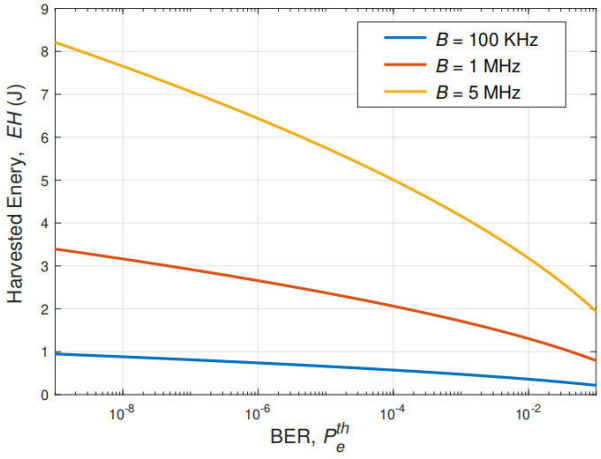


Fig. 3. Energy Harvesting Analysis as a Function of Bit Error Rate (BER) for Different Bandwidth Values, Assuming P_t = 40 dBm and M = 1.

explained by the fact that an increase in BER leads to a higher probability of errors in the transmitted signal, resulting in a greater percentage of wasted transmission power. For example, for a threshold BER of $P_e^{th} = 10^{-6}$ and $B = 1\,\text{MHz}$, the harvested energy is $EH = 2.8\,\text{J}$. This value decreases to $EH = 2\,\text{J}$ for $P_e^{th} = 10^{-4}$. Additionally, a significant impact of the system's bandwidth on the harvested energy is observed. By increasing the system's bandwidth, more ambient noise can

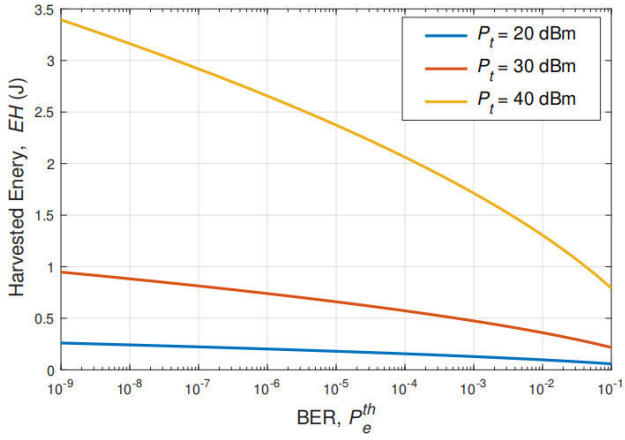


Fig. 4. Energy harvesting as a function of the bit error rate (BER) for different transmitted power values, with $B = 1\,\text{MHz}$ and $M = 1$.

be captured, which can then be converted into usable energy through harvesting techniques. For example, for a threshold BER of $P_e^{th} = 10^{-6}$, the harvested energy is about 1.9 J with $B = 5\,\text{MHz}$. This value increases to 2.8 J and 6.5 J for $B = 1\,\text{MHz}$ and $B = 100\,\text{kHz}$, respectively.

In Figure 4, we study the effect of the transmission electrical power on the harvested energy for a given bit error rate (BER) target, assuming $B = 1\,\text{MHz}$ and $M = 1$. It is observed that the harvested energy is strongly affected by the transmission power budgets. For example, considering $P_e^{th} = 10^{-6}$ and $P_t = 40\,\text{dBm}$, the harvested energy is given by $EH = 2.8\,\text{J}$. However, when the power budget decreases (i.e., $P_t = 30\,\text{dBm}$), the harvested energy decreases to $EH \approx 0.75\,\text{J}$, and for $P_t = 20\,\text{dBm}$, the harvested energy further decreases to $EH = 0.3\,\text{J}$.

To demonstrate the advantages of this V2V-VLC communication system with multi-hop relays, in Figure 5, we compared the harvested energy obtained in two scenarios: direct communication only and using a single relay. In this analysis, we used $B = 1\,\text{MHz}$, $P_t = 30\,\text{dBm}$, and a single relay $M = 1$. Figure 4 shows a significant improvement in the harvested energy using a multi-hop link compared to the direct link. To visualize the range improvement, consider a target bit error rate of $P_e^{th} = 10^{-6}$. The harvested energy with a single-hop relay is $EH = 0.75\,\text{J}$, which is 0.5 J greater than that of the direct link. Thus, it is clear that a multi-hop relay system is an alternative solution capable of compensating for the harvested energy. This improvement comes from the fact that the energy is relayed through other intermediate vehicles, and at each vehicle, the energy is harvested.

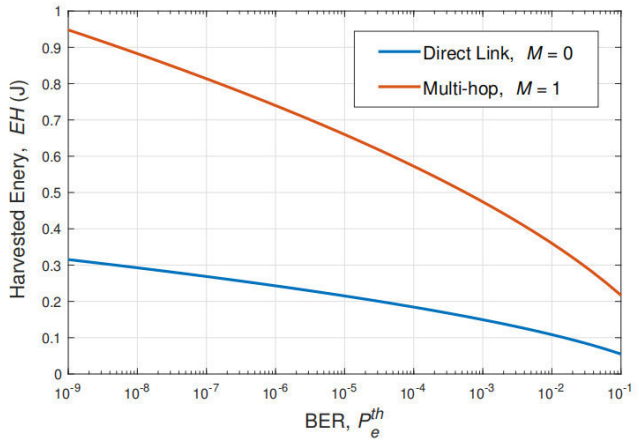


Fig. 5. Energy harvesting versus BER rate, with $B = 1\,\text{MHz}$ and $P_t = 30\,\text{dBm}$.

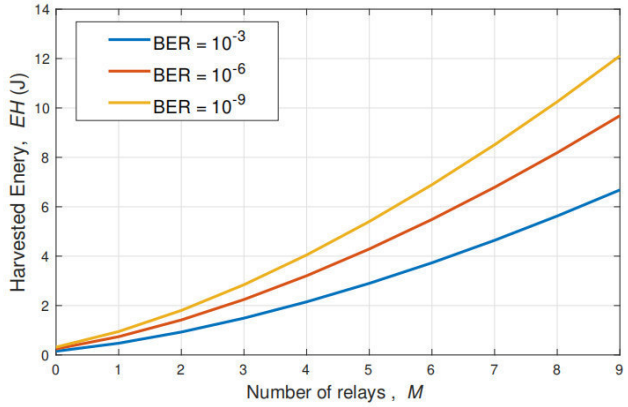


Fig. 6. Energy harvesting versus the number of relays, with $P_t = 30\,\text{dBm}$ and $B = 1\,\text{MHz}$.

In Figure 6, we study the effect of increasing the number of relays on the harvested energy. We consider $B = 1\,\text{MHz}$, $P_t = 30\,\text{dBm}$, and different BER targets with $P_e^{th} = 10^{-9}$, $P_e^{th} = 10^{-6}$, and $P_e^{th} = 10^{-3}$. It is observed that the harvested energy increases with the number of intermediate relays. For example, considering $P_e^{th} = 10^{-6}$, the harvested energy is given for $M = 1$ as $EH = 0.75\,\text{J}$. This increases to $EH = 4.2\,\text{J}$ and $EH = 9.9\,\text{J}$ for $M = 5$ and $M = 9$, respectively. It can also be observed that the target error rate influences the harvested energy. For example, considering $M = 5$, the ranges for $P_e^{th} = 10^{-9}$, $P_e^{th} = 10^{-6}$, and $P_e^{th} = 10^{-3}$ are given by $EH = 5.7\,\text{J}$, $EH = 4.2\,\text{J}$, and $EH = 3\,\text{J}$, respectively.

V. Conclusion

This paper has provided insights into the achievable energy harvesting capabilities of the V2V VLC system. Considering a realistic channel model that captures the asymmetric intensity profiles of the headlights, analytical expressions for the harvested energy under a target BER value have been derived. Additionally, the impact of the system bandwidth and the transmitted power budget on the system's performance has also been analyzed. The results showed that the proposed V2V VLC system can achieve an energy harvesting of 2.9 J when the BER performance is maintained at 10^{-6}, making it a promising solution for low-power wireless communication in V2V VLC scenarios.

Furthermore, the results revealed that using intermediate cars as relay nodes has the potential to extend the harvested energy to 6 J and enable vehicular applications with high data rates.

References

[1] R. I. Meneguette, R. De Grande, and A. Loureiro, "Intelligent transport system in smart cities," *Cham: Springer International Publishing*, 2018.

[2] O. Pribyl, P. Pribyl, M. Lom, and M. Svitek, "Modeling of smart cities based on ITS architecture," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 4, pp. 28–36, 2018.

[3] S. H. Bouk, S. H. Ahmed, D. Kim, and H. Song, "Named-data-networking-based ITS for smart cities," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 105–111, 2017.

[4] Z. MacHardy, A. Khan, K. Obana, and S. Iwashina, "V2X access technologies: Regulation, research, and remaining challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1858–1877, Feb. 2018.

[5] L. E. M. Matheus, A. B. Vieira, L. F. Vieira, M. A. Vieira, and O. Gnawali, "Visible light communication: concepts, applications and challenges," *IEEE Commun Surv Tutor*, vol. 21, no. 4, pp. 3204–3237, 2019.

[6] D. Karunatilaka, F. Zafar, V. Kalavally, and R. Parthiban, "LED based indoor visible light communications: State of the art," *IEEE Commun. Surv. Tutor.*, vol. 17, no. 3, pp. 1649–1678, 2015.

[7] A. Memedi and F. Dressler, "Vehicular visible light communications: A survey," *IEEE Commun. Surveys & Tut.*, vol. 23, no. 1, pp. 161–181, 2020.

[8] S. Refas, D. Acheli, S. Yahia, Y. Meraihi, A. Ramdane-Cherif, N. V. Van, and T. D. Ho, "Performance analysis of bidirectional multi-hop vehicle-to-vehicle visible light communication," *IEEE Access*, vol. 11, pp. 129 436–129 448, 2023.

[9] Z. Wang, D. Tsonev, S. Videv, and H. Haas, "On the design of a solar-panel receiver for optical wireless communications with simultaneous energy harvesting," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 8, pp. 1612–1623, 2015.

[10] G. Pan, J. Ye, and Z. Ding, "Secure hybrid VLC-RF systems with light energy harvesting," *IEEE Trans Commun.*, vol. 65, no. 10, pp. 4348–4359, 2017.

[11] J. S. Ramos, I. Demirkol, J. Paradells, D. Vössing, K. M. Gad, and M. Kasemann, "Towards energy-autonomous wake-up receiver using visible light communication," in *2016 13th IEEE Consum. Commun. Netw. Conf. (CCNC)*. IEEE, 2016, pp. 544–549.

[12] H.-q. Wang, X.-f. Chi, and L.-l. Zhao, "Energy adaptive MAC protocol for IEEE 802.15. 7 with energy harvesting," *Optoelectron. Lett.*, vol. 12, pp. 370–374, 2016.

[13] T. Rakia, H.-C. Yang, F. Gebali, and M.-S. Alouini, "Optimal design of dual-hop VLC/RF communication system with energy harvesting," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 1979–1982, 2016.

[14] A. M. Abdelhady, O. Amin, A. Chaaban, B. Shihada, and M.-S. Alouini, "Spectral-efficiency—illumination pareto front for energy harvesting enabled VLC systems," *IEEE Trans Commun.*, vol. 67, no. 12, pp. 8557–8572, 2019.

[15] S. Refas, D. Acheli, S. Yahia, Y. Meraihi, H. B. Eldeeb, T. D. Ho, L. Jiang, and S. Shimamoto, "Analysis of communication distance and energy harvesting for vehicular vlc using commercial taillights," in *2023 5th International Conference on Computer Communication and the Internet (ICCCI)*. IEEE, 2023, pp. 162–167.

[16] H. B. Eldeeb, E. Eso, M. Uysal, Z. Ghassemlooy, S. Zvanovec, and J. Sathian, "Vehicular visible light communications: The impact of taillight radiation pattern," in *2020 IEEE Photon. Conf. (IPC)*, 2020, pp. 1–2.

[17] A. Memedi, H.-M. Tsai, and F. Dressler, "Impact of realistic light radiation pattern on vehicular visible light communication," in *GLOBECOM 2017 - 2017 IEEE Global Commun. Conf.*, 2017, pp. 1–6.

[18] E. Sarbazi, M. Uysal, M. Abdallah, and K. Qaraqe, "Ray tracing based channel modeling for visible light communications," in *2014 22nd Signal Process. Commun. App. Conf. (SIU)*. IEEE, 2014, pp. 702–705.

[19] F. Miramirkhani and M. Uysal, "Channel modeling and characterization for visible light communications," *IEEE Photon. J.*, vol. 7, no. 6, pp. 1–16, 2015.

[20] Z. Nazari Chaleshtori *et al.*, "Utilization of an OLED-based VLC system in office, corridor, and semi-open corridor environments," *Sensors*, vol. 20, no. 23, p. 6869, 2020.

[21] H. B. Eldeeb, M. Elamassie, S. M. Sait, and M. Uysal, "Infrastructure-to-vehicle visible light communications: channel modelling and performance analysis," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2240–2250, 2022.

[22] S. Yahia, Y. Meraihi, S. Refas, A. B. Gabis, A. Ramdane-Cherif, and H. B. Eldeeb, "Performance study and analysis of MIMO visible light communication-based V2V systems," *Opt Quantum Electron.*, vol. 54, no. 9, p. 575, 2022.

[23] H. B. Eldeeb, E. Eso, E. A. Jarchlo, S. Zvanovec, M. Uysal, Z. Ghassemlooy, and J. Sathian, "Vehicular VLC: A ray tracing study based on measured radiation patterns of commercial taillights," *IEEE Photon. Technol. Lett.*, 2021.

[24] M. Karbalayghareh *et al.*, "Channel modelling and performance limits of vehicular visible light communication systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 6891–6901, 2020.

[25] S. Refas, D. Acheli, S. Yahia, Y. Meraihi, A. Ramdane-Cherif, A. B. Gabis, and H. B. Eldeeb, "Performance analysis of multi-hop v2v vlc system under atmospheric weather conditions," in *2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*. IEEE, 2022, pp. 414–419.

[26] A. Florea and H. Yanikomeroglu, "On the optimal number of hops in infrastructure-based fixed relay networks," in *IEEE Global Telecommun. Conf. (GLOBECOM)*. IEEE, 2005, pp. 3242–3247.

[27] C. Li, W. Jia, Q. Tao, and M. Sun, "Solar cell phone charger performance in indoor environment," in *2011 IEEE 37th Annu. Northeast Bioeng. Conf. (NEBEC)*. IEEE, 2011, pp. 1–2.

[28] S. Refas, D. Acheli, S. Yahia, Y. Meraihi, A. Ramdane-Cherif, and A. B. Gabis, "Energy harvesting based on slipt in v2v-vlc system under atmospheric weather conditions," in *2022 3rd International Conference on Human-Centric Smart Environments for Health and Well-being (IHSH)*. IEEE, 2022, pp. 13–17.

[29] P. D. Diamantoulakis, G. K. Karagiannidis, and Z. Ding, "Simultaneous lightwave information and power transfer (SLIPT)," *IEEE Trans. Green Commun.*, vol. 2, no. 3, pp. 764–773, 2018.

[30] S. Refas, D. Acheli, S. Yahia, and Y. Meraihi, "Energy harvesting based on slipt in i2v-vlc system," in *2023 International Conference on Advances in Electronics, Control and Communication Systems (ICAECCS)*. IEEE, 2023, pp. 1–5.

[31] Audi A5, https://www.audiusa.com/models/audi-a5-coupe, [Accessed: December 5, 2024].

Simulation and Optimization of Quality of Service in 5G Networks: A Rural Deployment Perspective

Souad Refas
LIST Laboratory
University of M'Hamed Bougara
Boumerdes, Algeria
s.refas@univ-boumerdes.dz

Anis Guettaf
LIST Laboratory
University of M'Hamed Bougara
Boumerdes, Algeria

Yassine Meraihi
LIST Laboratory
University of M'Hamed Bougara
Boumerdes, Algeria

***Abstract*—The emergence of 5G networks has brought unprecedented advancements in communication technologies, offering significantly improved performance in terms of throughput, latency, and reliability. However, achieving optimal Quality of Service (QoS) in rural environments remains a critical challenge due to low population density and unique infrastructural constraints. This paper addresses these challenges by focusing on the optimization of QoS in a rural 5G network.**

We consider a scenario where a single base station serves multiple user equipment within a wide coverage area. The study focuses on three key performance indicators: throughput, latency, and connection reliability, which are influenced by parameters such as user distance, network load, and signal-to-noise ratio (SNR). MATLAB is utilized for modeling and simulation, while optimization is performed using the Genetic Algorithm (GA), a robust method inspired by natural evolutionary processes.

Our results provide actionable insights into the trade-offs required to enhance QoS in rural 5G networks, demonstrating the effectiveness of evolutionary algorithm-based approaches for tackling complex optimization problems. These findings offer valuable contributions for designing efficient 5G systems and pave the way for future research in similar deployment contexts.

***Index Terms*—Quality of Service (QoS), 5G Networks, Genetic Algorithm (GA), Signal-to-Noise Ratio (SNR).**

I. Introduction

The advent of 5G networks marks a revolution in communication technologies, offering significantly higher performance compared to previous generations. Quality of Service (QoS) is central in this context as it ensures high throughput, low latency, and strong reliability, meeting the diverse needs of modern applications such as autonomous vehicles, telemedicine, and online gaming [1].

However, ensuring optimal QoS remains a challenge, particularly in rural areas. These environments are often characterized by low population density and infrastructure constraints, making standard deployment and optimization solutions less effective [2]. Addressing these challenges requires adopting realistic modeling approaches and efficient optimization tools.

In this paper, we focus on optimizing QoS in a rural 5G network by concentrating on three key performance indicators: throughput, latency, and reliability. This study framework is particularly relevant for rural areas, where a single base station (BS) serves multiple user equipment (UE) spread over a wide coverage area. Network performance under these conditions heavily depends on parameters such as the distance between the BS and users, network load, and the signal-to-noise ratio (SNR), which are often overlooked in traditional analyses [3], [4].

To address these issues, we use MATLAB, a well-established environment for modeling and simulation, along with the Genetic Algorithm (GA), an optimization technique inspired by the mechanisms of natural evolution. Genetic algorithms have proven particularly effective in solving complex optimization problems in communication networks [5]. For instance, GA-based optimization has been successfully employed to maximize throughput in heterogeneous networks [6] and reduce latency in multi-user systems [7].

By leveraging these methodologies, our study aims to provide concrete and actionable results to improve QoS in contexts where topographical constraints and user distribution pose unique challenges. Our results highlight the necessary trade-offs between different performance parameters, contributing to a better understanding of potential solutions for enhancing QoS in 5G.

The contributions of this paper are twofold: first, it presents a robust methodology for analyzing the performance of rural 5G networks; second, it demonstrates the effectiveness of evolutionary algorithm-based approaches in optimizing complex systems. These results may inspire future research to generalize these approaches to other deployment contexts.

The rest of this paper is organized as follows: In Section II, we describe the performance analysis. Section III illustrates the methodology. In Section IV, we present and discuss the simulation scenarios and numerical results. Finally, section V concludes the paper.

II. Performance Analysis

In this section, we analyze the relationships between key performance parameters of 5G networks: throughput, latency, and connection reliability. These metrics are evaluated as functions of the distance between user equipment (UE) and the base station (BS), the signal-to-noise ratio (SNR), and the network load.

1) **Throughput (D):** Throughput is influenced by several factors, including distance (d), SNR (S), and network

load (C). The mathematical relationship can be approximated as:

$$D = f(d, S, C) \tag{1}$$

Here, d represents the distance. As the distance between the UE and the BS increases, the throughput decreases due to signal attenuation, which is commonly modeled by an inverse power law [8]. S denotes the SNR. A higher SNR is associated with increased throughput because it indicates a stronger signal relative to background noise, enabling faster and more reliable data transmission [9]. C represents the network load. A higher load reduces the available throughput per user as resources are shared among a larger number of users and applications [10].

2) **Average Latency (L):** Latency refers to the time elapsed between sending and receiving data. Its relationship with the network parameters can be modeled as:

$$L = g(d, S, C) \tag{2}$$

A greater distance typically increases latency because the signal takes longer to travel between the UE and the BS [11]. A higher SNR reduces transmission errors, contributing to lower latency by ensuring more reliable and faster communication [12]. Additionally, a higher network load can increase latency due to congestion and longer data processing queues [13].

3) **Connection Reliability (R):** Reliability measures the likelihood of successful, error-free data transmission. It can be expressed as:

$$R = h(d, S, C) \tag{3}$$

As the distance increases, reliability may decrease due to signal attenuation and a higher likelihood of transmission losses [8]. Similarly, a higher SNR improves reliability by reducing transmission errors and enhancing the signal's robustness against noise [9]. A higher network load can reduce reliability by increasing data collisions and overloading network resources, leading to packet losses [10].

III. Methodology

In this study, we employed MATLAB to model and analyze various 5G network scenarios, incorporating different configurations of base stations (BS), users (UE), and network traffic loads. The primary Quality of Service (QoS) parameters evaluated include throughput, latency, and connection reliability. By examining these factors under diverse conditions, we aim to better understand how network design choices impact overall performance. The focus was placed on optimizing the key QoS indicators by adjusting parameters such as UE-BS distance, signal-to-noise ratio (SNR), and network load.

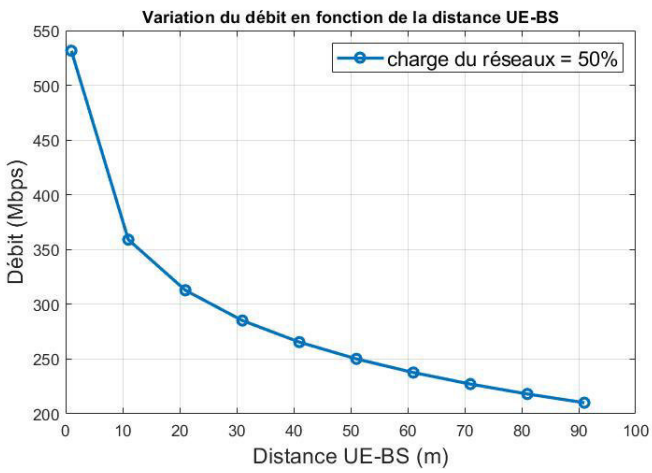


Fig. 1. Flow rate vs Distance

A. Simulation Environment

The simulation environment was implemented using MATLAB, which provides a robust platform for network modeling and performance analysis. The following key features were incorporated into the simulation:

- **Base Station (BS) Configuration:** The BS was modeled as the central node, with its coverage area defined based on the propagation characteristics of 5G networks. The BS supports multiple UE connections simultaneously, managing resources dynamically based on traffic load and signal conditions.
- **User Equipment (UE):** The UE devices were randomly distributed within the coverage area of the BS to mimic realistic deployment scenarios. Each UE generates traffic based on predefined patterns, simulating diverse user behavior and application demands.
- **Traffic Load:** The network load was varied across simulations to evaluate its impact on QoS metrics. Scenarios included low, medium, and high traffic loads to observe the system's performance under different levels of congestion.

IV. Simulation Scenarios and Numerical Results

To ensure a comprehensive evaluation of network performance, several simulation scenarios were examined. These scenarios include varying configurations of base stations (BS), user equipment (UE), and network load. The key Quality of Service (QoS) parameters analyzed in these scenarios are throughput, latency, and connection reliability.

A. Throughput Variation with UE-BS Distance

Figure 1 illustrates how throughput changes with the distance between the user equipment (UE) and the base station (BS). A noticeable decrease in throughput is observed as the distance increases, highlighting the critical role of network density and transmission power in maintaining high throughput over greater distances.

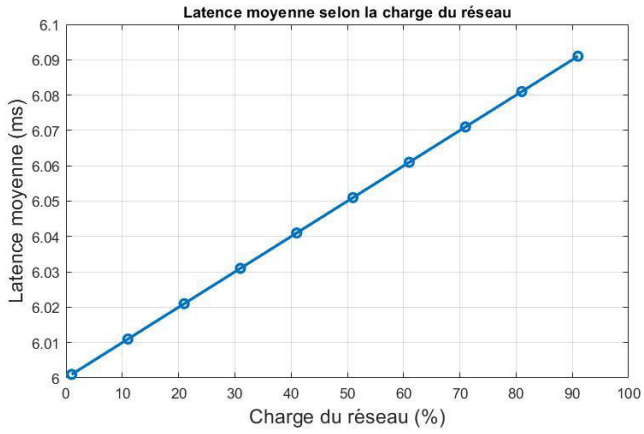


Fig. 2. Average Latency vs Network Load

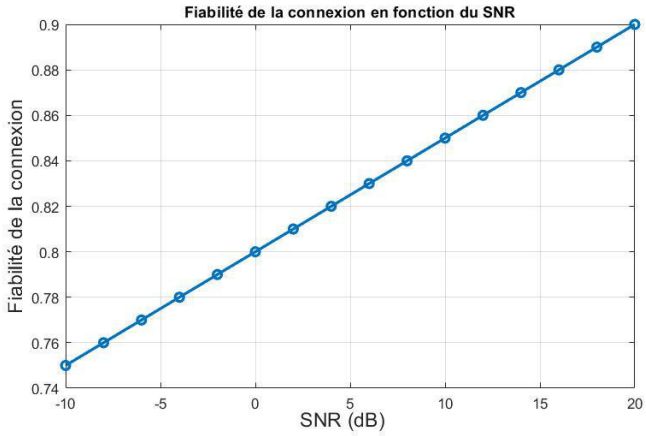


Fig. 3. Connection Reliability vs SNR

Figure 1 depicts the relationship between throughput (in Mbps) and the distance (in meters) between the UE and the BS. As the distance increases, the throughput decreases, which can be attributed to signal degradation over long distances, caused by propagation losses and interference. When users are located close to the BS, the signal strength is high, ensuring a high throughput. However, as the distance increases, the signal weakens, resulting in reduced throughput. This underscores the importance of distance in optimizing 5G networks to maximize QoS.

B. Average Latency as a Function of Network Load

Figure 2 illustrates the increase in average latency (in ms) as network load (in percentage) rises. When the network load is low, latency remains minimal, allowing for quick responses to data requests. However, as the load increases, network congestion leads to longer delays, as resources become insufficient to handle all requests immediately. This increased latency negatively affects the user experience, slowing down interactive applications, increasing buffering times for streaming, and reducing the performance of business applications. To maintain low latency, it is crucial to efficiently manage network load by enhancing network capacity and optimizing traffic management and resource allocation.

C. Connection Reliability vs. SNR

Figure 3 shows the relationship between connection reliability (in percentage) and the signal-to-noise ratio (SNR, in dB). Connection reliability measures the network's ability to maintain a stable, uninterrupted connection. A higher SNR indicates a stronger signal relative to the background noise, which improves communication quality. As the SNR increases, connection reliability improves because a clearer and stronger signal reduces transmission errors and data loss. This relationship is vital for network performance, as greater reliability leads to fewer disconnections, better call quality, and more robust data transmission. Optimizing SNR is therefore essential for network operators to deliver a more stable and satisfactory user experience.

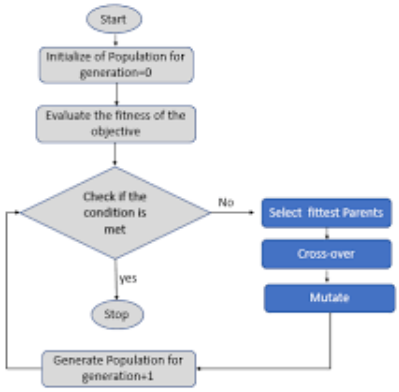


Fig. 4. Genetic Algorithm Optimization [14]

D. Optimization Framework Using Genetic Algorithm (GA)

To optimize QoS parameters, a Genetic Algorithm (GA) was integrated into the simulation framework. The GA was employed to determine the optimal configurations for BS placement, resource allocation, and traffic management strategies, with the goal of maximizing throughput, minimizing latency, and enhancing connection reliability.

The optimization process utilizes MATLAB's GA function, an optimization method inspired by the principles of natural selection. Potential solutions are treated as individuals in a population, and these individuals undergo selection, crossover, and mutation processes to create successive generations, each showing improved performance.

E. Genetic Algorithm Definition

The Genetic Algorithm (GA) is a heuristic optimization technique inspired by the natural selection process observed in biological organisms. The main steps of the algorithm include those shown in Figure 4:

- **Initialization:** Creation of an initial population of potential solutions.
- **Evaluation:** Calculation of the objective function for each individual in the population.

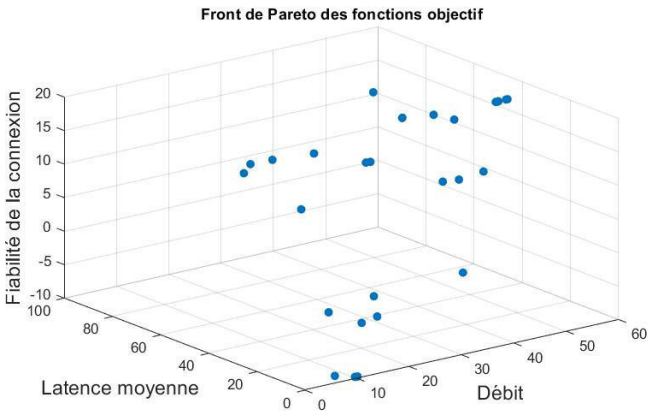


Fig. 5. parameter optimization

- **Selection:** Selection of the fittest individuals for reproduction.
- **Crossover:** Combination of selected individuals to produce a new generation.
- **Mutation:** Random alteration of some individuals to maintain genetic diversity.
- **Replacement:** Substitution of the current population with the new generation.

1) Optimization Implementation: In order to optimize the values of throughput, latency, and reliability, the three input parameters—distance, network load, and SNR—are defined. The objective function is designed to maximize throughput and reliability while minimizing latency.

Figure 5 demonstrates the three-dimensional optimization process using the genetic algorithm (GA). This representation shows how optimal values for throughput, average latency, and connection reliability are achieved based on the parameters of distance, network load, and SNR.

Following optimization, the optimal parameter values are obtained. The results show a significant improvement in throughput, a reduction in latency, and an increase in connection reliability. The optimal parameters found are as follows:

- UE-BS Distance: 5 km
- Network Load (%): 10.2841
- SNR (dB): 50.453

The objective function values at the optimal solution are:

- Throughput: 745.7981 Mbps
- Average Latency: 15.1462 ms
- Connection Reliability: 1.0523

These results indicate that the optimal distance between the UE and the BS is 5 km, with a network load of 10.2841 and an SNR of 50.453 dB. This configuration maximizes throughput, minimizes average latency, and enhances connection reliability, demonstrating the effectiveness of the genetic algorithm in optimizing 5G network performance for better Quality of Service (QoS).

V. Conclusion

This paper presented a comprehensive study on the simulation and optimization of Quality of Service (QoS) in 5G networks, with a focus on the rural deployment scenario. By analyzing the key performance indicators—throughput, latency, and connection reliability—based on critical parameters such as UE-BS distance, network load, and signal-to-noise ratio (SNR), we were able to demonstrate the impact of these factors on the network's performance.

The simulations revealed that throughput decreases with increasing distance, highlighting the importance of network density and transmission power. Additionally, the study showed that latency increases as network load rises, underscoring the need for efficient traffic management to maintain low latency. Connection reliability was found to improve with higher SNR, further emphasizing the significance of a strong signal in ensuring a stable and uninterrupted connection.

Moreover, the use of a Genetic Algorithm (GA) for network optimization proved to be effective in finding optimal configurations for maximizing throughput, minimizing latency, and enhancing connection reliability. The results indicated that the optimal configuration for the network in the given scenario was a UE-BS distance of 5 km, a network load of 10.28%, and an SNR of 50.45 dB, which resulted in improved QoS metrics.

In conclusion, this work demonstrates the potential of optimization techniques, such as the genetic algorithm, to enhance the performance of 5G networks, especially in rural environments. The findings provide valuable insights into network planning and resource allocation strategies that can be employed to optimize QoS in real-world 5G deployments.

References

[1] A. U. Rehman, I. Mahmood, M. Kamran, M. Sanaullah, A. Ijaz, J. Ali, and M. Ali, "Enhancement in quality-of-services using 5g cellular network using resource reservation protocol," *Physical Communication*, vol. 55, p. 101907, 2022.

[2] S. A. Hassan, M. S. Omar, M. A. Imran, J. Qadir, and D. Jayako, "Universal access in 5g networks, potential challenges and opportunities for urban and rural environments," *5G Networks: Fundamental Requirements, Enabling Technologies, and Operations Management*, 2017.

[3] I. Pyatin, J. Boiko, O. Eromenko, and I. Parkhomey, "Implementation and analysis of 5g network identification operations at low signal-to-noise ratio," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 21, no. 3, pp. 496–505, 2023.

[4] L. Chiaraviglio, N. Blefari-Melazzi, W. Liu, J. A. Gutiérrez, J. Van De Beek, R. Birke, L. Chen, F. Idzikowski, D. Kilper, P. Monti *et al.*, "Bringing 5g into rural and low-income areas: Is it feasible?" *IEEE Communications Standards Magazine*, vol. 1, no. 3, pp. 50–57, 2017.

[5] K. Deb, "Genetic algorithm in search and optimization: the technique and applications," in *Proceedings of International Workshop on Soft Computing and Intelligent Systems,(ISI, Calcutta, India)*. Proceedings of International Workshop on Soft Computing and Intelligent ..., 1998, pp. 58–87.

[6] R. Sachan, T. J. Choi, and C. W. Ahn, "A genetic algorithm with location intelligence method for energy optimization in 5g wireless networks," *Discrete Dynamics in Nature and Society*, vol. 2016, no. 1, p. 5348203, 2016.

[7] H. Ganame, L. Yingzhuang, A. Hamrouni, H. Ghazzai, and H. Chen, "Evolutionary algorithms for 5g multi-tier radio access network planning," *IEEE Access*, vol. 9, pp. 30 386–30 403, 2021.

[8] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[9] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.
[10] C. S. Inc., "Cisco annual internet report (2018–2023)," 2020.
[11] W. C. Lee, *Wireless and cellular communications*. McGraw-Hill Professional, 2005.
[12] J. G. Proakis and M. Salehi, *Digital communications*. McGraw-hill, 2008.
[13] T. Hoßfeld, M. Fiedler, and T. Zinner, "The qoe provisioning-delivery-hysteresis and its importance for service provisioning in the future internet," in *2011 7th EURO-NGI Conference on Next Generation Internet Networks*. IEEE, 2011, pp. 1–8.
[14] A. S. T. Hussain, M. K. Jarjes, and S. F. Ahmed, "Pid controller optimization using genetic algorithm based zeta converter application," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 4, pp. 558–562, 2020.

Particle Swarm Optimization for Solving the UAV-mounted Camera Network Placement Problem

Sylia Mekhmoukh Taleb
LIST Laboratory
University of M'Hamed Bougara
Boumerdes, Algeria
s.mekhmoukh@univ-boumerdes.dz

Karim Baiche
LIST Laboratory
University of M'Hamed Bougara
Boumerdes, Algeria
kbaiche@univ-boumerdes.dz

Yassine Meraihi
LIST Laboratory
University of M'Hamed Bougara
Boumerdes, Algeria
y.meraihi@univ-boumerdes.dz

***Abstract*—The efficient placement of UAV-mounted cameras for surveillance tasks significantly impacts the overall performance and cost of video surveillance systems. Identifying the optimal configuration of these cameras to maximize coverage poses a combinatorial optimization challenge. This paper employs the Particle Swarm Optimization (PSO) algorithm to determine the optimal camera configuration, including positions and orientations, within a multi-UAV network. The goal is to maximize the covered area while minimizing overlapping zones. The proposed algorithm is validated using map images across various scenarios and different numbers of UAV-mounted cameras. Experimental results demonstrate that PSO effectively maximizes coverage, outperforming Single Candidate Optimizer (SCO), Simulated Annealing (SA), and Grey Wolf Optimizer (GWO).**

I. Introduction

Unmanned Aerial Vehicles (UAVs) have emerged as a transformative technology across various sectors, including agriculture, logistics, environmental monitoring, and disaster response. Their ability to operate autonomously or remotely allows for efficient data collection and real-time monitoring of vast areas, making them invaluable in applications where human presence is challenging or impractical. UAVs can be equipped with a variety of sensors, including cameras, light detection and ranging, and thermal imaging, enabling them to gather diverse types of information for analysis and decision-making [1, 2, 3].

Among the most impactful applications of UAVs is the integration of camera systems for surveillance and monitoring tasks. UAV-mounted cameras provide a unique aerial perspective, allowing for enhanced visibility and coverage compared to traditional ground-based systems. This capability is particularly beneficial in scenarios such as security surveillance, wildlife monitoring, and search and rescue operations, where comprehensive situational awareness is crucial.

However, the effectiveness of UAV-mounted camera systems depends on the optimal placement of cameras to maximize coverage while minimizing redundancy and blind spots. This challenge requires sophisticated strategies for configuring cameras, including their positions, orientations, and fields of view, particularly in dynamic environments where conditions can change rapidly. Addressing this issue is critical to ensuring that UAVs perform surveillance tasks both efficiently and effectively.

This problem is proven to be NP-hard and can be effectively addressed using meta-heuristic optimization approaches with reasonable execution times. In this context, we employ the Particle Swarm Optimization (PSO) approach to solve the UAV-mounted camera placement problem, aiming to enhance performance and adaptability.

This paper employs the Particle Swarm Optimization (PSO) algorithm to solve the UAV-mounted camera network placement problem. The proposed approach is validated across multiple scenarios with varying numbers of UAV-mounted cameras. Simulation results demonstrate the efficiency of the PSO approach compared to Simulated Annealing (SA), Single Candidate Optimization (SCO), and Grey Wolf Optimizer (GWO).

The subsequent sections of the paper are organized as follows. Section 2 presents the formulation of the camera placement problem. The description of PSO and SA algorithms can be found in Section 3. Section 4 provides a detailed explanation of the structure of PSO. Section 5 presents the simulation findings. The work is concluded in Section 6.

II. Problem Formulation

The coverage optimization problem can be defined as maximizing the total coverage area while minimizing the number of cameras deployed.

Two primary approaches are used to model camera placement: **deterministic modeling** and **probabilistic modeling**.

- **Deterministic Modeling:** In this approach, coverage is explicitly defined based on fixed parameters such as camera range, angle, and field of view. Per-

formance evaluation is conducted by calculating the total area covered.

- **Probabilistic Modeling:** This approach describes coverage as a probabilistic function influenced by factors such as camera range, visibility, and orientation. It accounts for uncertainties in real-world conditions, making it suitable for dynamic environments.

To reduce computational complexity, we consider a triangular coverage area for each camera. As illustrated in Figure 1a, a camera positioned at the origin o has a range r and a horizontal field of view angle α.

The horizontal angle α of the field of view is given by:

$$\alpha = 2 \times \arctan\left(\frac{W_I}{2f}\right), \quad \tan\left(\frac{\alpha}{2}\right) = \frac{W_I}{2f} \tag{1}$$

where W_I represents the width of the image sensor, and f denotes the camera's focal length.

In Figure 1a, the coordinates of points A and B at the boundary of the field of view are (r, y) and $(r, -y)$, respectively. The value of y is calculated as:

$$y = r \times \frac{W_I}{2f} \tag{2}$$

When the camera's orientation is rotated by an angle θ, the new coordinates of the coverage area vertices are obtained using the following rotation matrix:

$$C = \begin{bmatrix} 0 & 0 \\ r & \frac{W_I}{2f} \\ r & -\frac{W_I}{2f} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3}$$

The main objective of this work is to optimize the covered area by cameras, with coverage being quantified by the number of covered pixels on an image map.

III. Particle Swarm Optimization

Particle swarm optimization algorithm (PSO) is a swarm-based meta-heuristic introduced by Kennedy and Eberhart in 1995 [4]. PSO concept is based on mimicking birds around food or movement of fishes. PSO has demonstrated its effectiveness in solving various optimization problems like the nodes placement problem [**? ?**].

Within the PSO algorithm, a particle is employed to represent a potential solution to the problem at hand. Each particle possesses its own position, denoted as X_i, and velocity, denoted as V_i. These values are updated iteratively throughout the optimization process, guided by a set of predefined rules. During the update process, each particle in the swarm takes into account two important factors: its personal best position (*Pbest*) and the best position achieved by any particle in the entire swarm (*Gbest*). These values are considered when updating the particle's position and velocity, enabling the particle to leverage both its own best performance and the collective best performance of the swarm. The personal best (*Pbest*) of each particle stores the best solution it has encountered so far. It serves as a reference for the particle to guide its movement towards better regions of the search space. On the other hand, the global best (*Gbest*) represents the best solution found by any particle in the entire swarm. It provides a global reference for all particles and influences their exploration and exploitation behavior.

$$V_i(t+1) = wV_i(t) + c_1 r_1 (Pbest_i(t) - X_i(t)) + c_2 r_2 (Gbest_i(t) - X_i(t)) \tag{4}$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{5}$$

where w represents the inertia weight parameter. r_1 and r_2 are random numbers in the range of [0 1]. c_1 and c_2 refer to the acceleration coefficients. The variables t and $t+1$ indicate the current and next iterations of the algorithm, respectively.

The PSO algorithm iteratively updates the positions and velocities of the particles based on these rules, continuously searching for improved solutions. This iterative process continues until a termination criterion is met, such as reaching a maximum number of iterations $tmax$. The pseudo-code of the PSO algorithm is presented in Algorithm 1, summarizing the main steps involved in the optimization process.

IV. Simulation results

In this section, we evaluated the performance of PSO algorithm against SA, SCO, and GWO for the UAV-mounted camera placement problem in a square room with dimensions $100m \times 100m$.

The objective was to maximize the covered area while varying the number of cameras from 2 to 10. The covered area was quantified as the number of non-white pixels obtained by projecting the coverage of UAV-mounted cameras onto an image map. Each camera was defined by two positional coordinates and an orientation, with orientations ranging from -180ř to 180ř.

Simulations were conducted using MATLAB R2020a on a Windows 11 machine with an Intel i5 processor. Each result represents the average of five runs, with a maximum of 150 iterations per run. The population size for both PSO and Grey Wolf Optimizer (GWO) is set to 10.

The results of the UAV-mounted camera placement PSO, SA, SCO, and GWO reveal significant differences in coverage efficiency. The PSO-based placement, shown in Figure 2, provided the most optimal configuration, maximizing coverage while minimizing overlap between the camera views. This demonstrates PSO's ability to effectively balance exploration and exploitation, leading to a superior solution. In contrast, the SA-based placement, depicted in Figure 3, achieved moderate coverage but suffered from noticeable overlap between the cameras, reducing its overall efficiency. The SCO-based placement, shown in Figure 4, resulted in significant coverage gaps, highlighting the limitations of this method in solving

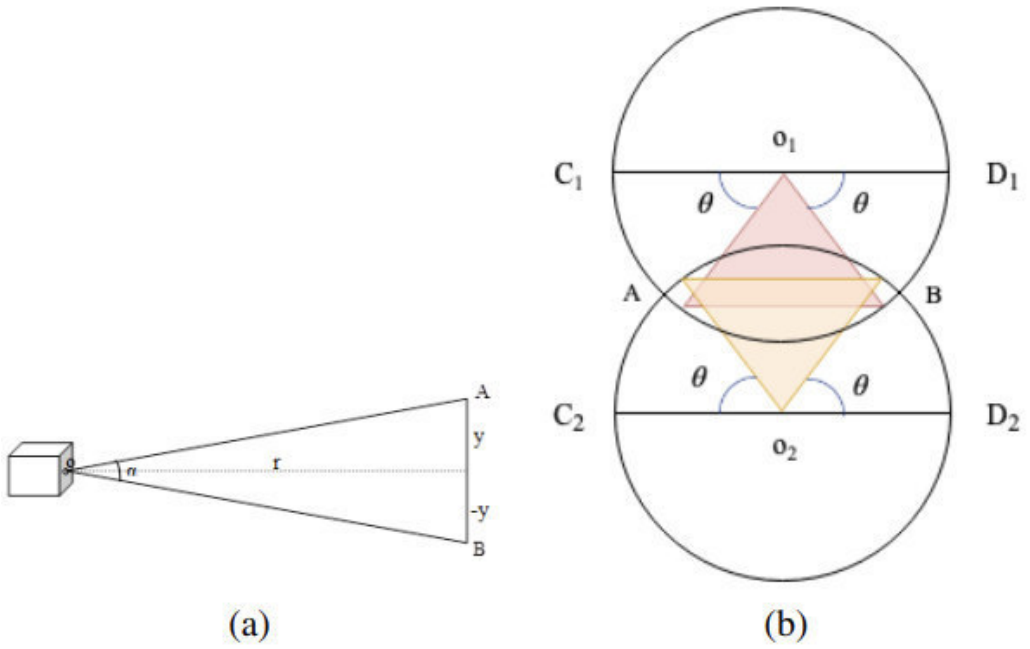


Fig. 1. Modeling the coverage area of a visual surveillance system: (a) Design of a triangular coverage area (b) Modeling coverage overlap between two cameras.

Algorithm 1 The pseudo-code of Particle Swarm Optimization (PSO)

Initialize PSO parameters: acceleration coefficients c_1 and c_2
Initialize the population of N particles (positions and velocities)
Evaluate the population and determine the best solution *Gbest*
t=1
while $(t < tmax)$ **do**
  for i= 1 to N **do**
    Update $V_i(t+1)$ using Equation 4
    Update $X_i(t+1)$ using Equation 5
    Calculate the new fitness value
    Do adaptation
  end for
  t=t+1
end while
Return the *Gbest* particle and its fitness value

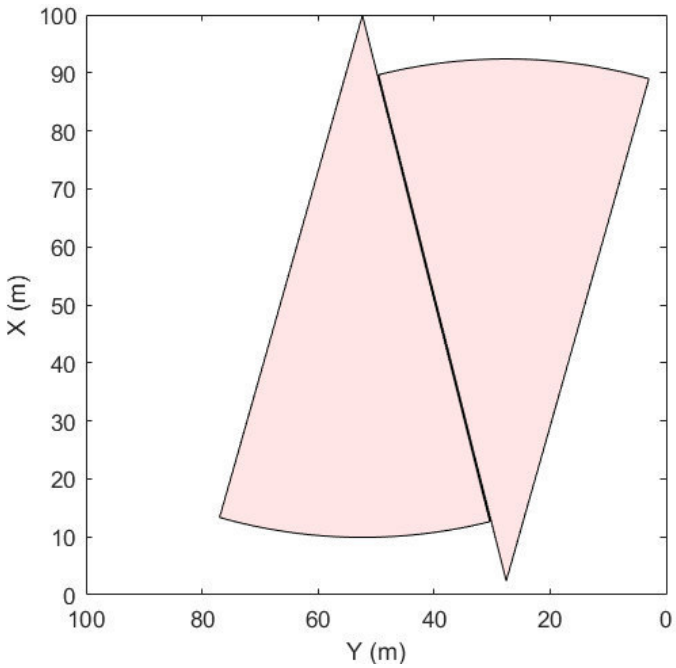


Fig. 2. Placement obtained using PSO

the camera placement problem. Overall, the PSO-based approach outperformed the other techniques, offering the best balance of comprehensive coverage and reduced redundancy, making it the most suitable approach for optimizing the placement of UAV-mounted cameras.

The results presented in Table I and Figure **??** provide a comprehensive overview of the performance of PSO, SA, SCO, and GWO algorithms in terms of coverage and stopping iterations across varying numbers of cameras (2, 4, 6, 8, and 10). The graphical representation is shown in Figure 6

The data indicate a clear positive correlation between the number of cameras and the coverage achieved by both algorithms, attributed to the increased capability of multiple cameras to capture a wider area.

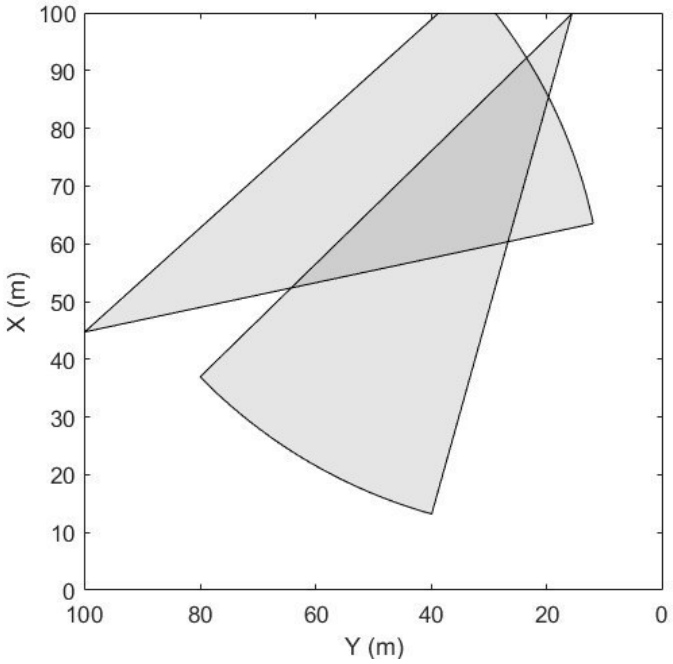


Fig. 3. Placement obtained using SA

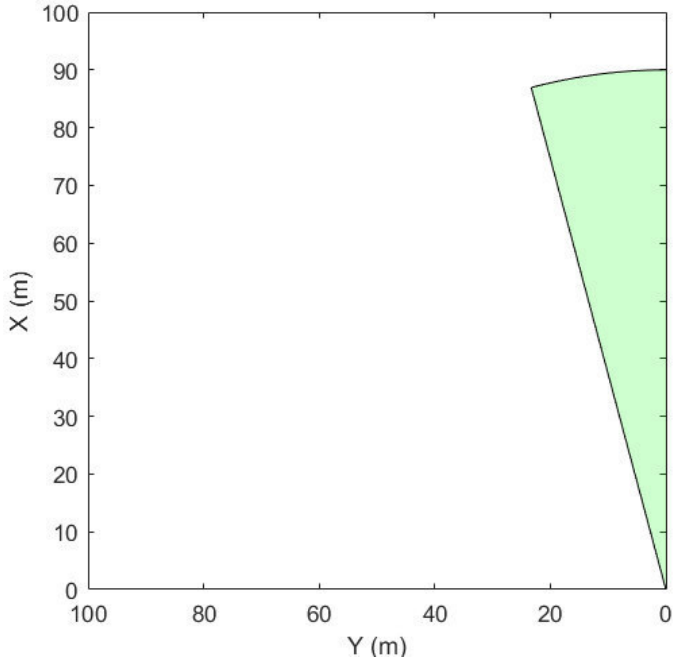


Fig. 4. Placement obtained using SCO

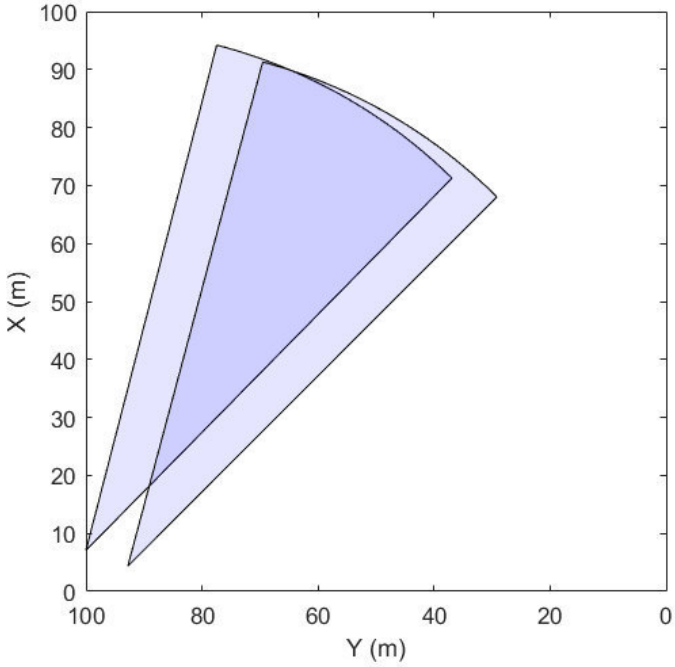


Fig. 5. Placement obtained using GWO

Number of cameras	2	4	6
Coverage (Number of covered pixels)			
PSO	924.4	1483.2	2043.8
SA	684	1393.4	2000.6
SCO	319	601	883
GWO	805	1260.8	2014.4
Stopping Iterations			
PSO	142.8	129.8	139.6
SA	123.6	141.2	148.6
SCO	5	6.4	4.6
GWO	66	133	100.2

TABLE I
COVERAGE AND STOPPING ITERATIONS FOR PSO, SA, SCO, AND GWO, UNDER VARIOUS NUMBERS OF CAMERAS.

Results show that PSO consistently achieves the highest coverage, demonstrating its effectiveness in maximizing coverage. SA also performs well, closely trailing PSO, particularly for larger numbers of cameras. GWO exhibits competitive performance, with coverage values between those of SA and PSO. In contrast, SCO achieves the lowest coverage, indicating its limited optimization ability in this context. These results highlight PSO as the most robust algorithm for achieving optimal coverage, while SCO's efficiency might come at the cost of solution quality.

Regarding stopping iterations, SCO converges the fastest across all scenarios, requiring significantly fewer iterations than the other algorithms, albeit with lower coverage. PSO and SA require more iterations but maintain stable performance across different cases. GWO strikes a middle ground, offering moderate iteration counts while

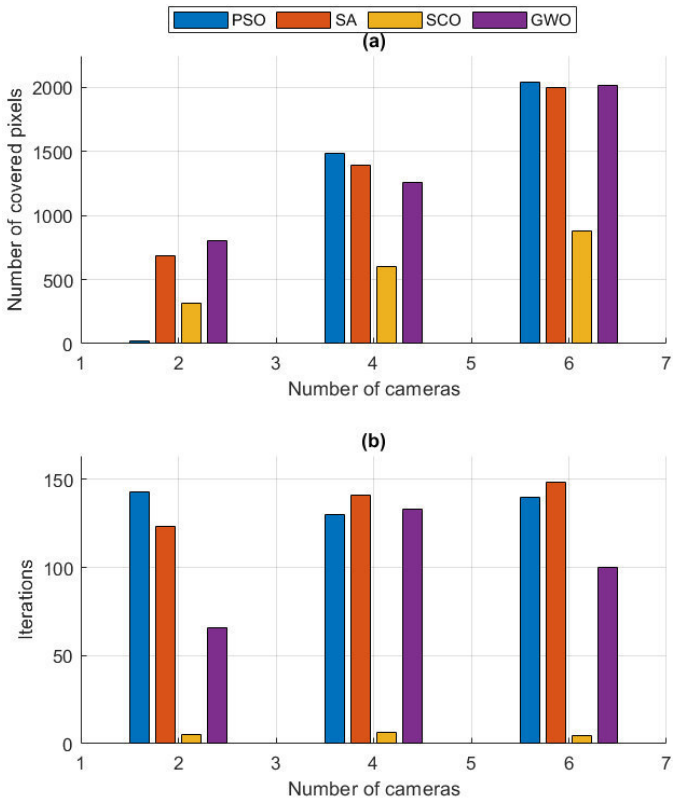


Fig. 6. Coverage and stopping iterations for PSO, SA, SCO, and GWO, under various numbers of cameras

still achieving reasonable coverage. This trade-off suggests that while PSO is best suited for applications prioritizing high coverage, GWO may be preferable for scenarios where both computational efficiency and reasonable coverage are critical. These findings emphasize the importance of aligning algorithm selection with specific application requirements.

V. Conclusion

In this chapter, we have presented PSO as an effective solution to the UAV-mounted camera placement problem. We formulated the problem and described the workings of the SA algorithm in detail. The evaluation of the proposed approach demonstrated its superior performance in terms of coverage compared to SA, SCO, and GWO. The results highlighted PSO's ability to explore the solution space more comprehensively, leading to better coverage outcomes, even though it requires more iterations. This analysis underscores the potential of PSO as a robust method for optimizing UAV-mounted camera placement, paving the way for further research into its application in various scenarios and the exploration of hybrid strategies that could enhance its efficiency.

References

[1] C. Zhang and J. M. Kovacs, "The application of small unmanned aerial systems for precision agriculture: a review," *Precision agriculture*, vol. 13, pp. 693–712, 2012.

[2] K. Anderson and K. J. Gaston, "Lightweight unmanned aerial vehicles will revolutionize spatial ecology," *Frontiers in Ecology and the Environment*, vol. 11, no. 3, pp. 138–146, 2013.

[3] I. Colomina and P. Molina, "Unmanned aerial systems for photogrammetry and remote sensing: A review," *ISPRS Journal of photogrammetry and remote sensing*, vol. 92, pp. 79–97, 2014.

[4] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, vol. 4. IEEE, 1995, pp. 1942–1948.

Lung and Colon Cancer Histopathological Image Classification Using Deep Learning Approaches

1st Oussama Bouguerra
LASS Laboratory
Department of Electronics
University of M'sila
M'sila, Algeria
oussama.bouguerra@univ-msila.dz

2nd Youcef Brik
LASS Laboratory
Department of Electronics
University of M'sila
M'sila, Algeria
youcef.brik@univ-msila.dz

3rd Bilal Attallah
LASS Laboratory
Department of Electronics
University of M'sila
M'sila, Algeria
bilal.attallah@univ-msila.dz

***Abstract*—This paper investigates the application of deep learning for multi-class classification of lung and colon cancer histopathological images using the LC25000 dataset. The dataset contains 25,000 images equally distributed across five classes: benign and adenocarcinoma for both lung and colon tissues, and squamous cell carcinoma for lung tissue. Preprocessing involved converting labels to one-hot encoding and removing 1280 identified duplicate images to prevent data leakage. Three models were trained and evaluated: a baseline CNN, an enhanced CNN with architectural modifications and regularization, and a transfer learning model utilizing ResNet50. The baseline model, using Adam optimizer and categorical cross-entropy loss, achieved a test accuracy of 62.6% and validation accuracy of 63.7%. The enhanced model, incorporating increased depth, adjusted kernel size, L2 regularization, dropout, and the Adamax optimizer with early stopping, reached a test accuracy of 97.9% and validation accuracy of 98.1%. Finally, the transfer learning model with ResNet50, fine-tuned with additional dense layers, dropout, and early stopping, achieved near-perfect performance with 98.9% test accuracy and 99.3% validation accuracy. This study demonstrates the significant performance improvement achieved through architectural enhancements and transfer learning, highlighting the potential of deep learning for automated diagnosis of cancer from histopathological images. The removal of duplicate images proved essential for accurate performance evaluation and preventing artificially inflated results due to data leakage.**

***Index Terms*—Histopathological Image Classification,Lung and Colon Cancer,Convolutional Neural Network (CNN),Transfer Learning,ResNet50.**

I. Introduction

Lung and colon cancers are among the leading causes of cancer-related mortality worldwide, underscoring the critical need for early detection to improve patient outcomes. Histopathological imaging plays a pivotal role in the diagnosis of these cancers, as it allows for the detailed examination of tissue samples to identify malignant cells. However, the high variability in tissue structures and staining techniques presents significant challenges in accurately interpreting these images. Studies have highlighted the importance of histopathological examination in colorectal cancer, emphasizing that early detection can significantly expand treatment options and reduce mortality rates [1]. Similarly, the manual analysis of histopathology images for lung cancer detection is labor-intensive and highly dependent on the pathologist's expertise, which can lead to variability in diagnostic accuracy [2]. These challenges necessitate the development of automated, reliable methods to enhance the precision and efficiency of cancer diagnosis. Deep learning, particularly convolutional neural networks (CNNs), offers promising solutions to these challenges by improving the accuracy of histopathological image analysis. CNNs have demonstrated impressive capabilities in image processing tasks, including medical imaging, by learning complex patterns and features from large datasets [2]. Transfer learning, using architectures like ResNet, has been particularly effective in leveraging pre-trained models to enhance classification performance, even with limited labeled data [1], [2]. This approach allows for the adaptation of models trained on large, diverse datasets to specific tasks, such as cancer detection, thereby improving generalization and reducing the need for extensive annotated datasets [3]. Despite these advancements, challenges remain, such as the need for models to generalize across different centers and staining protocols, which can lead to overfitting if not properly addressed [4]. Nonetheless, the integration of deep learning techniques in histopathological image analysis holds significant potential for advancing cancer diagnostics and improving patient care.

II. Related Work

Convolutional Neural Networks (CNNs) have revolutionized medical imaging, offering significant advancements in classification, segmentation, and detection tasks. Seminal architectures like AlexNet, VGG, and ResNet have laid the groundwork for these applications, demonstrating their efficacy in various medical domains. For instance, CNNs have been pivotal in enhancing diagnostic accuracy and efficiency in radiology and pathology, where they have been used to automate the analysis of complex medical images, thus reducing the reliance on manual interpretation and improving early disease detection [5]. In the context of cancer histopathology, CNNs have been employed to classify lung and colon cancer images, leveraging their ability to learn intricate patterns from large datasets.

Recent studies have focused on applying deep learning, particularly CNNs, to classify cancer from histopathological images, with a strong emphasis on lung and colon cancer.

For example, the BIC-SGODL method utilizes DenseNet and convolutional LSTM to capture complex features and spatiotemporal information, achieving high performance in cancer diagnosis [6]. Another study compared various pre-trained models, including VGG-16 and ResNet, on the LC25000 dataset, achieving accuracies up to 100% for certain categories, highlighting the effectiveness of transfer learning in this domain [2]. Additionally, novel architectures like the 1D Convolutional Channel-based Attention Networks have been proposed, achieving state-of-the-art performance with minimal computational resources [7]. These studies underscore the potential of deep learning models to significantly improve the accuracy and efficiency of cancer diagnosis.

Transfer learning has emerged as a powerful technique in histopathology, allowing models to leverage pre-trained weights from large datasets to improve performance on specific tasks. For instance, the Inception-ResNetV2 model, combined with local binary pattern features, achieved 99.98% accuracy in lung and colon cancer detection, demonstrating the utility of integrating texture-based features with deep learning [8]. Similarly, EfficientNet variants, when applied with transfer learning, achieved notable accuracy improvements, with EfficientNetB2 reaching 97% accuracy on the LC25000 dataset [9]. These approaches highlight the advantages of transfer learning in overcoming challenges such as limited labeled data and enhancing model generalizability.

Future research directions include developing more robust models that can handle diverse data sources and improving the interpretability of deep learning models to facilitate their integration into clinical workflows [5]. Additionally, exploring novel architectures and training strategies could further enhance the performance and applicability of these models in real-world settings.

III. Data Collection and Preprocessing

This section details the dataset used, preprocessing steps performed, and exploratory data analysis conducted.

A. Dataset Description

This study leverages the LC25000 Lung and Colon Cancer Histopathological Image Dataset introduced by [10], a publicly accessible resource containing 25,000 RGB histopathological images across five balanced classes: Colon Benign Tissue, Colon Adenocarcinoma, Lung Benign Tissue, Lung Adenocarcinoma, and Lung Squamous Cell Carcinoma, with each class comprising 5,000 images. Figure 1 showcases representative images from eachclass, visually highlighting the diversity of histopathological features within thedataset.

B. Data Preprocessing

Data cleaning and preprocessing were essential steps in preparing the histopathological images for effective model training. The following methods were applied to ensure data quality, standardize input, and optimize the dataset for Convolutional Neural Networks (CNNs):

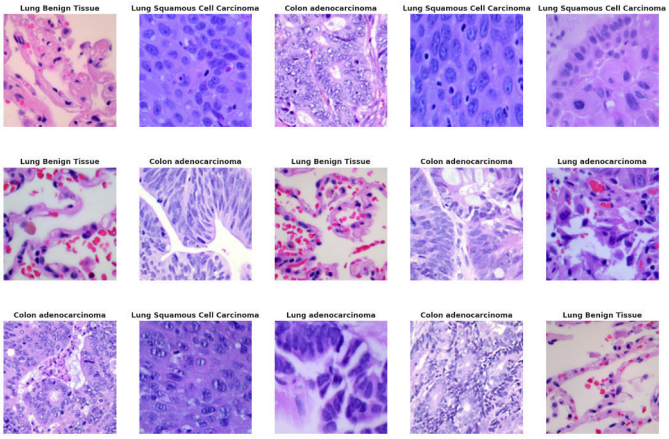


Fig. 1: Visual examples of histopathology slides from the LC25000 dataset.

- **Data Cleaning and Duplicate Removal**: An initial exploratory data analysis (EDA) assessed class distribution and identified duplicate images. This process removed 1,280 duplicates, resulting in a final dataset of 23,720 images (See Figure 2).
- **Image Resizing**: All images were resized to 120x120 pixels using LANCZOS resampling, balancing computational efficiency with preservation of critical histopathological details for feature extraction.
- **Normalization**: Although not explicitly applied in the code, normalization is recommended to scale pixel values between 0 and 1 (or -1 and 1). This practice stabilizes model training and prevents dominant features from skewing the learning process.
- **One-Hot Encoding of Target Classes**: To support multi-class classification, categorical target variables were transformed into binary vectors using one-hot encoding, preventing unintended ordinal relationships among diagnostic categories.
- **Train-Validation-Test Split**: The dataset was divided into training, validation, and test sets in a 60%-20%-20% ratio, ensuring balanced class representation across all phases (Figure 3).

IV. Methodology

This section details the methodologies employed for developing and evaluating the deep learning models for lung and colon cancer histopathological image classification. Figure 4 provides an overview of the proposed system, illustrating the workflow from dataset acquisition and preprocessing through model training, evaluation, and final classification.

A. Baseline Model

A baseline Convolutional Neural Network (CNN) model was established to serve as a benchmark for subsequent, more complex architectures. This model provides a foundation for evaluating the performance gains achieved through architectural modifications and transfer learning. The baseline

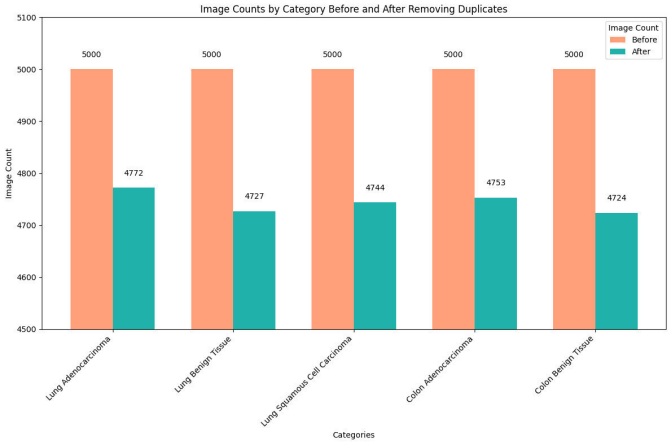


Fig. 2: Image Counts by Category Before and After Removing Duplicates.

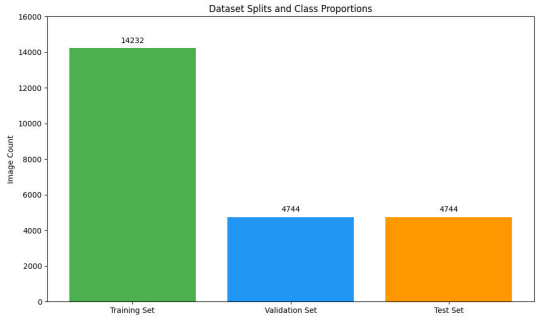


Fig. 3: Dataset Splits and Class Proportions.

model architecture, comprising convolutional, max pooling, and dense layers, is summarized in Figure 5.

B. Enhanced Model

To improve upon the baseline CNN, the enhanced model introduces additional convolutional layers, regularization techniques, and an optimized architecture to further capture the intricate patterns present in histopathological images. These modifications aim to reduce overfitting and enhance classification accuracy across the five cancer-related categories. The enhanced CNN architecture, incorporating increased depth, adjusted kernel size, and regularization, is detailed in Table I.

C. Transfer Learning Model (ResNet50)

we employed ResNet50 a deep residual network pre-trained on the ImageNet dataset, as the foundation for our transfer learning model. The architecture of the transfer learning model, utilizing a pre-trained ResNet50 base and custom classification layers, is depicted in Figure 6. This figure outlines the layer configurations and the distinction between trainable and non-trainable parameters.

D. Training Configuration

This section provides an overview of the key training configurations used across the baseline, enhanced, and transfer learning (ResNet50) models. The configurations include essential hyperparameters, optimizers, learning rates, batch sizes, and epochs, highlighting the specific settings employed to achieve optimal performance for each model (see Table II).

TABLE I: Enhanced Model Architecture

Layer (type)	Output Shape	Param #
conv2d_4 (Conv2D)	(None, 120, 120, 64)	1792
conv2d_5 (Conv2D)	(None, 120, 120, 64)	102464
max_pooling2d_4 (MaxPooling2D)	(None, 60, 60, 64)	0
conv2d_6 (Conv2D)	(None, 60, 60, 128)	73856
conv2d_7 (Conv2D)	(None, 60, 60, 128)	147584
max_pooling2d_5 (MaxPooling2D)	(None, 30, 30, 128)	0
conv2d_8 (Conv2D)	(None, 30, 30, 256)	295168
conv2d_9 (Conv2D)	(None, 30, 30, 256)	590080
max_pooling2d_6 (MaxPooling2D)	(None, 15, 15, 256)	0
conv2d_10 (Conv2D)	(None, 15, 15, 512)	1180160
conv2d_11 (Conv2D)	(None, 15, 15, 512)	2359808
max_pooling2d_7 (MaxPooling2D)	(None, 7, 7, 512)	0
flatten_2 (Flatten)	(None, 25088)	0
dense_6 (Dense)	(None, 256)	6422784
dropout (Dropout)	(None, 256)	0
dense_7 (Dense)	(None, 64)	16448
dense_8 (Dense)	(None, 5)	325
Total params:	11190469 (42.69 MB)	
Trainable params:	11190469 (42.69 MB)	
Non-trainable params:	0 (0.00 Byte)	

V. Result

The results of this study are based on the evaluation of three deep learning models—baseline CNN, enhanced CNN, and transfer learning with ResNet50—using a diverse set of performance metrics.

A. Model Performance Comparison

Each model was assessed on accuracy, precision, recall, and F1-score, providing a comprehensive view of performance across categories.The bar graph in Figure 7 further highlights these performance improvements across models, illustrating the ResNet50 model's significant gains over the baseline and enhanced CNN models.

B. Confusion Matrix Analysis

The confusion matrices for each model illustrate the model's effectiveness in differentiating between benign and malignant tissues. Misclassifications were notably reduced in the enhanced and ResNet50 models compared to the baseline.

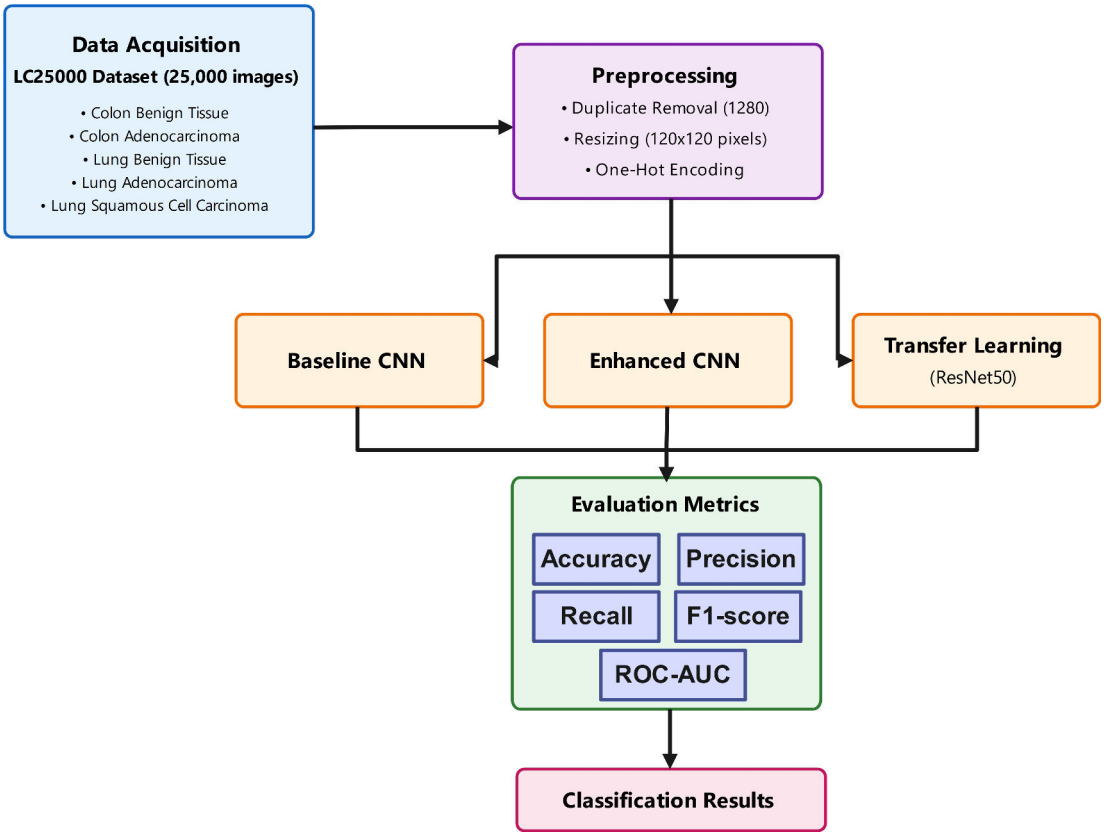


Fig. 4: Proposed System Diagram.

TABLE II: Training Configurations for Each Model

Model	Optimizer	Learning Rate	Batch Size	Epochs	Regularization	Early Stopping
Baseline Model	Adam	Default	32	10	None	No
Enhanced Model	Adamax	0.0005	32	30	L2 Regularization, Dropout (0.2)	Patience 2
Transfer Learning Model (ResNet50)	Adamax	0.0005	32	30	Dropout (0.2)	Patience 3

C. ROC-AUC Analysis

ROC-AUC curves were generated for each model to evaluate the models' abilities to distinguish between classes. The ResNet50 model achieved an AUC close to 1.0 across all categories, indicating strong discriminative power.

D. Discussion

The ResNet50 transfer learning model achieved the highest performance across metrics, followed by the enhanced CNN model. The baseline model, while effective as an initial benchmark, demonstrated limited accuracy and a higher rate of misclassification. The results underscore the impact of deep learning architectures and transfer learning in handling complex histopathological data, with the ResNet50 model emerging as the optimal approach for cancer classification in this study.

VI. Conclusion

This paper evaluated lung and colon cancer classification using three models: a baseline CNN, an enhanced CNN, and a ResNet50-based transfer learning model. The ResNet50 model achieved the highest performance, with 98.9% accuracy and strong precision, recall, and F1-scores, demonstrating its superior capability in feature extraction for complex histopathological images. These results confirm the value of transfer learning in medical image analysis and set a foundation for future advancements in automated cancer diagnosis.

Acknowledgment

The authors would like to extend their gratitude to the members of LASS Laboratory (University of M'sila, Algeria) for their valuable support and assistance in publishing this work.

Layer (type)	Output Shape	Param #
conv2d_2 (Conv2D)	(None, 120, 120, 128)	3584
max_pooling2d_2 (MaxPooling2D)	(None, 60, 60, 128)	0
conv2d_3 (Conv2D)	(None, 60, 60, 64)	73792
max_pooling2d_3 (MaxPooling2D)	(None, 30, 30, 64)	0
flatten_1 (Flatten)	(None, 57600)	0
dense_3 (Dense)	(None, 128)	7372928
dense_4 (Dense)	(None, 32)	4128
dense_5 (Dense)	(None, 5)	165

Total params: 7454597 (28.44 MB)
Trainable params: 7454597 (28.44 MB)
Non-trainable params: 0 (0.00 Byte)

Fig. 5: Summary of the baseline CNN architecture.

Model: "sequential_3"

Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 4, 4, 2048)	23587712
global_average_pooling2d_1 (GlobalAveragePooling2D)	(None, 2048)	0
dense_9 (Dense)	(None, 256)	524544
dropout_1 (Dropout)	(None, 256)	0
dense_10 (Dense)	(None, 64)	16448
dense_11 (Dense)	(None, 5)	325

Total params: 24129029 (92.04 MB)
Trainable params: 541317 (2.06 MB)
Non-trainable params: 23587712 (89.98 MB)

Fig. 6: Architecture of the transfer learning model using ResNet50.

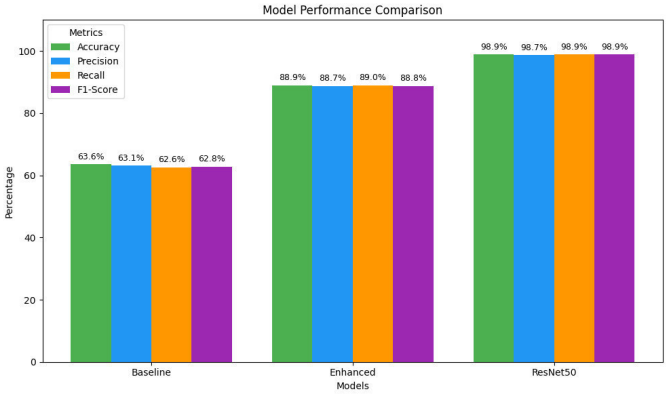
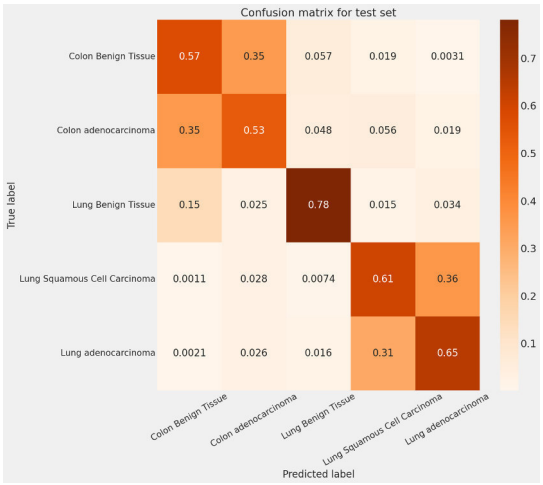
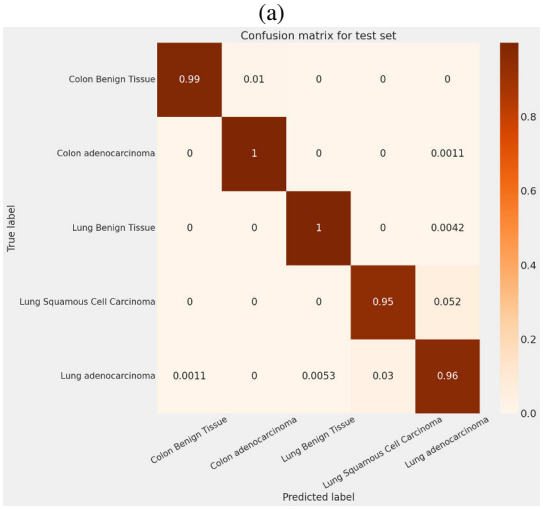


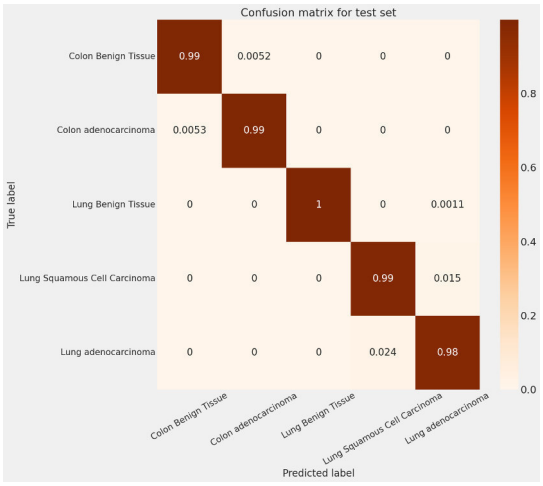
Fig. 7: Bar graph comparison of model performance metrics.



(a)

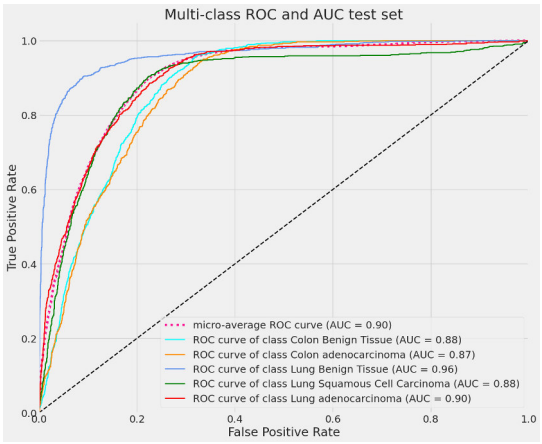


(b)

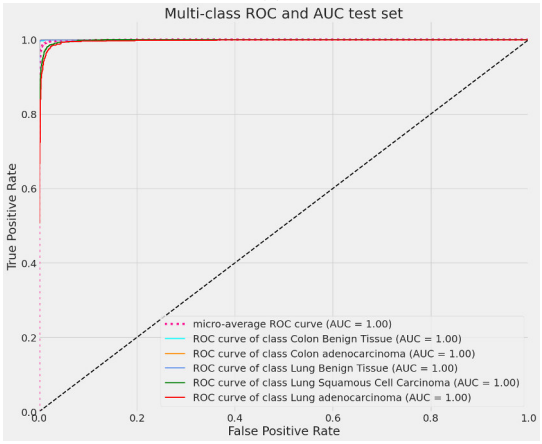


(c)

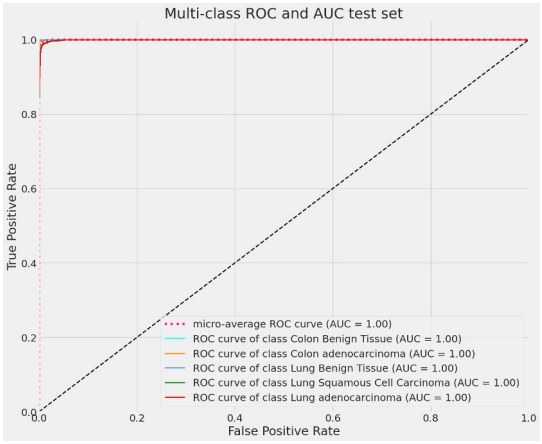
Fig. 8: Confusion matrix for (a:baseline model,b:enhanced model,c:ResNet50 model)



(a)



(b)



(c)

Fig. 9: ROC-AUC curves for (a:baseline model,b:enhanced model,c:ResNet50 model)

REFERENCES

[1] Dina M Ibrahim, Mohammad Ali A Hammoudeh, and Tahani M Allam. Histopathological cancer detection based on deep learning and stain images. *Indonesian Journal of Electrical Engineering and Computer Science*, 36(1):214–230, 2024.

[2] Abdelwahid Oubaalla, Hicham El Moubtahij, and Nabil EL AKKAD. Detection of lung and colon cancer from histopathological images: Using convolutional networks and transfer learning. *International Journal of Computing and Digital Systems*, 16(1):583–595, 2024.

[3] Andrey Ignatov, Josephine Yates, and Valentina Boeva. Histopathological image classification with cell morphology aware deep neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6913–6925, 2024.

[4] Maryam Asadi-Aghbolaghi, Amirali Darbandsari, Allen Zhang, Alberto Contreras-Sanz, Jeffrey Boschman, Pouya Ahmadvand, Martin Köbel, David Farnell, David G Huntsman, Andrew Churg, et al. Learning generalizable ai models for multi-center histopathology image classification. *NPJ Precision Oncology*, 8(1):151, 2024.

[5] Anirudh Atmakuru, Subrata Chakraborty, Oliver Faust, Massimo Salvi, Prabal Datta Barua, Filippo Molinari, UR Acharya, and Nusrat Homaira. Deep learning in radiology for lung cancer diagnostics: A systematic review of classification, segmentation, and predictive modeling techniques. *Expert Systems with Applications*, page 124665, 2024.

[6] Thiyagarajan Manoharan, Ramalingam Velvizhi, Tarun Kumar Juluru, Shoaib Kamal, Shrabani Mallick, and Ezudheen Puliyanjalil. Biomedical image classification using seagull optimization with deep learning for colon and lung cancer diagnosis. *Indonesian Journal of Electrical Engineering and Computer Science*, 2024.

[7] Nazmul Shahadat. Lung and colon cancer histopathological image classification using 1d convolutional channel-based attention networks. *The International FLAIRS Conference Proceedings*, 2024.

[8] Shtwai Alsubai. Transfer learning based approach for lung and colon cancer detection using local binary pattern features and explainable artificial intelligence (ai) techniques. *PeerJ Computer Science*, 10:e1996, 2024.

[9] Sunila Anjum, Imran Ahmed, Muhammad Asif, Hanan Aljuaid, Fahad Alturise, Yazeed Yasin Ghadi, and Rashad Elhabob. Lung cancer classification in histopathology images using multiresolution efficient nets. *Computational Intelligence and Neuroscience*, 2023(1):7282944, 2023.

[10] A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides. Lc25000 lung and colon histopathological image dataset, 2023. Retrieved from https://github.com/tampapath/lung$_c olon_i mage_s et$.

Breast cancer masses segmentation from mammography images using adaptive thresholding and morphological filtering methods

1st Lilia Radjef*, 2nd Tahar Omari†, 3rd Karim Baiche*
*Electrical Systems Engineering Department, M'hamed Bougara University, Boumerdes, Algeria
l.radjef@univ-boumerdes.dz, kbaiche@univ-boumerdes.dz
†Electrical Engineering and Telecommunications Department, Ain Temouchent University, Ain Temouchent, Algeria
omari.tahar@univ-boumerdes.dz

***Abstract*—Breast cancer (BC) is the second leading cause of death among women worldwide. Mammography(MM) is an effective X-ray imaging technique. It considerably increases the early detection and probability of patient survival. Mammographic image segmentation is a powerful tool for identifying cancerous areas, extracting features, and improving diagnosis. However, mammographic image's complex intensity distribution conditions need precise segmentation techniques. A simple algorithm has been implemented in a Matlab environment to segment cancerous masses from mammography images. The proposed method is based on adaptive thresholding and morphological filtering and is tested on 322 images from (MIAS) and 100 from (Mini-DDSM) databases. Adaptive thresholding involves dividing the image into smaller sub-regions and calculating a threshold value for each sub-region. Morphological filtering eliminates irrelevant small structures and enhances contour detection clarity. The images are classified into three categories: normal, benign, and malignant based on the geometric criteria of cancerous masses. For the (MIAS), the algorithm's accuracy is 97.45% and precision 98.41%. The accuracy and precision are 97% and 98.96 % respectively for the (Mini-DDSM) database. Results demonstrate the effectiveness of our segmentation approach.**

***Index Terms*—Breast cancer, Mammographic image, Cancerous mass segmentation, Adaptive thresholding, Morphological filtering, Signal processing, Matlab, Mass classification.**

I. Introduction

According to the World Health Organisation (WHO), cancer could overtake cardiovascular disease as the leading cause of premature death worldwide [1]. Breast cancer (BC) is a frequently diagnosed disease in women, accounting for 25% of all female cancers [2]. In Algeria, 14,000 new cases of breast cancer are diagnosed each year [3]. Early detection of this dangerous disease is essential to increase the chances of cure and extend patients' life expectancy [4].

A. Breast cancer

The breast constitutes lobules, ducts, connective tissue, adipose tissue, and lymphatic tissue [5]. Breast cancer (BC) results from the abnormal growth and multiplication of cells forming a mass, micro-calcifications, or architectural distortions called malignant tumours [6]. It develops in or around breast tissue, mainly in the ducts and milk glands [7]. Cancerous masses appear as dense regions of different sizes, contours, and properties[8]. Those with smooth, circumscribed edges are benign, while malignant masses generally have rough, fuzzy edges (Fig.1). It can be detected by a self-examination, clinical examination, or radiological diagnosis [9]. Most breast cancers are invasive, spreading beyond the breast tissue and metastases to other parts of the body[10].

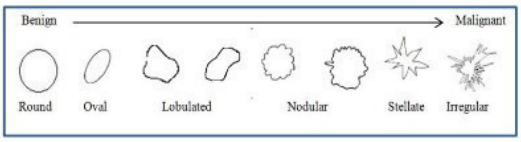


Fig. 1. Benign and malignant tumor contours.

Breast cancer symptoms include a palpable or visible lump in the breast, swollen lymph nodes, discharge from the nipple, heaviness, redness, swelling, deformity, or retraction of the breast [11]. It is associated with several risk factors, including family history, age, menopause, smoking, alcohol consumption, radiation exposure, and obesity [12]. External factors such as lifestyle, environment, and stress contribute to its development [13]. This pathology is morphologically and molecularly heterogeneous, and patients require different treatments, such as radiotherapy, chemotherapy, hormone therapy, immunotherapy, or targeted treatment [14-15-16]. The biopsy can be used as an invasive method for breast cancer detection. However, the tissue extraction procedure is painful and time-consuming favoring non-invasive methods [17].

B. Breast cancer imaging modalities

Various imaging modalities have been used for breast cancer screening, detection and diagnosis, such as magnetic resonance imaging (MRI)[18], ultrasound (US) [19], nuclear medicine imaging [20], computed tomography, positron emission tomography (PET) [21], and thermography [22] or their combinations.

Mammography(MM) is a low-energy (generally around 30 kVp) X-ray medical imaging technique[23]. It is widely used worldwide to screen breasts and anomalies diagnostics[24]. Mammography is commonly classified into film-screen mammography (FSM) and digital mammography (DMM)[25]. (FSM) is the standard imaging method because of its high sensitivity in detecting lesions[26]. Mammographic images are two-dimensional(2D) representations of the compressed breast, showing components such as fatty tissue, hydrated support, and glandular tissue [27]. They are classified into four views: left mediolateral oblique (LMLO) and left craniocaudal (LCC), right mediolateral oblique (RMLO) and right craniocaudal (RCC)[28]. Digital mammography has the advantages of better image quality, being less invasive, and offering several breast views over traditional methods, depending on the X-ray angle projection onto the tissue [29]. Screening mammography has the limitations of lower reliability in young women with dense breasts or who have undergone breast surgery [30]. Mammography images can be found in several formats, including LJPG, DI- COM, PGM, and TIFF[31].

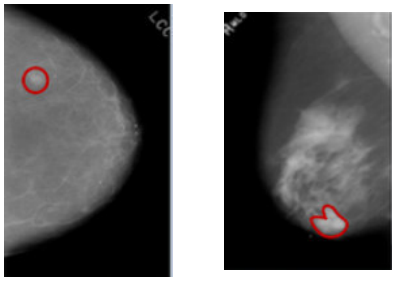


Fig. 2. Mammographic images tumours: (left) Benign and (right) malignant.

C. Mammographic image segmentation

Segmentation divides an image into regions with similar grey-level properties as colour, texture, brightness, and contrast[32]. Mammography image segmentation is crucial and is a key step in improving visual image quality, by providing homogeneous regions [33]. It reduces noise and accurately locates image features, region of interest(ROI), and the cancerous area's contours[34]. Three segmentation methods are currently used for mammography: (1) conventional [35]: Through edge-based methods (e.g. Canny edge detection, active edges), threshold-based methods (e.g. Otsu thresholding, adaptive thresholding), and region-based methods (e.g. watershed, region growth)[36]. (2) Machine learning[37]: including unsupervised methods (fuzzy clustering and k-means)[38-39], and supervised methods (support vector machines (SVMs) and extreme learning machines. (3) Deep learning [40]: such as (SegNet, U-Net[41], and fully convolutional neural networks (FCNs)[42]. The segmentation choice technique is crucial for detecting regions of interest (ROI)[43]. The selection of the appropriate technique depends on various factors, such as image modality, noise levels, and application[44].

Our work aims to segment cancerous masses from mammographic images. In the following section, we present a simple approach using image processing techniques: adaptive thresholding and morphological filtering to identify the mass area and contour's shape. This is used to classify images into three categories: normal, benign, and malignant.

II. Methodology

The proposed method is divided into four steps as shown in (Fig.3).

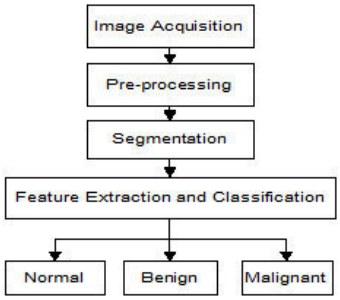


Fig. 3. Flowchart of the applied segmentation approach.

A. Image acquisition

For testing the proposed algorithm, we selected:

Mammographic Image Analysis Society Database (MIAS)

Mammographic Image Analysis Society database (MIAS) [45] results from the efforts of British research groups interested in mammography. It contains the left and right breast images of 161 patients, a total of 322 images. These images are classified into three (3) categories: normal, benign, and malignant. Pixels are coded on 8 bits and all images are cropped to 1024x1024. Mammogram names are in the following format: mdb XXXBS, where: - XXX: is the image number, from 001 to 322.

Digital Database Mammography (Mini-DDSM)

It was created in 1999 in the USA, the largest public mammography database in the world. It contains 2620 cases comprising two images (MLO and CC) of each breast, a total of 10480 pathologies, from normal images to those containing benign and malignant masses. Information on patient age, breast density, and classification (BI-RADS) is provided for each image[46].

B. Pre-processing

Image noise refers to random variations in brightness or color in images produced by the sensor and circuitry or a digital camera[47]. The image is processed at various stages to improve quality, and contrast, and reduce noise.

Median filter

The non-linear filter is an effective noise-removing method from mammographic images while preserving the contours[48].

Let S_{xy} represent the set of coordinates in a rectangular sub-image window of size $m \times n$ centred at the point (x, y). The arithmetic mean filtering process computes the average value of the corrupted image $g(x, y)$ in the area defined by S_{xy}. The value of the restored image $\hat{f}$ at any point (x, y) is simply the arithmetic mean computed using the pixels in the region defined by S_{xy}. In other words,

$$\hat{f}(x, y) = \frac{1}{mn} \sum_{(s,t) \in S_{xy}} g(s, t) \tag{1}$$

The median filter replaces the value of a pixel with the median of the grey levels in the vicinity of that pixel.

$$\hat{f}(x, y) = \text{median} \{g(s, t)\}_{(s,t) \in S_{xy}} \tag{2}$$

The original pixel value is included in the median calculation. Median filters are popular because they provide excellent noise reduction.

C. Segmentation

Adaptive thresholding

This method consists of dividing the image into regions and applying a different threshold for each area depending on the intensity level of the pixels in the region. This technique involves separating the image into sub-images and processing each with its threshold[49]. Adaptive thresholding differentiates dense structures in homogeneous areas[50].

$$\text{Pixel(x, y)} = \begin{cases} 1, & \text{if} I(x, y) > T(x, y) \\ 0, & \text{elsif} \end{cases}$$

- $I(x, y)$: Intensity of pixel at position(x, y) in the image.
- $T(x, y)$: Local adaptive threshold calculated around the pixel (x, y).

Morphological filters

Mathematical morphology was introduced in 1975 by Georges Matheron [51]. It is a powerful tool for signal and image analysis. A morphological operation (erosion, dilation, opening, closing) is applied to extract components from images that are useful for representation or description and to perform pre- or post-processing[52]. The use of morphological filters is more appropriate than linear filters for shape analysis and plays a role in geometry-based enhancement, detection, and noise reduction[53]. The morphological closure of an image A with a structuring element B is defined as follows :

$$A \bullet B = (A \oplus B) \ominus B$$

D. Feature extraction and classification

We classified the tumor as normal, malignant, or benign. According to the geometric criteria of the regions of interest(ROI): surface area and the contour shape of tumors. It has several shapes: round, oval, lobulated, nodular, stellate, and irregular. More irregular shapes are malignant, with a larger surface area. On the other hand, around, regular and smooth boundaries are benign.

E. Algorithm implementation

We implemented the breast cancer detection algorithm in MATLAB version 2019 (Fig.4). After image importation from (MIAS) and(Mini-DDSM) databases using the imread function, they are converted to greyscale using rgb2gray (if necessary) and displayed using imshow. To improve visibility, contrast is adjusted using imadjust. A median filter is applied to reduce noise. Adaptive thresholding segmentation using adaptthresh to obtain a binary image, and small noise regions are removed using bwareaopen. Then, morphological filtering eliminates irrelevant small structures and enhances contour detection clarity. The morphological features of the regions of interest, such as contour, shape, and surface area are extracted using regionprops. The images are classified into three categories: normal, benign, and malignant.

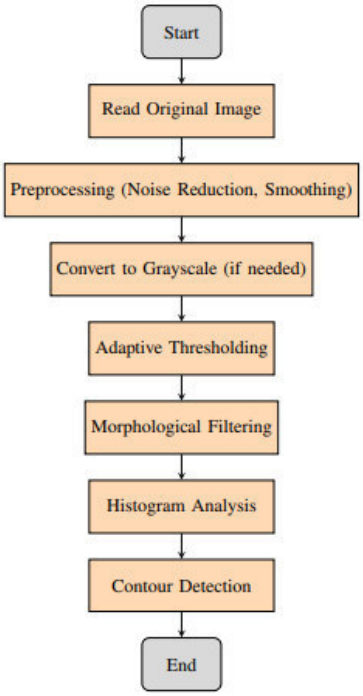


Fig. 4. Flowchart for image segmentation algorithm.

III. Results and Discussion

A. Segmentation results

Figures 5 and 6 illustrate the segmentation stages: (1) Original image. (2) Filtered image: Application of the median filter. (3) Adaptive thresholding is applied to generate a binary image. (4) Small area filtering: The binary image is filtered to eliminate irrelevant areas. (5)The greyscale histogram represents the distribution of intensities in the greyscale image after thresholding. (6) Segmentation: Regions of interest (cancer masses) are identified from the filtered image.

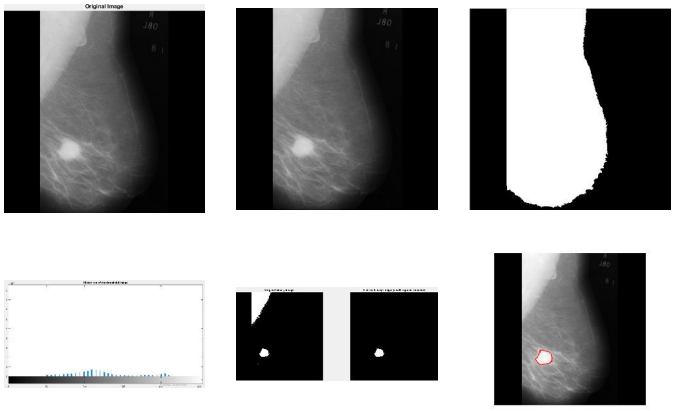


Fig. 5. Segmentation results of the mammographic image.

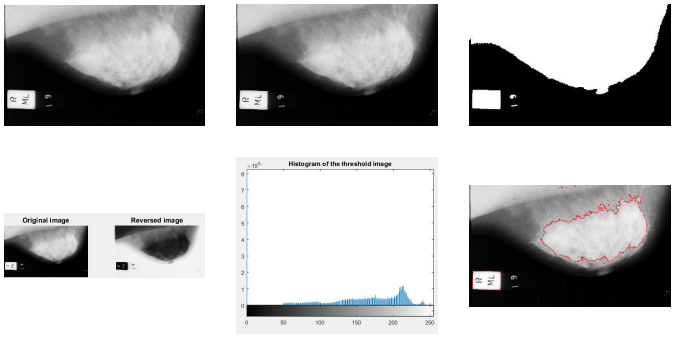


Fig. 6. Malignant tumor segmentation result.

B. Image classification

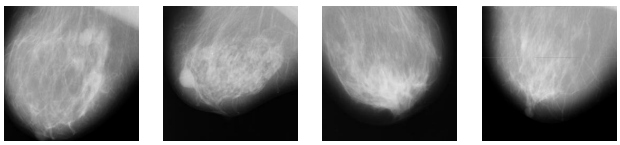


Fig. 7. Benign classification tumor.

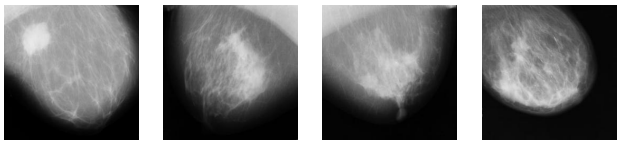


Fig. 8. Malignant classification tumor.

Figures 7, 8 and 9 show examples of mammographic image classifications based on geometric criteria of the segmented tumours.

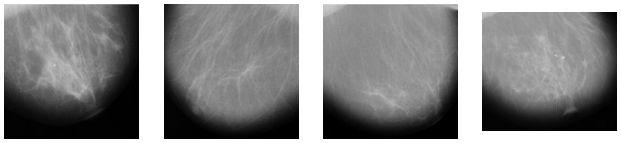


Fig. 9. Normal breast.

C. Segmentation performance

In our work, the algorithm segmented the breast masses on almost all the images.

1) Accuracy : The accuracy of the algorithm is expressed by the equation:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

TP is the correct segmentation.
TN is not correct segmentation.
FN is the under-segmentation.
FP is the over-segmentation.

2) Precision : Precision is defined as

$$P = \frac{TP}{TP + FP}$$

The result shows that out of 322 images of the (MIAS) database, 310 are correctly segmented, 3 are under-segmented, 5 are over-segmented, and 4 are not segmented. The algorithm's accuracy is 97.45% and precision 98.41%. For the Digital mammography database (Mini-DDSM), out of 100 images, 95 are correctly segmented, 2 are under-segmented 1 is over-segmented, and 2 are not segmented. The accuracy and precision are 97% and 98.96 % respectively.

D. Related review

Isa et al. [54] developed a system for segmentation to separate the mammary from the non-mammary area. They subsequently enhance the image to improve the contrast of the tissue structure. Constraint-based local statistical texture analysis to detect and segment masses in mammograms. It was evaluated on 322 mammograms from the Mammographic Image Analysis Society database. The results of the proposed technique have a sensitivity of 94.59%.

Elfadilb et al.[55] proposed a preprocessing technique that involves unwanted parts from the background removal of the mammogram, the pectoral muscle, and image enhancement. Based on Otsu's threshold and multi-level thresholding to segment the pectoral muscle. The three thresholding levels show perfect results of pectoral muscle segmentation. The proposed method was applied to 160 images from the (MIAS) database. The success rate is 96% for the mammogram preprocessing stage.

Soukaina et al. [56] proposed breast tumor segmentation and elimination of the pectoral muscle based on hidden Markov and region growth. The method separated the pectoral muscles from mammogram images and breast tumor feature extraction. It has two stages: (a) Otsu's thresholding and (b) image classification-based k-means. The mammography images were acquired from the (MIAS) database. The accuracy and error were reported to be 91.92% and 8.07%, respectively.

Mughal et al.[57] proposed an adaptive hysteresis thresholding method to detect mammogram masses. This method was applied to (MAIS) and (DDSM) datasets and gave sensitivity equal to 96.6%, and 96.4%, respectively.

Our method is simple and effective for identifying shapes, contours and cancerous mass regions. Unlike local thresholding methods, which give poor results, the contrast between objects and background is low. Our pre-processing approach improves the quality of the mammographic image. Adaptive thresholding adjusts a threshold value for each sub-region based on the local threshold, improving mass identification by processing each sub-region independently. The morphological filter eliminates small objects that do not correspond to the regions of interest. The images are classified as normal, benign and malignant based on the geometric criteria of the (ROI). Results are more accurate than the methods described in the literature[54-55-56].

IV. Conclusion

The article presented a simple and effective approach for cancerous mass region segmentation of mammographic images. Mammography(MM) is a commonly used imaging modality and is now an indispensable tool for any clinical examination relating to the breast. The developed algorithm is based on adaptive thresholding and morphological filtering. We tested our method on the Mammographic Image Analysis Society (MIAS) and Digital Mammograph (Mini-DDSM) Databases. The algorithm successfully segmented the regions of interest (ROI). Based on the geometric characteristics (contour, shape and surface), the mammographic images are classified into normal, malignant and benign. Results are promising for improving the early detection of breast cancer from mammography images.

References

[1] CAO, Bochen, BRAY, Freddie, ILBAWI, André, et al. Effect on longevity of one-third reduction in premature mortality from non-communicable diseases by 2030: a global analysis of the Sustainable Development Goal health target. The Lancet Global Health, 2018, vol. 6, no 12, p. e1288-e1296.

[2] SHEIKHPOUR, Elnaz, NOORBAKHSH, Parisa, FOROUGHI, Elnaz, et al. A survey on the role of interleukin-10 in breast cancer: A narrative. Reports of biochemistry & molecular biology, 2018, vol. 7, no 1, p. 30.

[3] BOUNEDJAR, Adda, MELZI, Mohamed Aimene, IDIR, Hassina, et al. General oncology care in Algeria. In: Cancer in the Arab World. Singapore: Springer Singapore, 2022. p. 15-30.

[4] AKRAM, Muhammad, IQBAL, Mehwish, DANIYAL, Muhammad, et al. Awareness and current knowledge of breast cancer. Biological research, 2017, vol. 50, p. 1-23.

[5] GUPTA, Madhur et GOYAL, Neeru. Applied anatomy of breast cancer. In: Breast Cancer: Comprehensive Management. Singapore: Springer Nature Singapore, 2022. p. 23-35.

[6] YADAV, Akshay R. et MOHITE, Shrinivas K. Cancer-A silent killer: An overview. Asian Journal of Pharmaceutical Research, 2020, vol. 10, no 3, p. 213-216.

[7] SHAIKH, Khalid, KRISHNAN, Sabitha, THANKI, Rohit, et al. An introduction to breast cancer. Artificial intelligence in breast cancer early detection and diagnosis, 2021, p. 1-20.

[8] MCCART REED, Amy E., KALINOWSKI, Lauren, SIMPSON, Peter T., et al. Invasive lobular carcinoma of the breast: the increasing importance of this special subtype. Breast Cancer Research, 2021, vol. 23, p. 1-16.

[9] BALALI, Gadafi Iddrisu, YAR, Denis Dekugmen, DELA, Vera Gobe Afua, et al. Breast cancer: a review of mammography and clinical breast examination for early detection of cancer. Open Access Library Journal, 2020, vol. 7, no 10, p. 1-19.

[10] THAHIR, M. P. V. R. A Clinicopathological Study with Mammographic Findings and Management of Benign Breast Diseases. 2019. Thèse de maîtrise. Rajiv Gandhi University of Health Sciences (India).

[11] OBEAGU, Emmanuel Ifeanyi et OBEAGU, Getrude Uzoma. Breast cancer: A review of risk factors and diagnosis. Medicine, 2024, vol. 103, no 3, p. e36905.

[12] FAKHRI, Nada, CHAD, Mohamed Amine, LAHKIM, Mohamed, et al. Risk factors for breast cancer in women: an update review. Medical Oncology, 2022, vol. 39, no 12, p. 197.

[13] WAKS, Adrienne G. et WINER, Eric P. Breast cancer treatment: a review. Jama, 2019, vol. 321, no 3, p. 288-300.

[14] HENNEQUIN, C., BELKACÉMI, Y., BOURGIER, C., et al. Radiotherapy of breast cancer. Cancer/Radiotherapie, 2022, vol. 26, no 1-2, p. 221-230.

[15] TRAYES, Kathryn P. et COKENAKES, Sarah EH. Breast cancer treatment. American Family Physician, 2021, vol. 104, no 2, p. 171-178.

[16] HE, Ziyu, CHEN, Zhu, TAN, Miduo, et al. A review on methods for diagnosis of breast cancer cells and tissues. Cell proliferation, 2020, vol. 53, no 7, p. e12822.

[17]HE, Ziyu, CHEN, Zhu, TAN, Miduo, et al. A review on methods for diagnosis of breast cancer cells and tissues. Cell proliferation, 2020, vol. 53, no 7, p. e12822. [18] EISEN, Andrea, FLETCHER, Glenn G., FIENBERG, Samantha, et al. Breast magnetic resonance imaging for preoperative evaluation of breast cancer: a systematic review and meta-analysis. Canadian Association of Radiologists Journal, 2024, vol. 75, no 1, p. 118-135.

[19] GUO, Rongrong, LU, Guolan, QIN, Binjie, et al. Ultrasound imaging technologies for breast cancer detection and management: a review. Ultrasound in medicine & Biology, 2018, vol. 44, no 1, p. 37-70.

[20]JACOB, Greeshma, JOSE, Iven, et SUJATHA, S. Breast cancer detection: A comparative review on passive and active thermography. Infrared Physics & Technology, 2023, p. 104932.

[21] Emission Tomography (PET)/Computed Tomography (CT) in the Management of Breast Cancer. Medicina, 2024, vol. 60, no 2, p. 321.

[22] JACOB, Greeshma, JOSE, Iven, et SUJATHA, S. Breast cancer detection: A comparative review on passive and active thermography. Infrared Physics & Technology, 2023, p. 104932.

[23] NICOSIA, Luca, GNOCCHI, Giulia, GORINI, Ilaria, et al. History

of mammography: analysis of breast imaging diagnostic achievements over the last century. In: Healthcare. MDPI, 2023. p. 1596.
[24] CHOTAI, Niketa, KULKARNI, Supriya, CHOTAI, Niketa, et al. Mammography. Breast Imaging Essentials: Case Based Review, 2020, p. 9-14.
[25]FIORICA, James V. Breast cancer screening, mammography, and other modalities. Clinical obstetrics and gynecology, 2016, vol. 59, no 4, p. 688-709.
[26] ALMALKI, Yassir Edrees, SOOMRO, Toufique Ahmed, IRFAN, Muhammad, et al. Computerized analysis of mammogram images for early detection of breast cancer. In: Healthcare. MDPI, 2022. p. 801.
[27] BICCHIERAI, Giulia, BUSONI, Simone, TORTOLI, Paolo, et al. Single center evaluation of comparative breast radiation dose of contrast enhanced digital mammography (CEDM), digital mammography (DM) and digital breast tomosynthesis (DBT). Academic Radiology, 2022, vol. 29, no 9, p. 1342-1349.
[28] MOHD NORSUDDIN, Norhashimah et KO, Zhi Xuan. Common Mammographic Positioning Error in Digital Era. In: 3rd International Conference for Innovation in Biomedical Engineering and Life Sciences: Proceedings of ICIBEL 2019, December 6-7, 2019, Kuala Lumpur, Malaysia. Springer International Publishing, 2021. p. 142-149.
[29] CHIKARMANE, Sona A., OFFIT, Lily R., et GIESS, Catherine S. Synthetic mammography: benefits, drawbacks, and pitfalls. Radiographics, 2023, vol. 43, no 10, p. e230018.
[30] VOURTSIS, Athina et BERG, Wendie A. Breast density implications and supplemental screening. European radiology, 2019, vol. 29, p. 1762-1777.
[31] JUSTANIAH, Eman, ALHOTHALI, Areej, et ALDABBAGH, Ghadah. Mammogram segmentation techniques: A review. International Journal of Advanced Computer Science and Applications, 2021, vol. 12, no 5.
[32]JASIM, Wala'a N. et MOHAMMED, Rana Jassim. A Survey on Segmentation Techniques for Image Processing. Iraqi Journal for Electrical & Electronic Engineering, 2021, vol. 17, no 2.
[33]JUSTANIAH, Eman, ALHOTHALI, Areej, et ALDABBAGH, Ghadah. Mammogram segmentation techniques: A review. International Journal of Advanced Computer Science and Applications, 2021, vol. 12, no 5.
[34] HAROUNI, Majid, KARIMI, Mohsen, et RAFIEIPOUR, Shadi. Precise segmentation techniques in various medical images. Artificial Intelligence and Internet of Things, 2021, p. 117-166.
[35] ABO-EL-REJAL, A., AYMAN, S., et AYMEN, F. Advances in breast cancer segmentation: A comprehensive review. Acadlore Transactions on AI and Machine Learning, 2024, vol. 3, no 2, p. 70-83.
[36] RAJOUB, Bashar, QUSA, Hani, ABDUL-RAHMAN, Hussein, et al. Segmentation of breast tissue structures in mammographic images. Artificial Intelligence and Image Processing in Medical Imaging, 2024, p. 115-146.
[37] MICHAEL, Epimack, MA, He, LI, Hong, et al. Breast cancer segmentation methods: current status and future potentials. BioMed research international, 2021, vol. 2021, no 1, p. 9962109.
[38]JARDIM, Sandra, ANTÓNIO, João, et MORA, Carlos. Graphical image region extraction with k-means clustering and watershed. Journal of Imaging, 2022, vol. 8, no 6, p. 163.
[39]SAKTHIVEL, Usha, JYOTHI, A. P., SUSILA, N., et al. Conspectus of k-Means Clustering Algorithm. In: Applied Learning Algorithms for Intelligent IoT. Auerbach Publications, 2021. p. 193-213.
[40]BALKENENDE, Luuk, TEUWEN, Jonas, et MANN, Ritse M. Application of deep learning in breast cancer imaging. In: Seminars in Nuclear Medicine. WB Saunders, 2022. p. 584-596.
[41]LI, Shuyi, DONG, Min, DU, Guangming, et al. Attention dense-u-net for automatic breast mass segmentation in digital mammogram. Ieee Access, 2019, vol. 7, p. 59037-59047.
[42] ABDELHAFIZ, Dina, BI, Jinbo, AMMAR, Reda, et al. Convolutional neural network for automated mass segmentation in mammography. BMC Bioinformatics, 2020, vol. 21, p. 1-19.
[43]MUHAMMAD, Muhammad, ZEEBAREE, Diyar, BRIFCANI, Adnan Mohsin Abdulazeez, et al. Region of interest segmentation based on clustering techniques for breast cancer ultrasound images: A review. Journal of Applied Science and Technology Trends, 2020, vol. 1, no 2, p. 78-91.
[44]IRANMAKANI, Sepideh, MORTEZAZADEH, Tohid, SAJADIAN, Fakhrossadat, et al. A review of various modalities in breast imaging: technical aspects and clinical outcomes. Egyptian Journal of Radiology and Nuclear Medicine, 2020, vol. 51, p. 1-22.
[45]https://www.mammoimage.org/databases/
[46]http://www.eng.usf.edu/cvprg/mammography/database.html
[47]SAGHEER, Sameera V. Mohd et GEORGE, Sudhish N. A review on medical image denoising algorithms. Biomedical signal processing and control, 2020, vol. 61, p. 102036.
[48] MICHAEL, Epimack, MA, He, LI, Hong, et al. Breast cancer segmentation methods: current status and future potentials. BioMed research international, 2021, vol. 2021, no 1, p. 9962109.
[49]TAHA, Taha Basheer. Adaptive Threshold-Based Tumor Detection Algorithm for Mammograms Images. EURASIAN JOURNAL OF SCIENCE AND ENGINEERING, 2023, vol. 9, no 2, p. 25-34.
[50]PATEL, Bhagwati Charan, SINHA, G. R., et SONI, Dilip. Detection of masses in mammographic breast cancer images using modified histogram based adaptive thresholding (MHAT) method. International Journal of Biomedical Engineering and Technology, 2019, vol. 29, no 2, p. 134-154.
[51] SERRA, Jean. Mathematical morphology. In: Encyclopedia of Mathematical Geosciences. Cham: Springer International Publishing, 2023. p. 820-835.
[52] ACHARJYA, Pinaki Pratim, BARMAN, Subhabrata, JOARDAR, Subhankar, et al. A literature review on mathematical morphology based image segmentation techniques. Recent Progress in Science and Technology, MV Irenee, Ed, 2023, vol. 8, p. 85-97.
[53]SHIH, Frank Y. Image processing and mathematical morphology: fundamentals and applications. CRC Press, 2017.
[54] ISA, Nor Ashidi Mat et SIONG, Ting Shyue. Automatic segmentation and detection of mass in digital mammograms. Recent researches in communications, signals and information technology, 2012, p. 143-146.
[55]OMER, Ashgan M. et ELFADIL, Mohammed. Preprocessing of digital mammogram image based on otsu's threshold. American Scientific Research Journal for Engineering, Technology, and Sciences, 2017, vol. 37, no 1, p. 220-229.
[56]EL IDRISSI EL KAITOUNI, Soukaina, ABBAD, Abdelghafour, et TAIRI, Hamid. A breast tumors segmentation and elimination of pectoral muscle based on hidden Markov and region growing. Multimedia Tools and Applications, 2018, vol. 77, no 23, p. 31347-31362.
[57]]MUGHAL, Bushra, MUHAMMAD, Nazeer, et SHARIF, Muhammad. Adaptive hysteresis thresholding segmentation technique for localizing the breast masses in the curve stitching domain. International Journal of Medical Informatics, 2019, vol. 126, p. 26-34.

Comparative Analysis of Classification Models for Bone Mineral Density

1st HAMIDAT Ibrahim
dept. of Electrical and Electronics
LIST Laboratory
University M'Hamed Bougara of Boumerdes,
35000 Boumerdes, Algeria

2nd HARRAR Khaled
dept. of Electrical and Electronics
LIST Laboratory
University M'Hamed Bougara of Boumerdes,
35000 Boumerdes, Algeria

3rd AOUINANE Bilal
dept. of electronics, LATSI Laboratory
University of Saad Dahlab Blida 1
09000 Blida, Algeria

***Abstract*—This research carries out a comparative study on four different machine learning classifiers, namely, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), and Multi-Layer Perceptron (MLP), for classifying the data of patients into three diagnostic categories related to the Bone Mineral Density (BMD) test required in diagnosing diseases like osteopenia and osteoporosis. The dataset is characterized by features such as age, height, weight, number of pregnancies, and T-score values. The model performance is evaluated using 5-fold cross-validation, and grid search is applied for the optimization of hyperparameters. The results indicate that Logistic Regression outperforms the other models, achieving an accuracy of 93.88%, followed by SVM at 91.84%. This work indicates how important hyperparameter tuning is in improving accuracy and further suggests that LR and kernel-based approaches (SVM) could be good suitors for healthcare classification tasks. Furthermore, performance metrics—accuracy, precision, recall, and F1-score—explain the benefits and drawbacks of each model, where ensemble techniques prove to have the fairest performance across all classes. Future research may try more advanced models, like Gradient Boosting, in order to improve the accuracy of classification and to cope with class imbalance better.**

***Index Terms*—BMD, SVM, LR, KNN, MLP, Osteoporosis, Osteopenia, 5-fold cross-validation.**

I. Introduction

The healthcare field has seen a wide use of machine learning because of its ability to process vast amounts of patient data and uncover patterns that might not be obvious using traditional methods [2]. In particular, classification models are useful for assisting medical professionals in making diagnoses of diseases, predicting patient outcomes, and personalizing treatment plans. With the increasing trend of data-driven decision-making in healthcare, accurate and robust classification models can enhance clinical decision-making, reduce human error, and ultimately improve patient care.

In this paper, we investigate the application of four popular classification algorithms—Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), and Multi-Layer Perceptron (MLP)—to a dataset of patient records containing key medical and demographic features. The dataset includes features such as age, height, weight, number of pregnancies, and bone density measurements (T-score and Z-score), all of which play important roles in assessing patient health status. The classification task involves assigning patients to one of three diagnostic categories, representing different health conditions.

While machine learning classifiers have been widely studied in healthcare applications, there is limited research on how different models perform when tuned for optimal hyperparameters using grid search in conjunction with cross-validation. Cross-validation is crucial in healthcare, as it ensures the generalizability of models and avoids overfitting to a single dataset. Similarly, grid search provides a systematic approach to identifying the best hyperparameters, which can significantly improve model performance.

Our study aims to compare the performances of these four classifiers with an extensive evaluation framework in terms of 5-fold cross-validation and grid search. We assume that the most complex models, SVM and MLP, able to capture nonlinear relationships between the input features, will present the best results compared with the classical models, KNN and Logistic Regression, especially after optimizing their hyperparameters. We also discuss the major trade-offs between model interpretability and its predictive power, another major consideration in healthcare, where decision transparency is paramount.

A growing body of literature has examined the use of machine learning models in health settings. For example, SVMs have found their application in various medical fields, such as cancer detection (Janee Alam et al., 2018) [7], with strong capability in handling high-dimensional data sets. Similarly, KNN, due to the simplicity and good performance for small and medium-sized datasets, has found applications in disease diagnosis and prognosis (X Zhou et all., 2022) [8]. Logistic Regression (LR) remains a widely used technique in healthcare, particularly for binary classification tasks like risk assessment for conditions such as diabetes and heart disease (M Laakso et al., 1998) [9]. Multilayer Perceptrons (MLP),

classified as a neural network, have garnered significant interest due to their capacity to represent intricate, non-linear associations within extensive datasets, thereby rendering them appropriate for sophisticated medical applications, including image analysis and predictive modeling (M Desai et al., 2021) [10].

Bone health is a critical aspect of overall well-being, and conditions like osteopenia and osteoporosis significantly increase the risk of fractures and other complications [2]. Bone Mineral Density (BMD) is a standard diagnosis conducted to assess an individual's bone health and further categorize them into normal, osteopenia, or osteoporosis, all based on their T-score.

Machine learning algorithms have become very popular in medical diagnosis for their ability to handle large datasets and extract patterns that may not be evident through traditional methods. However, correct classification of categories is still faced with many challenges in BMD, including:

- **Class imbalance:** Data for conditions like osteoporosis is often underrepresented compared to normal or osteopenia classes.
- **Feature variability:** Patient data includes diverse features such as age, weight, menopause status, and lifestyle factors, which require effective preprocessing.
- **Model selection and optimization:** Different classification models perform variably depending on the dataset, necessitating a comparative analysis.

The main objective of this study is to perform a comparative analysis of classification models for BMD classification using a real-world dataset. Specifically, we:

- Preprocess and clean patient data for analysis.
- Train and evaluate Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP) models.
- Handle class imbalance using techniques like SMOTE (Synthetic Minority Oversampling Technique) and class weights.
- Explore ensemble methods for improving classification performance.
- Compare model performance using metrics such as accuracy, precision, recall, and F1-score.

The rest of this paper is organized as follows. Section II outlines the methodology, including a detailed description of the dataset, data analysis, male and female BMD distribution, and Gender comparision, Section III presents preprocessing steps, and data splitting, Section IV contains the machine learning models used and evaluation metrics. and Section V presents the results of the classification experiments, followed by a discussion of their implications. Finally, Section V concludes with key findings and suggestions for future work.

II. Methodology

A. Dataset Description

The dataset used in this study consists of patient demographics and medical records. Table I below presents the statistical summary of the features such as:

- Age
- Height (meters)
- Weight (kg)
- Number of Pregnancies
- Maximum Walking Distance (km)
- T-score Value (bone density measure)
- Z-score Value (comparison to age-matched population)
- Body Mass Index (BMI)

The target variable is the diagnostic class, with three categories representing different patient health statuses. The goal is to classify each patient into one of these three categories.

The summary statistics highlight key characteristics of the dataset:

Age shows a wide range, with a maximum of 107 years and a mean of 51.13 years, as shown in Fig. 1. Height and Weight are normally distributed, with reasonable variation. There are missing values in the number of Pregnancies and Maximum Walking Distance, which were handled appropriately in preprocessing. T-score and Z-score values are already

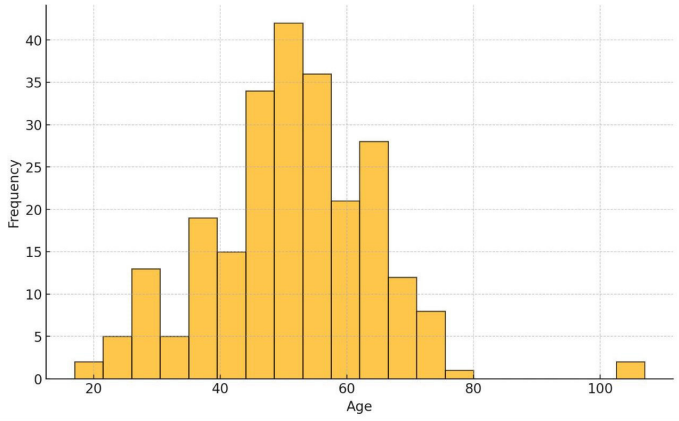


Fig. 1. Age distribution of gender

standardized. BMI ranges from 16.1 (underweight) to 42.75 (obese), with a mean value of 27.59, indicating the dataset covers a broad spectrum of body mass indices. These features are crucial as they represent important health indicators that the machine learning models will use for classification tasks.

B. Data analyzing

The BMD of the participants was examined in relation to their age, and the results were visualized separately for male and female participants, as shown in Fig. 2. The analysis of the BMD distributions reveals significant insights:

C. Female BMD Distribution

- The age of female participants ranges from **30 to 70** years, with the majority concentrated between **40 and 60** years.
- There is a noticeable **decline in BMD with increasing age**, particularly after the age of **50**, which aligns with known physiological changes, such as menopause, that significantly affect bone health in women.

TABLE I
STATISTICAL SUMMARY OF FEATURES

Feature	Count	Mean	Std Dev	Min	25%	50%	75%	Max
Age	243	51.13	13.23	17	44.5	50	60	107
Height (m)	243	1.59	0.097	1.37	1.52	1.57	1.65	1.82
Weight (kg)	243	69.05	9.88	39	63	69	74.5	98
Number of Pregnancies	124	3.44	1.47	1	2	3	4	7
Max Walking Distance (km)	239	1.94	1.99	0.1	0.5	1.0	3.0	10
T-score Value	240	-1.83	0.69	-2.99	-2.32	-2.01	-1.29	-0.16
Z-score Value	240	-1.33	0.68	-2.99	-1.82	-1.19	-0.85	0.73
BMI	240	27.59	4.04	16.14	24.96	27.28	30.21	42.75

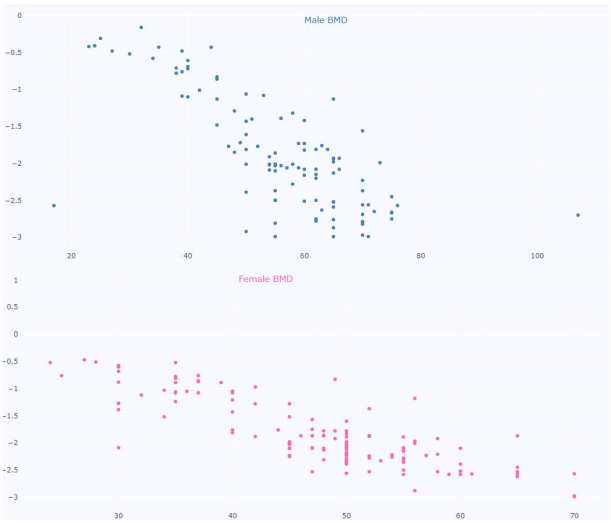


Fig. 2. Gender distribution of BMD

- The BMD values primarily range from **0 to -3**, with lower values (indicative of reduced bone density) concentrated in older participants.

D. Male BMD Distribution

- The age of male participants ranges from **20 to 100** years, offering a broader dataset for analysis.
- Like females, males exhibit a **decline in BMD with age**, though the decline is less steep. This difference may be attributed to the slower rate of bone loss in men compared to women.
- The BMD values are similarly distributed between **0 and -3**, but with a more even spread across age groups. There are notable outliers in the higher age range (above **80 years**).

E. Gender Comparison

- The overall trend of **decreasing BMD with age** is observed in both genders, but the decline is more pronounced in females, particularly after the age of **50**.
- This gender-specific trend suggests that **age and gender** are key factors in predicting BMD and potentially identifying individuals at higher risk of conditions like osteoporosis.

III. DATASET AND PREPROCESSING

The dataset consists of 243 patient records, with each record categorized into one of three BMD diagnoses. The distribution of classes highlights a slight imbalance, with fewer samples in the "osteoporosis" category.

A. Data Cleaning and Handling Missing Values

The dataset contained missing values in both numerical and categorical features. Missing values were addressed as follows:

- For numerical features (e.g., age, weight), missing values were imputed using the median of the respective feature.
- For categorical features (e.g., gender, smoker), missing values were filled with the mode of the respective feature.

B. Data Standardization

To ensure numerical stability and improve model performance, all numerical features were standardized using Z-score normalization:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

Where x is the original feature value, μ is the mean, and σ is the standard deviation of the feature.

C. Class Encoding

Categorical features were encoded into numerical values using Label Encoding. The target variable (diagnosis) was also encoded into three classes:

- Normal (0)
- Osteopenia (1)
- Osteoporosis (2)

D. Train-Test Split

The dataset was split into training (80%) and testing (20%) sets, ensuring that the distribution of target classes was preserved using stratified sampling.

E. Class Imbalance Handling

Class imbalance was addressed using two approaches:

- **SMOTE (Synthetic Minority Oversampling Technique):** This technique generates synthetic samples for minority classes by interpolating between existing samples and their nearest neighbors.
- **Class Weights:** Models like Support Vector Machines (SVM) and Logistic Regression were trained with class weights proportional to the inverse of class frequencies.

IV. Models and Evaluation Metrics

A. Classification Models

The following classification models were implemented and evaluated in this study:

1) Logistic Regression: Logistic Regression models the probability of a binary or multi-class outcome using the sigmoid or softmax function. For binary classification:

$$P(y = 1|x) = \sigma(w^T x + b) = \frac{1}{1 + e^{-(w^T x + b)}} \tag{2}$$

Where w and b are the model parameters, and $\sigma(z)$ is the sigmoid function.

For multi-class classification, the softmax function generalizes this:

$$P(y = k|x) = \frac{e^{w_k^T x + b_k}}{\sum_{j=1}^{K} e^{w_j^T x + b_j}} \tag{3}$$

2) Support Vector Machine (SVM): SVM finds a hyperplane that maximizes the margin between two classes. The optimization problem is:

$$\min \frac{1}{2} w^2 \tag{4}$$

Subject to:

$$y_i(w^T x_i + b) \geq 1, \quad \forall i \tag{5}$$

Where w is the norm of the weight vector.

For non-linear problems, the kernel trick maps the input data into a higher-dimensional space:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \tag{6}$$

3) K-Nearest Neighbors (KNN): KNN predicts the label of a test sample x based on the majority vote of its k nearest neighbors in the feature space:

$$y = \text{argmax}_c \sum_{i=1}^{k} I(y_i = c) \tag{7}$$

Where $I(y_i = c)$ is an indicator function that equals 1 if the neighbor's label is c, otherwise 0.

4) Multilayer Perceptron (MLP): The MLP is a neural network that computes the output as:

$$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)}, \quad a^{(l)} = f(z^{(l)}) \tag{8}$$

Where $W^{(l)}$ and $b^{(l)}$ are the weights and biases of layer l, and $f(z)$ is the activation function. The loss function is minimized using backpropagation with gradient descent.

5) Ensemble Model: An ensemble model combines predictions from multiple models to improve performance. For soft voting:

$$P(y = c|x) = \frac{1}{N} \sum_{i=1}^{N} P_i(y = c|x) \tag{9}$$

Where $P_i(y = c|x)$ is the probability predicted by the i-th model.

B. Evaluation Metrics

The performance of the models was evaluated using the following metrics:

1) Accuracy:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{10}$$

2) Precision:

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \tag{11}$$

3) Recall:

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{12}$$

4) F1-Score: The harmonic mean of precision and recall:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

5) Macro and Weighted Averages:

- **Macro Average:** Arithmetic mean of metrics across all classes.

$$\text{Macro Metric} = \frac{\sum_{i=1}^{n} \text{Metric}_i}{n} \tag{14}$$

- **Weighted Average:** Weighted by the number of instances in each class.

$$\text{Weighted Metric} = \frac{\sum_{i=1}^{n} (\text{Support}_i \cdot \text{Metric}_i)}{\sum_{i=1}^{n} \text{Support}_i} \tag{15}$$

V. Results and Discussion

This section presents the experimental results of the classification models, followed by a comparative analysis of their performance.

A. Model Performance Without Class Imbalance Handling

Initially, the models were trained without addressing class imbalance. Table II summarizes the test set evaluation results.

Logistic Regression demonstrated the highest accuracy of 93.88%, with a balanced macro precision, recall, and F1-score. SVM performed comparably, achieving 91.84% accuracy. KNN struggled with class imbalance, resulting in a lower macro recall of 76%.

B. Impact of Class Imbalance Handling

To address the class imbalance, two techniques were applied: Synthetic Minority Oversampling Technique (SMOTE) and class weighting. The results are shown in Table III.

The results indicate that SMOTE significantly improved macro recall for Logistic Regression (96%), while class weighting provided a balanced performance for SVM.

TABLE II
PERFORMANCE OF CLASSIFICATION MODELS (BASELINE)

Model	Accuracy	Macro Precision	Macro Recall	Macro F1-Score
Logistic Regression	93.88%	91%	93%	92%
Support Vector Machine (SVM)	91.84%	90%	90%	90%
K-Nearest Neighbors (KNN)	85.71%	89%	76%	81%
Multilayer Perceptron (MLP)	87.76%	84%	84%	84%

TABLE III
PERFORMANCE AFTER CLASS IMBALANCE HANDLING

Model	Technique	Accuracy	Macro Precision	Macro Recall
Logistic Regression	SMOTE	91.84%	87%	96%
SVM	Class Weights	91.84%	89%	92%

C. Ensemble Model Performance

The ensemble model, combining Logistic Regression and SVM, achieved the best overall performance:

- **Accuracy:** 93.88%
- **Macro Precision:** 91%
- **Macro Recall:** 95%
- **Macro F1-Score:** 93%

The ensemble model leverages the strengths of Logistic Regression and SVM, particularly excelling in minority class recall while maintaining high precision.

D. Visualizing Model Performance

Fig. 3 compares the performance metrics of the models.

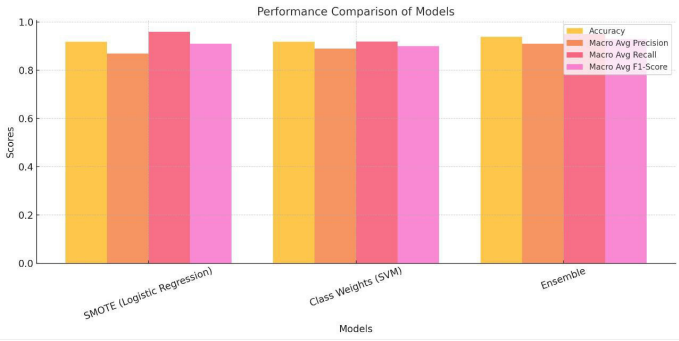


Fig. 3. Performance Comparison of Models Across Metrics

The ensemble model demonstrates superior accuracy and balanced macro averages, making it the best overall approach for BMD classification.

E. Discussion

The results highlight the importance of addressing class imbalance in medical datasets. Logistic Regression and SVM were the most robust individual models [1], with ensemble learning providing further improvements. The SMOTE technique greatly improved the recall of minority classes, while class weighting was a far simpler alternative that gave similar results. KNN and MLP were less competitive, suggesting that instance-based and simple neural network approaches may not be well-suited for this dataset without additional tuning or feature engineering.

VI. CONCLUSION

This study carried out a wide comprehensive comparison of the classification models used to diagnose BMD, including Logistic Regression, SVMs, KNN, and the MLP. The following conclusions were drawn:

- Logistic Regression and SVM emerged as the most reliable individual models, showing high accuracy and balanced performance across metrics such as precision, recall, and F1-score.
- Class imbalance handling has significantly improved the recall of the minority class. Among these, SMOTE was the most effective technique that improved recall, while class weighting was a simpler yet similarly effective technique for SVM.
- The ensemble learning of combining Logistic Regression and SVM had the best overall performance, with an accuracy of 93.88%, and balanced macro-precision, recall, and F1-score. This approach leveraged the strengths of individual models to improve generalization.
- The result from KNN and MLP performance was very poor, which suggests that these approaches may not be well-suited for datasets with diverse features and imbalanced classes without extensive tuning.

This study highlights various directions of further study:

- Explore more sophisticated ensemble techniques, such as stacking and boosting, to further improve performance.
- Explore feature selection and engineering methods to reduce the dimensionality and improve model interpretability.
- Apply the proposed methods to larger and more diverse datasets to validate their robustness across different populations.
- Integrate additional clinical features, such as genetic markers or biochemical parameters, that might considerably enhance their predictive power.

The findings of this study underline the potential of machine learning in medical diagnostics, particularly for conditions like

osteopenia and osteoporosis. By leveraging robust classification models and addressing data imbalances, clinicians can make more accurate and reliable assessments of bone health.

REFERENCES

[1] Cortes, C., and Vapnik, V. "Support-vector networks." *Machine Learning*, vol. 20, no. 3, 1995, pp. 273-297.

[2] Siris, E. S., et al. "Bone mineral density thresholds for pharmacological intervention to prevent fractures." *Nature Reviews Endocrinology*, vol. 6, no. 1, 2010, pp. 1-6.

[3] Esteva, A., et al. "A guide to deep learning in healthcare." *Nature Medicine*, vol. 25, no. 1, 2019, pp. 24-29.

[4] Chawla, N. V., et al. "SMOTE: Synthetic minority over-sampling technique." *Journal of Artificial Intelligence Research*, vol. 16, 2002, pp. 321-357.

[5] Opitz, D., and Maclin, R. "Popular ensemble methods: An empirical study." *Journal of Artificial Intelligence Research*, vol. 11, 1999, pp. 169-198.

[6] Fix, E., and Hodges, J. L. "Discriminatory analysis: Nonparametric discrimination." *The Annals of Mathematical Statistics*, vol. 27, no. 4, 1951, pp. 1163-1171.

[7] Alam, Janee, Sabrina Alam, and Alamgir Hossan. "Multi-stage lung cancer detection and prediction using multi-class svm classifie." 2018 International conference on computer, communication, chemical, material and electronic engineering (IC4ME2). IEEE, 2018.

[8] Zhou, Xiaorong, et al. "Application of kNN and SVM to predict the prognosis of advanced schistosomiasis." Parasitology Research 121.8 (2022): 2457-2460.

[9] Laakso, Markku. "Glycemic control and the risk for coronary heart disease in patients with non-insulin-dependent diabetes mellitus: the Finnish studies." Annals of Internal Medicine 124.1_Part_2 (1996): 127-130.

[10] Desai, Meha, and Manan Shah. "An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN)." Clinical eHealth 4 (2021): 1-11.

Chest CT Scan Super Resolution using Deep Residual Attention Fusion Network

Rania Saoudi
ETA Laboratory,Electronic Department
University of Mohamed El Bachir El Ibrahimi–Bordj Bou Arreridj
Bordj Bou Arreridj, Algeria
rania.saoudi@univ-bba.dz

Abstract— **Acquiring high-resolution (HR) medical images in clinical fields presents significant challenges. These challenges often result in low-resolution (LR) images, which can lead to misinterpretations and hinder accurate diagnoses due to hardware limitations and prolonged scanning times. To address these issues, super-resolution (SR) methods has been increasingly explored the field of medical imaging. This paper introduces the Deep Residual Attention Fusion Network (DRAFN), a novel deep learning approach specifically designed to enhance the quality of chest CT images through advanced SR techniques. Our model incorporates an Advanced Feature Extraction Module that employs hierarchical multi-scale extraction along with sophisticated channel attention mechanisms. This innovative design allows the model to capture features across various scales while prioritizing the most critical information during the feature extraction process. Our model was specifically applied to chest CT imaging, demonstrating substantial improvements over traditional interpolation methods and existing deep learning SR baseline techniques in terms of both image quality metrics and visual quality. These findings underscore the DRAFN potential as a valuable tool for chest CT image SR in clinical evaluations, enhancing diagnostic accuracy and improving patient care.**

Keywords— medical images, CT, super-resolution, deep learning, Chest scans.

I. INTRODUCTION

Computed Tomography (CT) has become a cornerstone in medical imaging, particularly due to its ability to produce detailed cross-sectional images of the body using X-ray technology. Unlike Magnetic Resonance Imaging (MRI), which relies on magnetic fields, CT utilizes X-ray radiation combined with computerized processing to generate intricate images of organs, tissues, and cells. This technique has gained significant traction in both clinical settings and research environments because of its precision and versatility in diagnosing various medical conditions [1],[2].

The process of obtaining HR CT images presents several challenges that can impact patient comfort and diagnostic accuracy [3]. Extended scanning times are often required to achieve HR imaging, which can lead to patient anxiety and discomfort. This discomfort frequently results in excessive movement during the procedure, exacerbated by hardware limitations, physical constraints, and physiological factors [2]. Such movement can significantly degrade image quality, leading to blurriness and artifacts that compromise the diagnostic utility of the images. Consequently, the resolution of CT images acquired using these costly devices may not be adequate for accurate disease diagnosis and progression assessment, as the limited structural information provided can hinder these critical tasks. Additionally, variations in scanning protocols and patient-related factors further complicate the acquisition of high-quality images, underscoring the need for improved techniques to enhance CT imaging outcomes [4].

In order to overcome these hardware, physical and physiological issues and enhance the speed of acquisition and maintain the quality of MRI scans, the field of medical imaging has explored the use of super-resolution (SR) techniques. These techniques aim to reconstruct a HR image from a LR image input, thereby increasing the resolution of MRI scans without the need for additional updates and preserving the original content and details to the greatest extent possible.

In this context, this paper presents an advanced Convolutional neural network, a novel deep leaning method for CT scans SR using Deep Residual Attention Fusion Network (DRAFN) framework. This approach not only improves resolution but also extends its application to high-level tasks in medical imaging. Our approach introduces a hierarchical multi scale feature extraction strategy that effectively addresses the challenges associated with enhancing CT scans resolution.

The DRAFN is distinguished by its unique sequence of Residual Attention Fusion Blocks (RAFBs) and an integrated advnced attention mechanism. A key innovation in the DRAFN architecture is the combination of hierarchical multi scale extracted features within Channel attention mechanism. This structure provides the necessary multi-scale feature extraction capabilities, along side refines the feature representations by highlighting the most relevant information and ensuring that critical details are not ignored during the reconstruction process.

Multiple experiments conducted on Chest CT dataset, have demonstrated that the hierarchical multi-scale and attention nature of our model exhibits an enhanced performance compared to traditional interpolation methods and widely recognized deep learning SR models, excelling in both image clarity metrics and visual quality.

The remainder of this paper is structured as follows: in section II, we define a detailed mathematical SR model. Section III, offers an overview of the advancement of the most cutting-edge SR techniques applied on natural and CT scan fields. In section IV, we present a detailed description of our proposed DRAFN highlighting its innovative features and methodologies. The obtained findings from our research are reported in section V and in section VI, we state our conclusion.

II. SUPER RESOLUTION MODEL

In processing image filed, the Single Image Super Resolution task seeks to recover a corresponding HR image from its LR counterpart. The LR image is typically the output of a degradation process of the original HR image. Mathematically, this process is formulated as [5]:

$$I_{xLR} = d(I_{yHR}, \delta) \quad (1)$$

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Where I_{xLR} represents the LR image, I_{yHR} denotes the input HR image or reference image. The variable d signifies the degradation function that transforms the HR image into its LR counterpart, and δ stands an associated parameter with the degradation process (additive noise). The degradation process is unavailable in a real-world environment, and I_{xLR} is the only the given information input. Serval studies assume that the degradation of images in reality is influenced by various factors including blurring, downsampling and additive noise that should be considered in practical applications of image processing. Below are the mathematical formulations for each component of the degradation process.

The blurring process can be represented as a convolution operation $\otimes$ between the HR image I_{yHR} and bluring kernel *K*. This is mathematically expressed as:

$$I_{xblur} = (I_{yHR} \otimes K) \quad (2)$$

After the blurring step, the image undergoes downsampling step $\downarrow s$ by a scale factor s. This can be represented as:

$$I_{xdown} = I_{xblur} \downarrow s \quad (3)$$

The degradation process also includes additive noise, The degradation process also includes additive noise, which is typically modeled as Gaussian noise n described as follows:

$$I_{xLR} = I_{xdown} + n \quad (4)$$

Combining these components, the overall degradation model can be expressed as [5]:

$$I_{xLR} = D(I_{yHR}, \delta) = (I_{yHR} \otimes K) \downarrow s + n \quad (5)$$

Each component plays a critical role in how images are processed and restored in practical applications.

The SR process is theoretically aimed to provide the inverse solution d^{-1} of the original degradation mapping function to recover the HR image from the LR version. So, the SR process can be described mathematically as follows [5]:

$$g(I_{xLR}, \delta) = d^{-1}(I_{xLR}) \approx I_{ySR} \quad (6)$$

The function g represents the SR function, where d represents the input degradation parameters and I_{ySR} denotes the resulting SR value.

The earlier conventional image SR algorithms, including interpolation techniques [6] enhancing the quality of a LR image by estimating the newly pixel values by interpolating information from neighboring pixels based on multiple range of formultas. While these techniques effective for basic upscaling, they often struggle to preserve fine details and introduce artifacts and demonstrate a limited enhancement in image detail .

To overcome these challenges, deep learning (DL) SR methods have been developed, these models have demonstrated superior performance on a variety of SR benchmarks. These techniques involve the direct acquisition of relationship between an LR image and its corresponding HR version. The deep learning-based SR reconstruction algorithm under consideration asserts its capability to extract a multitude of intricate details from a LR image, thereby surpassing the utility of conventional algorithms[6].

III. Related Works

In this section, we briefly present several deep learning-based SR networks that have been explored for both natural and medical CT scans.

a) Natural images SR methods

In the field of natural image SR task, various deep learning architectures have emerged [7], [8], [9], utilize CNN demonstrating their effectiveness in spatial domain processing, These networks excel at extracting spatial features from images, employing convolutional layers to learn hierarchical representations of image data. Generative Adversarial Networks (GANs) have also gained traction for SISR by generating HR images that are indistinguishable from real images; their adversarial training mechanism significantly enhances the realism of generated outputs as evidenced by several researches [10], [11]. The U-Net approach, originally designed for biomedical image segmentation, has been adapted as a diffusion model in SISR. Its encoder-decoder structure effectively captures both local and global features, making it particularly suitable for enhancing image resolution, as cited in various studies [12], [13]. Recently, several studies [ref], [ref] have employed the Swin Transformer architecture in their SR processes. This architecture has gained attention for its ability to capture various features while maintaining computational efficiency. Utilizing a hierarchical design with shifted windows allows it to process images at multiple scales effectively. This adaptability makes the Swin Transformer a powerful complement to traditional CNNs, GANs, and U-Net models in SR tasks.

b) Medical CT images SR methods

In the medical field, particularly concerning CT images, various deep SR studies have emerged to address the challenges posed by limited LR CT data.

One prominent approach [3] have explored SR methods based on sparse representation and dictionary learning. These approaches utilize pairs of LR and their corresponding HR image patches to train dictionaries that can effectively reconstruct high-quality images from low-dose CT scans. This method is particularly beneficial for clinical applications where minimizing radiation exposure is critical 3. By leveraging sparse coding techniques, these models can enhance image quality while reducing computational costs. Some studies [1] use Enhanced Deep Residual Network (EDSR) [14] which has been shown to outperform other SR algorithms in generating HR CT slices. By fusing different planes of CT studies and applying SR models to reconstruct a third plane, researchers achieved superior image quality, particularly in producing thin-slice images from thicker slices 2. Another effective technique [15] has been developed specifically for coronary CT angiography (CCTA). This method significantly improves image quality by reducing noise and blooming artifacts while enhancing contrast-to-noise ratio (CNR). The results indicate that SR-DLR provides better spatial resolution and object detectability compared to traditional reconstruction techniques, making it a promising option for clinical practice. A novel approach [4] involves the use of U-Net architectures for mapping LR to HR images. This study proposed a SR U-Net that utilized paired LR and HR cone-beam CT (CBCT) images to improve image quality. This method demonstrated significant improvements in

evaluation metrics when compared to traditional interpolation methods, highlighting the effectiveness of deep learning in reconstructing sharper images from lower-quality inputs.

IV. Proposed Works

Unlike the SR-based CNN and GAN methods discussed earlier, which typically extract features at a one scale. In this paper we propose a novel deep learning model specifically designed for Chest CT images SR called Deep Residual Attention Fusion Network (DRAFN). Our approach integrates Hierarchical multi-scale with attention mechanisms, significantly enhancing the feature extraction process from (LR) images.

The workflow of the suggested model is illustrated in Fig.1. the proposed approach can be categorized into three principal elements: (1) Primary Feature Extraction Module, (2) Deep Feature Extraction Modules, and (3) Reconstruction Module.

In the Primary Feature Extraction Module, we employ 3×3 convolutional filters with the Rectified Linear Unit (ReLU) activation function. This initial module is designed to extract 64 shallow features from the input sequence of lower-resolution images. The set of 64 shallow features is then transmitted to the Deep Feature Extraction Module, where more robust features are extracted through a series of processing stages.

The Deep Feature Extraction Module comprises six consecutive blocks of Residual Attention Fusion Blocks (RAFBs), as shown in Fig.2, each RAB consists of four parallel convolutional branches, utilizing different kernel sizes (1x1, 3x3, 5x5, and 7x7) to capture features at multiple scales. To further enhance feature extraction, we integrate a Channel Attention Module (CAM) [16] within each scale, accompanied by convolutional layers. This integration significantly augments the model's ability to focus on essential and informative features while filtering out non-essential ones. Specifically, CAM emphasizes

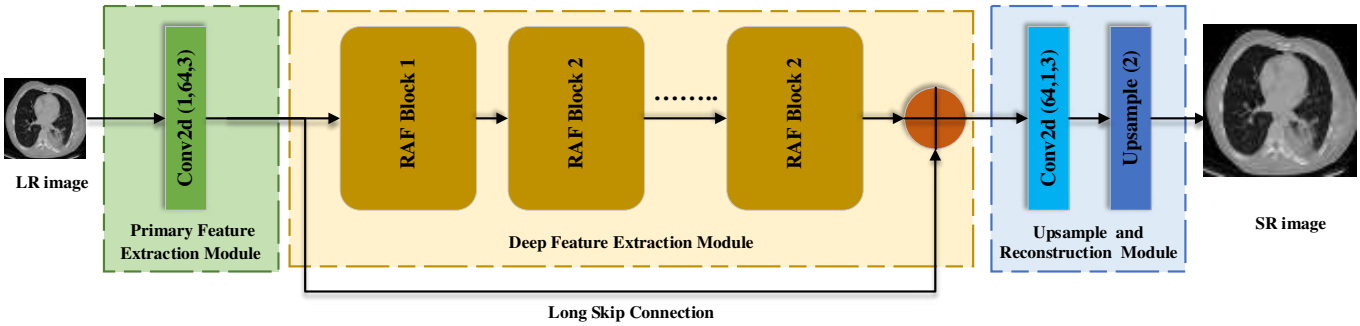


Fig.1. Deep Residual Attention Fusion Network (DRAFN) for Chest CT image SR architecture.

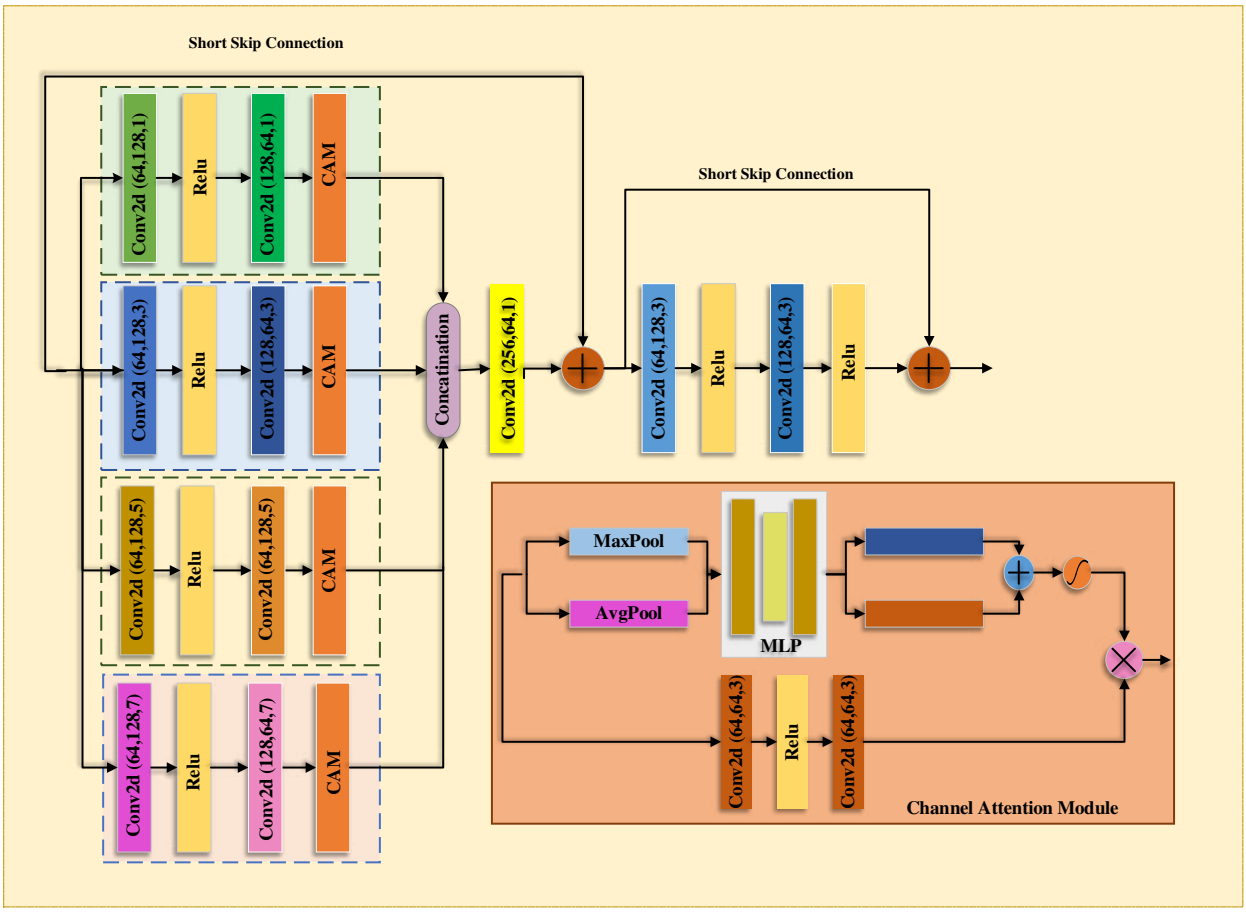


Fig.2. Residual Attention Fusion Blocks (RAFBs) structure.

interdependencies among feature channels, dynamically adjusting the importance of different channels based on their relevance.

The 64 extracted features from the four branches are concatenated together and sent through a convolutional layer with 3x3 kernel size filters to lower the number of dimensions from 256 concatenated features to 64 features. We apply residual connections between the original input and the processed output. Additionally, we employ residual blocks as a final component to further combat overfitting and support deeper network training.

To address potential issues related to network degradation, we implement short skip connections that merge both inputs and outputs of the residual layers with shallow feature fusion, enriching information flow. Long skip connections are also utilized to integrate local depth features, enhancing the network's ability to capture fine details critical for accurate image reconstruction.

In the Reconstruction Module, the 64 resultant features from the Deep Feature Modules are fed into a convolution layer with a kernel size of 3, producing an output size of 1 (as we specifically address grayscale images). The feature map is subsequently upscaled by a scale factor of 2 using an upsampling layer, to produce high-quality super-resolved images.

In our proposed model, the process of SR reconstruction is faster and easier thanks to the exclusion of batch normalization results in reduced complexity and a 40% decrease in memory space occupation as stated in reference [14]

V. EXPERIMENTAL RESULTS

In this section, we provide a detailed overview of the implementation specifics, followed by an analysis of the datasets and evaluation metrics utilized in our study. then, we demonstrating our proposed model outperforms existing deep SR methods through both quantitative and qualitative comparison analyses.

A. Implementation details

The proposed methodology was implemented using PyTorch version 1.6, with the training and testing processes conducted on a Google Colab, utilizing a Tesla V4 GPU equipped with 16 GB of RAM. The neural network underwent training for a total of 30 epochs, employing a batch size of 16.

We utilized the mean square error (MSE) loss, which is a widely recognized metric for assessing the performance of regression models. In the optimization process, we use Adam optimizer with a learning rate of 0.001 and we set their parameters momentum term $\beta 1$ and RMSprop term $\beta 2$ empirically to 0.9 and 0.999 respectively. These settings were specifically chosen to enhance both convergence speed and stability of our model during the training phase.

B. Datasets Details

The proposed model as well as the considered SR algorithms was applied to Chest CT dataset utilized for the purpose of training and evaluating various models of Chest CT scans SR task , this datasets is publicly available at [17]. The dataset comprises 1,000 images divided into four classes: one normal class and three abnormal classes, with 80% allocated for training and 20% reserved for testing. To mitigate the challenges posed by the limited dataset size, we implemented a data augmentation technique that enhances the diversity of the training dataset. This strategy incorporates several techniques, including random rotation, vertical and horizontal flipping, and cropping of the most informative sections of the images.

The HR images have a size of 256 × 256 pixels with a single channel and the LR images were generated from the original HR images using bicubic downsampling with a scale factor of 2.

C. Image Evaluation metrics

We use Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) metrics, in addition to visual quality to evaluate the efficiency of all models as used in the deepest SR outcomes designed for Chest CT.

D. Results

To evaluate the effectiveness of our proposed network (DRAFN), which integrates residual attention and a multi-scale feature extraction architecture, we conducted both quantitative comparative analyses and visual comparisons against classical bicubic interpolation and three advanced deep learning SR algorithms including VDSR[18], MDSR[14], and EDSR[14]. To ensure a fair comparison, all models were trained using the same dataset and for an identical number of epochs. The evaluation was conducted for an upscaling factor of 2, with optimal outcomes highlighted in bold. Additionally, the parameters for each network are provided in the same table.

The funding results clearly demonstrate that our network achieves superior PSNR and SSIM scores compared to the other methods for Chest CT SR task. Specifically, our network achieved an impressive PSNR of 43.91 dB and an SSIM of 0.9988 compared to the next best approach, EDSR, recorded a PSNR of 42.24 dB and an SSIM of 0.9994. The MDSR algorithm achieved a PSNR of 42.51 dB with an SSIM of 0.9973, while VDSR yielded a PSNR of 41.80 dB and an SSIM of 0.9961. In contrast, classical bicubic interpolation resulted in a significantly

TABLE I. PSNR & SSIM AVERAGE VALUES OF OUR PROPOSED MODEL VS BICUBIC INTERPOLATION, VDSR, MDSR, EDSR AND APPLIED ON CHEST CT DATASET.

SR Methods	*Scale Factor*	*Parameters*	*PSNR*↑	*SSIM*↑
Bicubic Interpolation	x2	/	34.50	0.9095
VDSR	x2	0.7 M	41.80	0.9961
MDSR	x2	1.7 M	42.51	0.9973
EDSR	x2	1.2 M	43.12	0.9984
Proposed	x2	2.9 M	**43.91**	**0.9988**

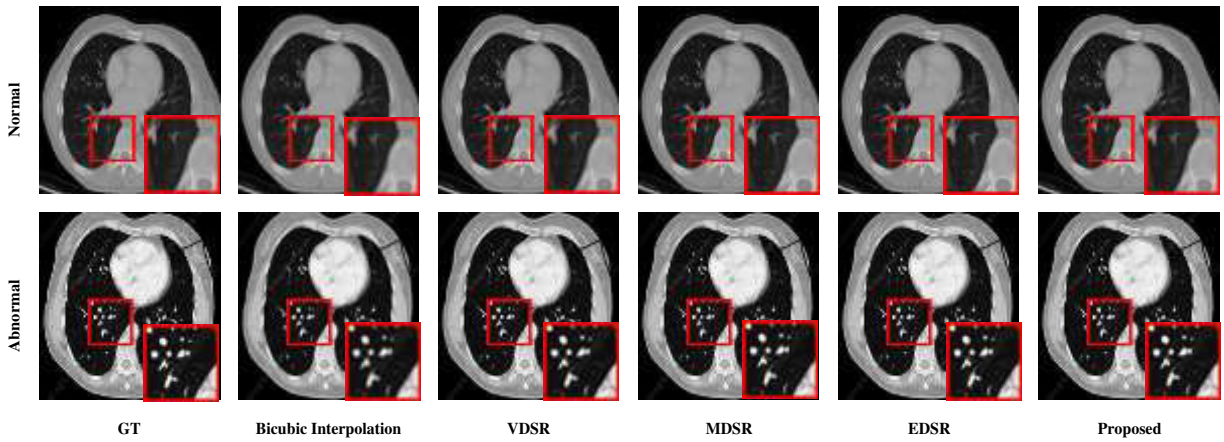


Fig.3. Visual comparison of different SISR models on Chest CT benchmark datasets with 2× upscaling factor.

lower PSNR of 34.50 dB and an SSIM of 0.9195.

Based on the statistical analyses conducted, it is evident that our proposed model significantly outperforms the other SR networks, demonstrating its effectiveness in enhancing image quality.

Fig.3 presents a visual comparison of the mentioned methods in the Table .1 applied to Chest CT images, utilizing an upscaling factor of 2×. We showcase the ground truth (GT) images alongside the results generated by bicubic interpolation, as well as the performances of VDSR, MDSR, EDSR, and our proposed model.

To illustrate the visual superiority of our proposed model, we provide an enlargement of the region highlighted by the red rectangle, which is discernible in both normal and abnormal chest CT images.

In the normal case of images, the analysis of the visual results highlights that bicubic interpolation results indicate that the reconstruction of this particular area was incomplete. Conversely, VDSR and MDSR results have shown some improvement but they are unable to achieve a completely accurate detailed texture representation. EDSR results show some improvement in this regard. However, our model offers accurate representation of the ground truth, with well-defined edges.

Similar findings can also be seen in abnormal case, demonstrating its effectiveness in enhancing image quality, making it a valuable contribution to the field of medical image processing practically in telemedicine filed by delivering clearer images for off-site specialists, ultimately improving patient outcomes. These practical applications underscore the proposed method's significance in advancing medical imaging and enhancing patient care.

VI. Conclusion

In this paper, we introduced the Deep Residual Attention Fusion Network (DRAFN), an advanced CNN designed to improve the resolution of CT images using SR techniques. By integrating residual multi-scale feature extraction with attention mechanisms, our model effectively captures critical features from LR images, leading to significant improvements in CT scan resolution. We evaluated the DRAFN's performance using publicly available chest CT datasets and conducted extensive experiments comparing it to three SR techniques. The results demonstrate the superiority of our model, as indicated by improved PSNR and SSIM metrics. These findings highlight the effectiveness of the DRAFN for SR tasks involving chest CT scans, underscoring its potential to enhance diagnostic accuracy and patient care.

References

[1] H.-S. Chao, Y.-H. Wu, L. Siana, and Y.-M. Chen, "Generating High-Resolution CT Slices from Two Image Series Using Deep-Learning-Based Resolution Enhancement Methods," *Diagnostics (Basel)*, vol. 12, no. 11, p. 2725, Nov. 2022, doi: 10.3390/diagnostics12112725.

[2] Z. Zhang, S. Yu, W. Qin, X. Liang, Y. Xie, and G. Cao, "Self-supervised CT super-resolution with hybrid model," *Computers in Biology and Medicine*, vol. 138, p. 104775, Nov. 2021, doi: 10.1016/j.compbiomed.2021.104775.

[3] C. Jiang, Q. Zhang, R. Fan, and Z. Hu, "Super-resolution CT Image Reconstruction Based on Dictionary Learning and Sparse Representation," *Sci Rep*, vol. 8, no. 1, p. 8799, Jun. 2018, doi: 10.1038/s41598-018-27261-z.

[4] W. Y. R. Fok, A. Fieselmann, M. Herbst, L. Ritschl, S. Kappler, and S. Saalfeld, "Deep learning in computed tomography super resolution using multi-modality data training," *Medical Physics*, vol. 51, no. 4, pp. 2846–2860, 2024, doi: 10.1002/mp.16825.

[5] S. M. A. Bashir, Y. Wang, M. Khan, and Y. Niu, "A comprehensive review of deep learning-based single image super-resolution," *PeerJ Computer Science*, vol. 7, p. e621, Jul. 2021, doi: 10.7717/peerj-cs.621.

[6] F. Zhu, "A Review of Deep Learning Based Image Super-resolution Techniques," Jan. 21, 2022, *arXiv*: arXiv:2201.10521. Accessed: Aug. 15, 2023. [Online]. Available: http://arxiv.org/abs/2201.10521

[7] C. Tian, Y. Yuan, S. Zhang, C.-W. Lin, W. Zuo, and D. Zhang, "Image Super-resolution with An Enhanced Group Convolutional Neural Network," Jul. 31, 2022, *arXiv*: arXiv:2205.14548. Accessed: Oct. 23, 2024. [Online]. Available: http://arxiv.org/abs/2205.14548

[8] H. Yan, Z. Wang, Z. Xu, Z. Wang, Z. Wu, and R. Lyu, "Research on Image Super-Resolution Reconstruction Mechanism based on Convolutional Neural Network," in *International Conference on Artificial Intelligence, Automation and High Performance Computing*, Zhuhai China: ACM, Jul. 2024, pp. 142–146. doi: 10.1145/3690931.3690956.

[9] R. Sharma, B. Deka, V. Fusco, and O. Yurduseven, "Integrated convolutional neural networks for joint super-resolution and classification of radar images," *Pattern Recognition*, vol. 150, p. 110351, Jun. 2024, doi: 10.1016/j.patcog.2024.110351.

[10] W. Chen, Y. Ma, X. Liu, and Y. Yuan, "Hierarchical Generative Adversarial Networks for Single Image Super-Resolution," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA: IEEE, Jan. 2021, pp. 355–364. doi: 10.1109/WACV48630.2021.00040.

[11] S. AlTakrouri, N. M. Noor, N. Ahmad, T. Justinia, and S. Usman, "Image Super-Resolution using Generative Adversarial Networks

with EfficientNetV2," *IJACSA*, vol. 14, no. 2, 2023, doi: 10.14569/IJACSA.2023.01402100.
[12] B. B. Moser, A. S. Shanbhag, F. Raue, S. Frolov, and S. Palacio, "Diffusion Models, Image Super-Resolution And Everything: A Survey".
[13] Z. Chen *et al.*, "Binarized Diffusion Model for Image Super-Resolution".
[14] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA: IEEE, Jul. 2017, pp. 1132–1140. doi: 10.1109/CVPRW.2017.151.
[15] Y. Nagayama *et al.*, "Improving image quality with super-resolution deep-learning-based reconstruction in coronary CT angiography," *Eur Radiol*, vol. 33, no. 12, pp. 8488–8500, Dec. 2023, doi: 10.1007/s00330-023-09888-3.
[16] W. Wu, G. Lv, Y. Duan, P. Liang, Y. Zhang, and Y. Xia, "DCANet: Dual Convolutional Neural Network with Attention for Image Blind Denoising," Jun. 17, 2023, *arXiv*: arXiv:2304.01498. Accessed: Oct. 25, 2024. [Online]. Available: http://arxiv.org/abs/2304.01498
[17] "Chest CT-Scan images Dataset." Accessed: Nov. 29, 2024. [Online]. Available: https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images
[18] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," Nov. 11, 2016, *arXiv*: arXiv:1511.04587. Accessed: Jul. 18, 2023. [Online]. Available: http://arxiv.org/abs/1511.04587

Phonocardiogram Segmentation Based on HMM Modelling combined with LWE: Application for Heart Valve Disorder Classification

Rima Touahria
LMSE Laboratory
University of Bordj Bou Arreridj, Bordj BouArréridj, Algeria
touahria.rimaa@gmail.com

AbdenourHacine-Gharbi
LMSE Laboratory
University of Bordj Bou Arreridj, Bordj Bou Arréridj, Algeria
gharbi07@yahoo.fr

Philippe Ravier
PRISME Laboratory
University of Orleans, Orléans, France
philippe.ravier@univ-orleans.fr

Karim Abed-Meraim
PRISME Laboratory
University of Orleans, Orléans, France
karim.abed-meraim@univ-orleans.fr

Fateh Ghazali
LMSE Laboratory
University of Bordj Bou Arreridj, Bordj Bou Arréridj, Algeria
laboauto@hotmail.fr

***Abstract*— This study explores the use of Phonocardiogram (PCG) signals for diagnosing cardiovascular disorders and introduces a robust system for automated PCG signal segmentation and classification. It proposes a new framework for segmenting and classifying the PCG signal by dividing it into regions corresponding to S1, systole, S2, and diastole, and classifying these regions as either normal (healthy) or abnormal classes. To achieve this, the system employs Hidden Markov Models (HMM) combined with Gaussian Mixture Models (GMM) for modeling the four regions classes corresponding to class normal of PCG signal and the four regions corresponding to class abnormal. Feature extraction is conducted using Discrete Wavelet Transform (DWT). Performance evaluation using the PASCAL Classifying Heart Sounds Challenge database highlights the system's efficiency. The results showed that four-state Hidden Markov Models (HMM), combined with two-component Gaussian Mixture Models (*GMM*) and the Logarithmic Wavelet Energy (*LWE*) descriptor using a *Db2* mother wavelet with a decomposition level of 7, achieved the best performance.** of 93.68%.

KEYWORDS— ***Segmentation, PCG signal, features extraction, gaussian mixture model, hidden markov model,* Logarithmic Wavelet Energy (*LWE), classification, Toolkit HTK***

I. INTRODUCTION

Cardiac activity has long been the subject of a number of studies, including cardiologists first test of their patients, and heartbeat auscultation through an ordinary stethoscope. The beating sequence can be monitored using electrocardiogram (ECG), phonocardiogram (PCG), which monitor the electrical, audible and blood flow characteristics of heart beat respectively [1].

Among the different modalities mentioned above, PCG is the oldest method for the analysis of the heart sound signal because of its favorable features that include easy to cost-effectiveness and operate [1].

Heart sounds result from certain processes occurring in the heart muscle, such as closing of a valve or tension in the chordae tendons. The main normal heart sounds are the first heart sound (S1), which is produced by the vibrations resulting from the closing of the tricuspid valve and the mitral valve. It corresponds to the beginning of ventricular contraction and the end of diastole and precedes the rise of the carotid artery pulse. The second heart sound (S2) is the sound of the semilunar valves closing in the end of the contraction phase of the right and left ventricles. The third heart sound (S3) can sometimes be normal, but it may also be pathological, while the fourth heart sound (S4) is always considered pathological. Heart sounds can be described according to their intensity, pitch, location, type, and timing in the heart cycle.

Automated heart sound analysis begins with segmenting first and second heart sounds to identify systolic and diastolic regions, enabling reliable classification of pathologies. This process, focused on phonocardiogram (PCG) signals, has been extensively studied over the past decades.

The segmentation and classification of biomedical signal has been studied in [37], in which the authors have proposed the segmentation and the classification of the surface EMG signal for Parkinson's Disease Based on HMM Modeling.

In the present work, we propose to apply the Hidden Markov Models HMM to model separately the S1, systole, S2 and diastole segments of the Normal class PCG signal from those of the Abnormal class during the training phase. This discrimination of normal and abnormal region class models facilitates the decision on the class of the PCG signal during the testing and diagnostic phase.

The remainder of this paper is organized as follows. In section 2, we present the related work. We introduce the proposed approach and details each algorithm step, in the section 3. Section 4, demonstrates the experimental results and discussion. Section 5 highlights conclusions and perspectives concerning future work.

II. RELATED WORK

Most studies on heartbeat sounds classification focus on the pattern recognition approach, which involves three main steps: first, segmentation of PCG signals; second, extracting distinctive features for the PCG classification task; and third, building a mathematical model to classify healthy and pathological heartbeat sounds.

The identification of positions of the fundamental heart sounds (FHSs) is a first step in the automatic classification of heart sound recordings. A lot of different approaches to heart sound segmentation have been presented in the past using techniques such as: Expert Frequency-Energy Based Metric [2], high frequency signatures [3], multi-level wavelet coefficient features [4], Deep Recurrent Neural Networks [5], mel-frequency cepstral coefficients (MFCCs) and deep neural network (DNN) [6]. In addition to the above studies, many other researchers have also contributed to automated segmentation of heart sounds based on the envelope from the PCG. For example, Hilbert transform (HT) [7], [8]using S-transform, [9] using Homomorphic filtering and K-means clustering, [10] empirical mode decomposition (EMD) Wavelet transformation (WT) algorithm with Shannon energy (SE), [11] using Short-term log energy, [12] using sound energy, [13] using lung sounds entropy. Other approaches include algorithms based on statistical models such as hidden Markov models. In research, [14] [15] (DDHMM) [16], [17] method addressed the segmentation of noisy heart sounds using a HSMM and wavelet transform (WT), extended with the use of logistic regression for emission probability estimation. The present findings confirm that all those studies discussed before in matter of the segmentation of PCG signal, show clearly the predominance of envelop method and the probabilistic models. The second step involves feature extraction, wherein each signal is converted into a feature vector or sequence of feature vectors extracted from segmented regions. These features serve as input for the classifier. Wavelet transform is the widespread and the efficient one that applied in signal processing to obtain the information which are applied in many applications. Some studies in this field [18], In [19] [20] [21] [22] [23].

The final step is classification, aiming to choose a suitable classifier capable of making correct decisions based on the extracted features. In heartbeat sounds classification, commonly used methods include neural networks (NN) [24]. Milani et al. [25]applied deep learning techniques for heart sound classification tasks; however, this remains challenging due to the absence of a large authoritative open heart sound dataset. Li et al. [26]introduce a novel method for heart sound classification, utilizing enhanced mel-frequency cepstrum coefficient features and deep residual learning. Saracoglu et al. [27] employed an unconventional approach by fitting a Hidden Markov Model (HMM) to the frequency spectrum extracted from entire heart cycles. Hidden Markov Models (HMM) have been utilized for pathology classification in Phonocardiogram (PCG) recordings, as discussed in reference [28]. Wang et al. [29] utilizing Hidden Markov Models (HMM), the classification of Phonocardiogram (PCG) recordings has been carried out.

This work makes several key contributions to the classification of Phonocardiogram (PCG) signals. First, it exploits the advantage of wavelet decomposition, which is particularly well-suited for extracting impulsive features from the different regions of PCG signals, namely S1, systole, S2, and diastole. Second, it adapts Hidden Markov Models (HMM) for the automatic segmentation and classification of PCG signals. The idea was inspired by the work [37]. HMM is recognized as an effective tool for modeling state transitions in signals, making it ideal for supervised cardiovascular classification tasks [33]. Third, the proposed approach is rigorously evaluated using the PASCAL Classifying Heart Sounds Challenge database, achieving high performance based on standard evaluation metrics. Finally, the methodology incorporates a structured process with distinct learning and testing phases: the learning phase focuses on modeling the four PCG regions, while the testing phase evaluates the classification system's performance, using extracted discriminative parameters to differentiate between the two primary classes.

III. HIDDEN MARKOV-BASED SEGMENTATION AND CLASSIFICATION SYSTEM

A. Database

In order to test our methods, we used the PASCAL Classifying Heart Sounds Challenge database [31]. Two database were provided for this method. Dataset A, it was obtained from the general public via the iStethoscope Pro iPhone app. Dataset B, it was obtained from a clinical trial in hospitals using the digital stethoscope DigiScope. To

evaluate this work, only 621 cardiac cycles "beat" including 204 pathological cardiac cycles were used (see table I). then, each cycle was resampled to 16kHz. This segmentation and registration are performed using the PRAAT software [35]. All of these files are divided into two sub-databases. The first is used for the learning phase (70% heartbeat sound signal) while the second is used for the testing phase (30% heartbeat sound signal). In addition, we assigned to each sound file, a labeling file containing a transcription on the class of the heart sound. Each of these files takes the same name of the sound file with the extension 'lab'. These transcript files are used in the class modeling step and the system evaluation step.

TABLE I. DISTRIBUTION OF DATA IN TRAINING AND TESTING DATASETS.

Classes	Normal	Abnormal
Number	417	204
Test/Train	121/296	58/146

B. *Proposed System*

The hidden Markov Models have been used in several speech recognition systems implemented using the HTK toolbox [30] which was designed to segment the speech signal into several levels such as phoneme, word or sentence using respectively HMM model, lexical model and grammar model. Hence, the speech signal can be transcribed into sequence of phonemes using their trained HMM models, or into sequence of words using their phonemes decomposition described in dictionary, or into sequence of sentences using grammar models that describe the constraints of the acceptable sentences.

In this work, we use the HTK toolbox for segmenting the PCG signal on their components S1, Systole, S2 and Diastole. where each one is modeled with HMM models used to classify them on classes normal and abnormal.

This system consists of a learning and a testing phase. Each phase requires an acoustic analysis step to extract relevant features for this task. So, split the dataset into two sets with 70 percent for training and 30 percent for testing [31]. The following figure show the block diagram of an automatic classification system.

The training phase consists to modelling each class by an HMM of $\mathbf{N_{States}}$,; Each state is modeled by a GMM with $\mathbf{N_{Gaussians}}$ .The reestimation of the features of the HMM models are implemented by an algorithm named the Baum-Welch algorithm. The application of this algorithm is carried out using the HErest command of the HTK box [32].

In the segmentation and classification phase, the HVITE command from the HTK tool uses the trained HMM models along with the constraint model (language model) to transcribe each input sequence of feature vectors into a sequence of classes (S1n, Sysn, S2n, and Diasn) or a sequence of classes (S1an, Sysan, S2an, and Diasan). It also detects the boundaries of these segments [32]. Following this, a classification decision is made for the PCG signal by evaluating the class of the obtained sequence, which is classified as either AN (Abnormal) or N (Normal).

Generally MFCC coefficients are widely used for this task. But MFCC representation requires a large vector of dimensional features. More specifically, we propose to apply an acoustic analysis by wavelets [33].

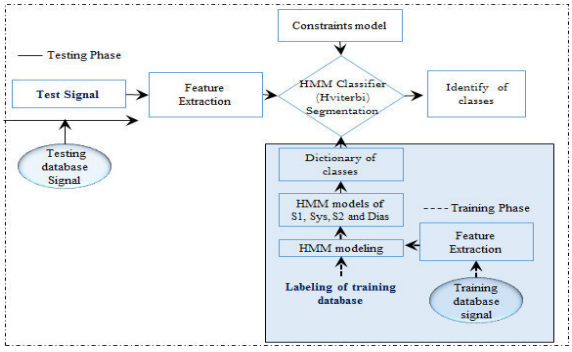


Fig. 1. An automatic segmentation classification system of PCG signal based on HMM models.

C. *WCC feature extraction*

Our method is based on extracting a type of coefficient called wavelet cepstral coefficient (WCC) with low dimensionality of the features vector. Figure 2 shows the flowchart of DWE, LWE and WCC extraction method [33],[36].

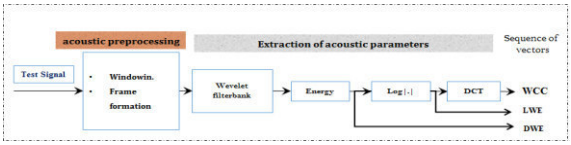


Fig.2. Principle of calculation of the WCCs coefficients, DWEs and LWEs features extraction with hamming windowing.

D. *Performance evaluation*

To assess the performance of the classification task, criterion is taken into account: accuracy ($\boldsymbol{Acc_{HTK}}$). The Acc is defined in Equation (1) and is utilized to assess the detection of PCG signal normal or abnormal by means of the *HRESULTS* command in the HTK tool [32].

- ***Accuracy***

$$Acc_{HTK} = \frac{N-D-S-I}{N} \quad (1)$$

where N denotes the total count of segment labels in the reference transcriptions of PCG signals, D represents the number of labels removed, S indicates the number of labels substituted, and I refer to the number of labels inserted.

IV. Experimental results

A. Experiments

The HMM-based approach, as proposed, was employed to automatically segment and classify PCG signals. Subsequently, this section details the performance evaluation tools used to objectively compare the segmentation and classification results. Following that, an experimental setup is executed to determine the optimal configuration for the segmentation and classification system. Classification results are included, and these results are compared with those from state-of-the-art methods documented in reference [33].

This expertise was carried out as part of the PASCAL Classifying Heart Sounds Challenge database, as referenced in [31].

B. Results and Discussion

- *Performance comparison of the descriptors*

In this section, we present the evaluation results of the LWE (Log Wavelet decomposition-based Energy) descriptor for classifying PCG signals. We conduct a performance analysis and compare it to the DWE descriptor and the WCC descriptor. Throughout this experiment, our objective is to identify the optimal parameter configuration that maximizes the accuracy ($\boldsymbol{Acc_{HTK}}$).

this section, we describe a series of experiments conducted to identify the optimal framework configuration that delivers the best performance. These experiments aim to pinpoint the optimal parameters for the HMM, including the number of states and the number of Gaussian components.

Four experiments are conducted sequentially:

- to assess and compare the performance of the DWE, LWE, and WCC descriptors;
- to determine the ideal duration for the window size;
- to identify the optimal combination of mother wavelet and decomposition level;
- to analyze the performance outcomes of our framework and make comparisons with results obtained from other existing state-of-the-art methods.

To comprehensively assess the performance of our method, we conducted the first experiment, varying the number of HMM states ($\boldsymbol{N_{States}}$), the number of Gaussians for GMM modeling ($\boldsymbol{N}_{\text{Gaussian}}$), and different levels of wavelet decomposition ($\boldsymbol{L}_{\text{decomp}}$). We selected the 'db2' mother wavelet with a 20ms window size, as suggested in [33] for PCG pattern recognition. In the second experiment, we employed the best descriptors with their optimal parameters from the first experiment to study the effect of window duration. Subsequently, we examined the optimal wavelet configuration. Finally, we conducted a comparative analysis with state-of-the-art methods.

- *Optimal configuration of the system*

This section is dedicated to outlining the experimental setup used to determine the most effective configuration for our system. It involves assessing the performance of the proposed HMM method on two signal types: PCG signals from healthy subjects and PCG signals from abnormal subjects.

For each descriptor, we systematically varied the number of states in each experiment, using $\boldsymbol{N_{States}}$= (2, 3, 4, 5, 6, 7, 8, 9, 10,11,12), as well as the number of components in GMM modeling with $\boldsymbol{N}_{\text{Gaussian}}$= (1, 2, 3, 6, 12, 24, 48). The window size is set to a fixed 20 ms, and the mother wavelet selected is 'db2' with a decomposition level of $\boldsymbol{L}_{\text{decomp}}$ =7 [34]. The results obtained are provided in Table II, which illustrates the optimal configurations regarding the number of Gaussians ($\boldsymbol{N}_{\text{Gaussian}}$) and the number of states (N_{States}) for each wavelet descriptor, including DWE, LWE, and WCC.

The results presented in Table II asserts better performance of wavelet descriptors, which achieve a best accuracy for the number of the Gaussians equal to 2 and number of the stat equal to 4. However, the top-performing descriptor is LWE, with an accuracy of 90.23%. The next best candidate, with a lower accuracy, is DWE, which still outperforms WCC. Consequently, in the subsequent sections, we select the LWE descriptor with a configuration of 4 states and 2 Gaussians.

TABLE II. PERFORMANCE COMPARISON OF Acc_{HTK} FOR DWE, LWE, AND WCC FEATURES USING 'DB2' AND $\boldsymbol{L}_{\text{DECOMP}}$= 7, WITH AN WINDOW SIZE OF 20MS, CONSIDERING THE OPTIMAL HMM PARAMETERS, $\boldsymbol{N}_{\text{GAUSSIAN}}$, AND $\boldsymbol{N_{States}}$.

	DWE	**LWE**	**WCC**
$\boldsymbol{N_{States}}$	8	4	8
$\boldsymbol{N}_{\text{Gaussian}}$	3	2	3
$\boldsymbol{Acc_{HTK}}$ (%)	88.51	90.23	89.66

- *The impact of window size*

After selecting the LWE descriptor, this section explores the appropriate analysis frame duration. The goal of this study is to investigate performance improvements while taking into account the advantages of suitable wavelet analysis for PCG signals. To achieve this, we vary window size using the 'db2' wavelet. The other parameters are set as follows: the number of Gaussian Mixture Models ($\boldsymbol{N}_{\text{Gaussian}}$) is 2, the number of

states ($\boldsymbol{N_{States}}$) is 4, and the decomposition level ($\boldsymbol{L}_{\text{decomp}}$) is set to 7 for the LWE descriptor.

Table III displays the accuracy values (Acc_{HTK}) for various window size. The best performance is achieved with a window size of 30ms, resulting in Acc_{HTK} values of **93.68**%.

TABLE III. COMPARISON OF PERFORMANCE IN Acc_{HTK} FOR THE LWE DESCRIPTOR USING 'DB2' FOR DIFFERENT ANALYSIS FRAME DURATIONS.

	50ms	**40ms**	**30ms**	**20ms**
$\boldsymbol{Acc_{HTK}}$ (%)	87.93	85.80	**93.68**	90.23

- *Optimal Mother Wavelet and decomposition Level*

This part of the study aims to select the optimal order of mother wavelets within its family for an analysis frame duration of 30ms, a number of GMM ($\boldsymbol{N}_{\text{Gaussian}}$) equal to 2, a number of states ($\boldsymbol{N_{States}}$) equal to 4, with a decomposition level ranging from 1 to maximum level.

In this study, we consider the following wavelet families: Daubechies (Db1, Db2, ..., Db8); Coiflets family (Coif1, Coif2..., Coif5); and the Symlets (Sym1, Sym2, ..., Sym8).

The Acc_{HTK} values and $\mathbf{L}_{\text{decomp}}$ are reported in Tables IV,V, and VI for each of the three wavelet families, respectively. The obtained results highlight the robustness of the proposed approach in terms of Acc value. In particular, we also observe that the 'db2' wavelet with a decomposition level of $\boldsymbol{L}_{\text{decomp}}$= 7 leads to the maximum Acc_{HTK} = 93.68%.This latter result demonstrates an improvement in performance in terms of accuracy compared to the performance of LWE, as indicated in table II.

TABLE VI. THE PERFORMANCE IN TERMS OF Acc_{HTK} (%) AND OPTIMAL $\boldsymbol{L}_{\text{DECOMP}}$ FOR FEATURES FOR LWE USING THE DAUBECHIES WAVELET FAMILY

Daubechies		**db 1**	**db 2**	**db 3**	**db 4**	**db 5**	**db 6**	**db 7**	db8
30 ms	$\mathbf{L_{decom}}$	2	**7**	7	4	3	6	5	3
	$\boldsymbol{Acc_{HT}}$	86.21	**93.68**	92.53	89.66	91.43	90.23	89.08	91.38

TABLE V. THE PERFORMANCE IN TERMS OF Acc_{HTK} (%) AND OPTIMAL $\boldsymbol{L}_{\text{DECOMP}}$ FOR FEATURES FOR LWE USING THE SYMLETS WAVELET FAMILY

Symlets		**Sym1**	**Sym2**	**Sym3**	**Sym4**	**Sym5**	**Sym6**	**Sym7**	**Sym7**
30 ms	$\mathbf{L_{decom}}$	1	**2**	6	5	4	3	4	3
	$\boldsymbol{Acc_{HT}}$	87.36	**93.10**	89.66	88.51	89.66	89.08	89.66	89.66

TABLE VI: THE PERFORMANCE IN TERMS OF Acc_{HTK} (%) AND OPTIMAL $\boldsymbol{L}_{\text{DECOMP}}$ FOR FEATURES FOR LWE USING THE COIFLETS WAVELET FAMILY

Coiflets		**Coif 1**	**Coif 2**	**Coif 3**	**Coif 4**	**Coif 5**
30ms	$\mathbf{L_{decomp}}$	4	4	**5**	3	1
	$\boldsymbol{Acc_{HTK}}$	88.51	90.80	91.38	90.23	85.06

From these results, Daubechies wavelet of decomposition level of 7 and order 2 give the preferable Acc_{HTK}(93.68%) (see table II). Also, this latter experiment was implemented using the Symlets and the Coiflets wavelet families (table 3, 4 respectively). The Symlets wavelet with level 2 and order 2 gives the highest Acc_{HTK} of 93.10% inside her families. Also, the Coiflets with level 5 and order 3 gives the best value of Acc_{HTK} equal to 91.38%.

We can conclude from these experiments that LWE features gives the best performance results in the low order and high level (order 2 with level 7), whatever the type of the wavelet family (Daubechies, Symlets or Coiflets).

V. Conclusions

The classification and diagnosis of diseases play a crucial role in clinical practice. This paper presents a new system designed to address the classification problem of cardiovascular diseases. The proposed system based on wavelet analysis utilizes the HMM modeling for segmenting the PCG signal on S1, Systole, S2, and Diastole and classifying the signal on classes of Normal (Healthy) and abnormal. The results demonstrate that the HMM-based system achieved high classification accuracy for segmenting the PCG signal on S1, Sys, S2, and Dias regions, as well as for overall signal classification, making it well-suited for clinical applications. The proposed system yields improved results, achieving an accuracy of 93.68% using the LWE with Db2 and level 7. Our proposed method exhibits strong detection capabilities suitable for clinical applications, making it a valuable tool for the analysis of recorded PCG signals.

In future research, we plan to assess the reference system using a larger dataset. Additionally, the LWEs will be tested in various noise environments to evaluate their robustness against noisy PCG signals.

REFERENCES

[1] I. Shahid , A. Malik , C. Muhammad Usman Akram and I. Siddiqi , Localization and classification of heartbeats using robust adaptive algorithm", . Biomedical Signal Processing and Control, pp.57–77, 2019.

[2] H. Naseri and M. Homaeinezhad , Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric. of Biomedical Engineering pp.279-92.2012.

[3] D. Kumar , P. Carvalho , M. Antunes and J. Henriques , Detection of S1 and S2 heart sounds by high frequency signatures. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. (IEEE). 2006.

[4] J. Vepa , P. Tolay and A. Jain , Segmentation of heart sounds using simplicity features and timing information. International Conference on Acoustics(IEEE), Speech and Signal Processing, pp.469-72, .2008.

[5] E. Messner and F. Pernkopf , Heart Sound Segmentation An Event Detection Approach Using Deep Recurrent Neural Networks. IEEE Transactions on Biomedical Engineering. 2018.

[6] T. Chen , S. Yang , L. Ho , K. Tsai , Y. Chen , Y. Chang and Y. Lai , S1 and S2 heart sound recognition using deep neural networks. : Transactions on Biomedical Engineering (IEEE), pp.372–380, 2017.

[7] S. Sun , Z. Jiang , H. Wang and Y. Fang , Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified Hilbert transform. Comput Methods Programs Biomed, pp.219-30, . 2014.

[8] A. Moukadem , A. Dieterlena , N. Hueberb and C. Brandtc , A robust heart sounds segmentation module based on S-transform. Biomed Signal Process Control, pp.273–81 2013.

[9] C. Gupta, R. Palaniappan and S. Swaminathan, Neural network classification of homomorphic segmented heart sounds. in Appl Soft Comput.7. 286-97,2007.

[10] H. Sun , W. Chen and J. Gong , An improved empirical mode decomposition-wavelet algorithm for phonocardiogram signal denoising and its application in the first and second heart sound extraction. Biomedical Engineering and Informatics.; 2013.

[11] M. Zia, B. Griffel and J. Semmlow, Robust detection of background noise in phonocardiograms. Biomedical Engineering (IEEE), pp. 130–33. 2011.

[12] X. Wan, "Detection of the first and second heart sound using heart sound energy," 2009.

[13] A. Yadollahi and Z. Moussav, "A robust method for heart sounds localization using lung sounds entropy," 2006.

[14] L. Gamero and R. Watrous , Detection of the first and second heart sound using probabilistic models. Medicine and Biology Society (IEEE), pp.2877-80. 2003.

[15] A. Ricke, R. Povinelli and M. Johnson, Automatic Segmentation of Heart Sound Signals Using Hidden Markov Models. Computers in Cardiology 2005;32:953−956. 2005.

[16] S. Schmidt , C. Holst-Hansen, C. Graff , E. Toft and J. Struijk , Segmentation of heart sound recordings by a duration-dependent hidden Markov model. Physiol Meas , pp.513-29, 2010.

[17] D. Springer, L. Tarassenko and G. Clifford, Logistic regression-HSMM-based heart sound segmentation. (pp. 822-32). Transactions on BiomedicalEngineering (IEEE). 2016.

[18] P. Wang , Y. Kim and B. Soh , Feature extraction based on mel-scaled wavelet transform for heart sound analysis. Medicine and Biology - Proceedings; 2005.

[19] S. Rajan , E. Budd , M. Stevenson and R. Doraiswami , "Unsupervised and Uncued Segmentation of the Fundamental Heart Sounds in Phonocardiograms Using a Time-Scale Representation. Medicine and Biology – Proceedings(IEEE), 2006.

[20] P. Jain and A. Tiwari, A Robust Algorithm for Segmentation of Phonocardiography Signal Using Tunable Quality Wavelet Transform. Medical and Biological Engineering. 2018.

[21] T. Oskiper and R. Watrous, Detection of the first heart sound using a time-delay neural network . Computers in Cardiology (IEEE), (29) pp. 537–540. 2002.

[22] I. Daubechies , J. Lu and H. Wu , "Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool," 2010.

[23] V. Nivitha , K. Ramachandran and K. Soman , Wavelet-based fundamental heart sound recognition method using morphological and interval features. Healthcare Technol. Lett, pp.81–87. 2018.

[24] B. Barschdorff, A. Bothe and U. Rengshausen, "Heart sound analysis using neural and statistical classifiers a comparison," 1989.

[25] M. Milani, P. Abas and L. Silva, " A critical review of heart sound signal segmentation algorithms.," 2022.

[26] S. Li, F. Li, S. Tang and F. Luo, " Heart sounds classification based on feature fusion using lightweight neural networks," 2021.

[27] R. Saracoglu , "Hidden Markov model-based classification of heart valve disease with PCA for dimension reduction.," 2012.

[28] D. Springer , L. Tarassenko and . G. Clifford, "Support vector machine hidden semi-markov model-based heart sound segmentation," 2014.

[29] P. Wang, C. S. Lim, S. Chauhan, J. Y. Foo and V. Anantharaman, "Phonocardiographic signal analysis method using a modified hidden Markov mode," 2007.

[30] . S. Young, D. Kershaw, . J. Odell and D. Ollason, "The HTK Book," 1999.

[31] P. Bentley, . G. Nordehn, M. Coimbra, S. Mannor and R. Getz, " The PASCAL classifying heart sounds challenge.," 2011.

[32] S. Young, D. Kershaw, J. Odell and D. Ollason, The HTK Book. Cambridge: Entropic Ltd. 1999

[33] Touahria.R , A. Hacine-Gharbi and P. Ravier, Discrete Wavelet based Features for PCG Signal Classification using Hidden Markov Models. International Conference on Pattern Recognition Applications and Methods; 2021.

[34] B. Lekram and M. Abhishek, Heart Sounds Segmentation Analysis Using Daubechies Wavelet (db). International Journal of Research in Advent Technology,pp.2321-9637 . 2014.

[35] "praat.fr.softonic.com/., P. (. (s.d.).".

[36] P. Ravier and A. Hacine-Gharbi, Wavelet Cepstral Coefficients for Electrical Appliances Identification using Hidden Markov Models. International Conference on Pattern Recognition Applications and Methods ICPRAM.,2018.

[37] H Bengacemi,P. Ravier and A. Hacine-Gharbi, P. Ravier, K Abed-Meraim and O.Buttell, Surface EMG Signal Segmentation and Classification for Parkinson's. International Conference on Pattern Recognition Applications and Methods ICPRAM.,2024.

Implementation of VGG for Pneumonia Detection and Brain Tumor Classification Using Raspberry Pi

Cherifa ABDELAOUI
Department Engineering of Electrical Systems
Faculty of Technology, LIST Laboratory
35000 Boumerdes, Algeria
Research Center in Industrial Technologies (CRTI); P.O.Box 64, Cheraga 16014 Algiers, Algeria
c.abdellaoui@univ-boumerdes.dz

Samia BELKACEM
Department Engineering of Electrical Systems
Faculty of Technology
35000 Boumerdes, Algeria
s.belkacem@univ-boumerdes.dz

Noureddine MESSAOUDI
Department Engineering of Electrical Systems
Faculty of Technology, LIST Laboratory
35000 Boumerdes, Algeria
n.messaoudi@univ-boumerdes.d

Redouane DRAI
Research Center in Industrial Technologies (CRTI); P.O.Box 64, Cheraga 16014 Algiers, Algeria
r.drai@crti.dz

Mounia BACHI
Department Engineering of Electrical Systems
Faculty of Technology
35000 Boumerdes, Algeria
mounia.bachi@gmail.com

***Abstract*— Medical imaging has become an indispensable tool in diagnosing various diseases. Traditionally, medical image analysis heavily relies on manual interpretation. However, with the advent of deep learning techniques, automated image classification has emerged as a powerful tool. In this work, we utilized the VGG16 architecture to develop a medical image classification model capable of detecting brain tumors and pneumonia. To enhance accessibility and portability, the models was implemented on a Raspberry Pi4. A user-friendly graphical interface was developed to facilitate the visualization and interpretation of results. The model was trained on a comprehensive dataset and evaluated using standard performance metrics like accuracy, precision, and F1-score. The results demonstrated the model's ability to accurately classify medical images. This project highlights the potential of deep learning and embedded systems to revolutionize medical image analysis. By enabling rapid and accurate diagnosis, this technology can significantly improve patient care and healthcare outcomes.**

Keywords*— *Medical image, Deep learning, VGG16, Brain tumor, Pneumonia, Classification.

I. INTRODUCTION

Diagnosis and treatment of diseases have become essential thanks to medical imaging, which provides detailed, non-invasive visualizations of the human body's interior. However, the increasing volume of medical image data has created an urgent need for efficient methods for their analysis and evaluation.

The field of medical imaging has undergone significant changes in recent years due to major advancements in deep learning and computer vision. Medical imaging and artificial intelligence now play a key role in the treatment of various diseases [1].

Recent advancements in machine learning, particularly in deep learning, present significant opportunities to improve medical image segmentation and classification through the provision of more precise and reproducible automated analysis tools. Deep learning is currently receiving considerable attention for its application with extensive healthcare data.

Deep learning is a powerful machine learning technique that involves training artificial neural networks with multiple layers to learn complex patterns from data. This approach has revolutionized various fields, including speech recognition, object detection, and many others. Training can be supervised, where the model is trained on labeled data, or unsupervised, where the model learns patterns without explicit labels [2].

In today's world, medical specialists are inundated with vast amounts of data. The human capacity to analyze and process this data is limited, leading to potential fatigue and reduced efficiency in patient care. A significant portion of this data, approximately 90%, consists of medical images that require careful analysis for accurate diagnosis and treatment planning. The increasing demand for medical image analysis has opened up significant opportunities for the development of innovative healthcare solutions [3].

To achieve these objectives, we bridge two fields: medicine and artificial intelligence. We propose the VGG16 architecture, based on deep learning, to classify images in our dataset into two classes for each database: "Pneumonia" and "Normal" for the pneumonia dataset, "Yes" and "No" for the brain dataset.

The primary goal of medical image classification is to automatically categorize images into different classes or groups based on their visual characteristics and the information they contain. This enables healthcare professionals to diagnose diseases more quickly and accurately, identify anomalies, and monitor treatment progress [4].

Enhanced Diagnosis: By classifying images, specific elements like tumors or fractures can be detected, accelerating the diagnostic process.

Treatment Monitoring It allows for the comparison of images over time to evaluate the efficacy of a treatment.

Research and Development: Classification can be used for research, helping to identify patterns or common characteristics in groups of images.

Machine Learning and AI: Machine learning techniques, especially convolutional neural networks (CNNs), are at the heart of modern medical image classification, enabling deeper and more accurate analysis [4].

II. RELATED WORKS

In this section, we will present some research works on medical image classification:

Amelia Ritahani Ismail et al. [5] proposed the VGG16 architecture, renowned for its ability to extract important features. The model has demonstrated effectiveness in identifying anomalies in medical images, particularly in the context of brain tumor datasets. Evaluation metrics such as sensitivity, specificity, and F1-score highlight the model's ability to accurately distinguish between various medical image diseases. Notably, the VGG-16 model has shown promising results in predicting brain tumors. The Brain Tumor MRI dataset yielded significantly superior accuracy compared to the imbalanced Alzheimer's dataset. This underscores the significant impact of dataset characteristics and complexity on model performance.

Yogesh Kumaran et al [6] suggested an integrated deep learning approach was proposed, leveraging pre-trained models—VGG16, ResNet50, and InceptionV3—within a unified framework to enhance diagnostic accuracy in medical imaging. The method specifically targets lung cancer detection, utilizing images resized and standardized for optimal performance and consistency across datasets. The proposed model achieved an impressive accuracy of 98.18%, with notably high precision and recall rates across all classes. This significant improvement underscores the potential of integrated deep learning systems in medical diagnostics, offering a more accurate, reliable, and efficient means of disease detection. The model demonstrated exceptional performance in terms of precision, recall, and F1-scores across three categories: Benign, Malignant, and Normal cases, solidifying its potential as a reliable diagnostic tool.

Chiranjibi Sitaula et al [7], explored a novel attention-based deep learning model, incorporating VGG-16, to diagnose COVID-19 using chest X-ray (CXR) images. By employing an attention module, the model captures spatial relationships between regions of interest (ROIs) within CXR images. Additionally, the integration of a specific convolutional layer (the 4th pooling layer) from the VGG-16 model further enhances the model's ability to perform fine-tuned classification. The promising classification performance achieved by this proposed method suggests its suitability for COVID-19 diagnosis using CXR images.

Haoyan Yang1 et al [8], designed a novel method for crop variety identification, using peanuts as a case study. Traditional image processing techniques for peanut variety identification often rely on extracting numerous features, which can be subjective and lack generalization ability. To address these limitations, the study improves the VGG16 deep convolutional neural network and applies it to the identification and classification of 12 peanut varieties. Experimental results demonstrate the effectiveness of the improved VGG16 model in accurately identifying and classifying peanut pods of different varieties, confirming the feasibility of convolutional neural networks for crop variety identification and classification tasks. This research holds significant promise for exploring similar applications in other crop varieties. The successful improvement of the VGG16 model highlights the potential of convolutional neural networks in the field of crop variety identification and classification.

III. VGG16

VGG16 is a widely used deep learning architecture introduced by Oxford University [9]. Figure 1 illustrates its prominent architecture, composed of 16 weight layers, 13 convolutional layers (Conv.), and 3 fully connected (FC) layers. VGG16 utilizes small 3x3 convolutional kernels with a stride of one on all convolutional layers. Max pooling layers follow each convolutional layer. The network is designed to process fixed-size 224x224 images with three colors channels.

The first two fully connected layers have 4096 channels, while the final FC layer has 1000 channels, corresponding to the number of classes in the ImageNet dataset. The output layer is a softmax layer, which assigns a probability to each class for the input image.

VGG16 [9], like other pre-trained models, demands significant training time if initialized with random weights. To mitigate this, Convolutional Neural Networks (CNNs) often leverage transfer learning (TL). TL involves adapting a model trained on one task to a related task. In this context, a CNN model is trained on a similar problem with the same input type but potentially different output classes. For instance, VGG16 is trained on the ImageNet dataset, which comprises numerous real-world object images.

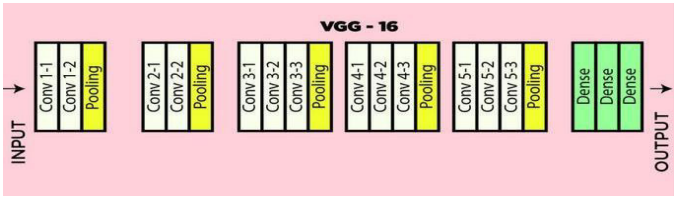


Fig. 1. VGG16 architecture

IV. MATERIALS AND METHOD

A. *Evaluation metrics*

This section outlines various metrics used to assess the model developed for classifying tumor brain and pneumonia, these metrics provide a comprehensive understanding of the model's accuracy, reliability, and diagnostic ability. Which includes the number of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) predictions [6].

TABLE I. DISCRIPTION EVALUATION METRICS

Indicator	Description	Equation
Accuracy	Measures the overall correctness of the model and is defined as the ratio of correctly predicted observations to the total observations	$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$
Precision	Measures the accuracy of positive predictions. It is defined as the ratio of true positive predictions to the total predicted positives	$Precision = \frac{TP}{TP + FP}$
Recall	Indicates the ability of the model to find all relevant cases within a dataset. It is defined as the ratio of true positives to the actual number of positives	$Recall = \frac{TP}{TP + FN}$
F1 score	Is the harmonic mean of precision and recall it is particularly useful when the class distribution is uneven. The score takes both false positives and false negatives into account and is a better measure of the incorrectly classified cases than the Accuracy Metric	$F1 = \frac{2 * TP}{2 * TP + FN + FP}$

B. Materials

For this project, we used a set of equipment that has the following characteristics:

TABLE II. MATERIALS USED IN THE PROJECT

Device	Model
Implementation model	Raspberry Pi 4 B, 2GB Ram
Screen	7-inch LCD touch screen HDMI for Raspberry ultra-clear resolution of 1024x600
Battery	For alimentation of raspberry and LCD screen 12V 11A

C. DATABASE

For this study, we used two datasets: one for brain tumors and the other for pneumonia. The first dataset contains 1085 'YES' images (with tumors) and 980 'NO' images (without tumors), while the second dataset contains 992 'Normal' images and 1924 'Pneumonia' images. Both datasets were obtained from the open-source platform Kaggle/dataset [10] [11].

The classification model demonstrates very strong overall performance with high precision (0.94) and recall (0.94). It appears to be particularly effective at detecting cases of pneumonia (F1-score of 0.96) but shows a slight decrease in recall for normal cases (0.88), indicating that some normal cases are incorrectly classified as pneumonia. Nevertheless, the high scores demonstrate that the model is well-balanced and reliable for this classification task. About MRI data The model exhibits very good performance on the classification task, as evidenced by the classification report. All major performance metrics, such as precision, recall, and F1-score, are very close to 1.00 for both the "no" and "yes" classes. Furthermore, the overall accuracy of the model is 100%, meaning that all predictions made by the model are correct.

TABLE III. DATABASE DISCRIPTION AND RESULTS

	X-ray		MRI	
	Pneumonia	Normal	Yes	No
Presecion	0.94	0.95	1	0.99
Recall	0.98	0.88	0.99	1
F1 score	0.96	0.92	1	1
Accuracy	0.9964		1	
Number of Images	2846		2065	

V. IMPLEMENTATION AND RESULTS

A. RESULTS

After following the previous steps, we tested the pre-trained model using images to perform classification. The figure.2 presents the obtained results.

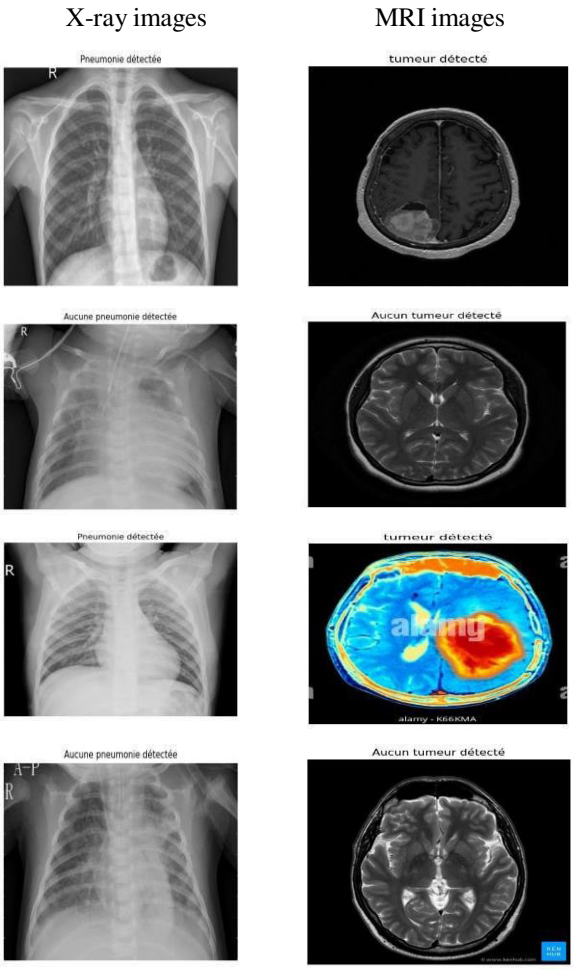


Fig. 2. Classification results

B. *IMPLEMENTATION*

This section is dedicated to the implementation of an image classification solution using the VGG16 model, along with its deployment on the Raspberry Pi 4 microcomputer using the obtained model. Subsequently, we will define the selected criteria to evaluate the performance of this model. Finally, we will analyze the results obtained by implementing the VGG16 model on a Raspberry Pi and The graphical user interface (GUI) will be used to display the results.

In this work, we used the VGG16 architecture to train different datasets and selected the best model to perform our tests. We upload the selected image via a USB device to the Raspberry Pi and perform image classification with the materials used as shown in the figure3.

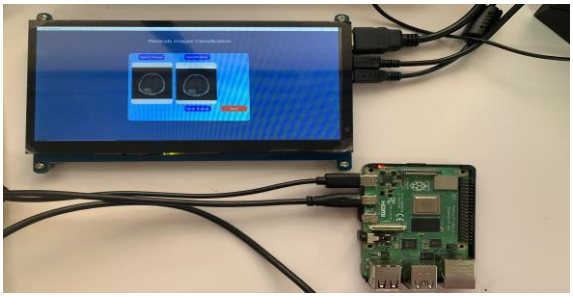


Fig. 3. IMPLEMENTATION on Raspberry PI4.

The developed interface is shown in the figure (4). The button“Select image” allows us to select the image that will be classified. While the click on the button” classification” will do the classification task according to the VGG16 architecture.



Fig. 4. Graphical Interface.

VI. DISCUSSION

The objective of this study was to develop a medical image classification model using the VGG16 architecture to detect brain tumors and cases of pneumonia. The accuracy of these classifications is crucial for providing reliable diagnostic tools that can assist healthcare professionals in making informed decisions.

Tumor brain detection :

- Avec une précision de 99.68%, le modèle montre une très haute fiabilité dans la détection des tumeurs cérébrales.
- The model's low False Positive and False Negative rates indicate excellent performance, as evidenced by the minimal number of incorrect predictions. This suggests that the model is well suited for generalization to new data.

Pneumonia Detection :

- Une précision de 94.29% est indicative d'une capacité solide du modèle à différencier les images de poumons atteints de pneumonie des poumons sains.

Even with a slightly lower accuracy, this model can be extremely useful for screening pneumonia, helping to reduce diagnostic errors and accelerate the diagnostic process.

The results highlight the VGG16 architecture's strong performance on MRI images, particularly those depicting brain lesions. The somewhat lower accuracy on chest X-rays may be attributed to greater variability within this dataset or to subtle distinctions between healthy and diseased lungs, increasing the complexity of classification tasks.

VII. CONCLUSION

In technology, deep learning plays a very important role in image segmentation. The various studies confirm that applications of deep neural networks in the task of medical image classification.

The Raspberry Pi implementation has proven that embedded systems can be effectively used for medical image classification. This portable solution holds immense promise for rapid diagnostics, especially in resource-constrained clinical settings and it can be easily transported like an instrument.

In conclusion, this project will serve as application in using deep learning for classification of tumor brain and pneumonia detection in fast way by implementation, this application can be used to create reports and automate processes.

VIII. REFERENCES

[1] Deep Learning in Medical Imaging: General Overview, June-Goo Lee et all 2017.

[2] Brain Tumor Classification Using Deep Learning, Ahmad Saleh et all 2020.

[3] A deep learning approach for brain tumor classification using MRI images, Muhammad Aamir et all 2022.

[4] The most efficient CNN architectures in 2021 for deep learning classification in medical imaging, Sarah Madeleine 2022.

[5] Utilising VGG-16 of Convolutional Neural Network for Medical Image Classification 2019.

[6] Explainable lung cancer classification with ensemble transfer learning of VGG16, Yogesh Kumaran S1, J. Jospin Jeya2, Mahesh T R1*, Surbhi Bhatia Khan3, Saeed Alzahrani4 and Mohammed Alojail4 2024.

[7] Attention-based VGG-16 model for COVID-19 chest X-ray image classification, Chiranjibi Sitaula1 · Mohammad Belayet Hossain1 2021.

[8] A novel method for peanut variety identification and classification by Improved VGG16, Haoyan Yang1, Jiangong Ni2, Jiyue Gao2, Zhongzhi Han2* & Tao Luan1* , 2021.

[9] Dheeb Albashish, Rizik Al-Sayyed et all, Deep CNN Model based on VGG16 for Breast Cancer Classification 2021.

[10] https://www.kaggle.com/datasets/andrewmvd/pediatric-pneumonia-chest-xray.

[11] https://www.kaggle.com/datasets/arifmia/brain-tumor-dataset