Hajji Bekkay · Adel Mellit ·
Antonio Gagliano · Abdelhamid Rabhi ·
Mohammed Amine Koulali   *Editors*

Proceedings of the 3rd International Conference on Electronic Engineering and Renewable Energy Systems

ICEERE 2022, 20–22 May 2022, Saidia, Morocco

Springer

Lecture Notes in Electrical Engineering

Volume 954

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering—quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Editorial Director (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

** This series is indexed by EI Compendex and Scopus databases. **

Hajji Bekkay · Adel Mellit · Antonio Gagliano ·
Abdelhamid Rabhi · Mohammed Amine Koulali
Editors

Proceedings of the 3rd International Conference on Electronic Engineering and Renewable Energy Systems

ICEERE 2022, 20–22 May 2022, Saidia, Morocco

Springer

Editors
Hajji Bekkay
National School of Applied Sciences
Mohamed Premier University
Oujda, Morocco

Adel Mellit
Faculty of Sciences and Technology
Jijel University
Jijel, Algeria

Antonio Gagliano
Department of Electrical Engineering,
Electronics and Computer Engineering
University of Catania
Catania, Italy

Abdelhamid Rabhi
University of Picardie Jules Verne
Amiens, France

Mohammed Amine Koulali
National School of Applied Sciences
Mohammed Premier University
Oujda, Morocco

Organization

Honorary Committee

Mohamed Ali Habouha, Governor of Berkane Province, Morocco
Yassine Zarloul, President of Mohammed First University, Morocco
Mohamed Ibrahimi, President of Berkane Urban Commune, Morocco
Hassan Ettahiri, Director of Colaimo, Oujda, Morocco

General Chairs

Hajji Bekkay, ENSA-Oujda, Mohammed First University, Morocco
Abdelhamid Rabhi, University of Picardy Jules Verne, France
Adel Mellit, Jijel University, Algeria

General Co-chairs

Mohamed Amine Koulali, ENSA-Oujda, Morocco
Antonio Gagliano, University of Catania, Italy
Giuseppe Marcio Tina, University of Catania, Italy
Michele Calì, University of Catania, Italy

Steering Committee

Loubna Bouselham, ENSAOujda, Morocco
Mohammed Amine Koulali, ENSA-Oujda, Morocco

Mohammed Essoufi, ENSAOujda, Morocco
Mohamed Hajji, ENSAOujda, Morocco
Hanae Azzaoui, ENSA-Oujda, Morocco

Students Committee

Chaimae El Fouas, ENSA-Oujda, Morocco
Safae Margoum, ENSA-Oujda, Morocco
Essoufi Mohamed, ENSA-Oujda, Morocco
Ismail Mir, ENSA-Oujda, Morocco
Mohammed Benzaouia, ENSA-Oujda, Morocco
Oussama El Manssouri, ENSA-Oujda, Morocco
Ayat Yassine, ENSA-Oujda, Morocco
Nabil Ayadi, ENSA-Oujda, Morocco

Technical Program Committee

A. Massi Pavan, University of Trieste, Italy
S. Mekhilef, University of Malaysia
A. Elahi, Southern Connecticut State University, USA
E. Quaranta, European Commission, Joint Research Centre, Italy
Adel Mellit, University of Jijel, Algeria
Zineb Bougroun, ENSA-Oujda, Morocco
Abdellah Mechaqrane, FST-Fes-Morocco
J. Bosche, University of Picardie, France
S. Safak, University of Marmara, Turkey
M. Benghanem, University of Madinah, KSA
Şafak Sağlam, Marmara University, Turkey
A. Gagliano, Université de Catane, Italy
Y. Al Younes, United Arab Emirates
J. Launay, LAAS-CNRS, France
M. Ben Ammar, ENIS, Tunisia
F. Farmakis, University of Thrace, Greece
P. Temple-Boyer, LAAS-CNRS, France
N. Msirdi, LSIS-UMR France
F. Tadeo, Universidad de Valladolid, Spain
A. Migan-dubois, SUPELEC, France
A. Lakhssassi, University of Quebec, Canada
K. Khodja, University of Sciences and Technology, Algeria
Abdelhamid Rabhi, University of Picardie, France
A. Boualit, URAER, Algeria

M. Heyouni, ENSA-Oujda, Morocco
G. M. Tina, University of Catania, Italy
I. Abdi Hadi, Université, Djibouti
A. Mpanda, ESIEE, Amiens, France
A. Aissat, Blida University, Algeria
Omar Moussaoui, EST-Oujda, Morocco
L. Bouselham, ENA-Oujda, Morocco
Vicenc Puig, Universitat Politècnica de Catalunya, Spain
E. Ogliari, Politecnico di Milano, Italy
D. Benhaddou, University of Houston, USA
M. Belkheiri, Université Amar Telidji de Laghouat, Algeria
C. Jean-Yves, University of Laval Canada
A. Boutayeb, FSO-UMP, Morocco
D. M. Grasso, University of Catania, Italy
M. Alami Marktani, ENSA-FES, Morocco
Naamane Abdelaziz, Université d'Aix Marseille, France
M. Zekriti, Euromed University-FES, Morocco
Augustin Mpanda, ESIEE-Amiens, France
Francesco Nocera, University of Catania, Italy
T. Jarou, Ibn Tofail University, Morocco
Nasiruddin Muhammad, Universiti Sains, Malaysia
D. Ishak, Universiti Sains, Malaysia
Mohamed Khafallah, ENSEM, Morocco
Boualit Hamid, CDER, Algeria
F. Chekired, UDES, EPST-CDER, Algeria
Laili Djaafer, Jijel University, Algeria
A. Mbarki, ENSA-Oujda, Morocco
Y. Reggad, Med First University, Morocco
M. Hajji, ENSA-Oujda, Morocco
A. El Moussaouy, The Regional Center for the Professions of Education and Training, Oujda, Morocco
O. Mommadi, FSO-UMP, Morocco
El-Kaber Hachem Faculté des Sciences, Université Moulay Ismail, Meknès, Morocco
M. Saber, ENSA-Oujda, Morocco
Boukenoui Rachid, Blida University, Algeria
M. G. Belkasmi, ENSA-Oujda, Morocco
A. El Moussati, ENSA-Oujda, Morocco
Y. G. Dessouky, AASTMT, Egypt
B. Oral, Marmara University, Turkey
A. Drighil, Université Hassan II, Morocco
Haddad Sofiane, Jijel University Algeria
Daoudi Salah, FSO-Oujda, Morocco
A. Alami, USMBA, FES, Morocco
M. Belkheiri, Université Amar Telidji de Laghouat, Algeria
A. Aissat, University of Blida, Algeria

A. Messaoudi, EST-Oujda, Morocco

H. Aitabbas, UATL, Algeria

Ilhame El Farissi, ENSA-UMP, Morocco

Ouadoudi Zytoune, University Ibn Tofail, Kénitra, Morocco

M. Emharraf, ENSA-UMP, Morocco

Driss Bria, Med First University, Morocco

A. El Ougli, ENSA-Oujda, University Med First, Morocco

Zyane Abdellah, ENSA-Safi, University Med First, Morocco

H. Zahboune, EST-Oujda, Morocco

Senhorinha Teixeira, University of Minho, Portugal

Kerkour-El Miad, EST-Oujda, Morocco Stefano Aneli University of Catania, Italy

Amraqui Samir, EST-Oujda, Morocco

Imad El Mahi, ENSA-Oujda, Morocco

Benslimane Anas, ENSA-Oujda, Morocco

M. Heyouni, ENSA-Oujda, Morocco

El Houssaine El Boudouti, FSO-Oujda, University Med First, Morocco

Abdelali Ed-Dahhak, Moulay Ismail University, Morocco

A. Mbarki, ENSA-Oujda, Morocco

Mohammed Amine Koulali, ENSA-Oujda, Morocco

Ait Madi Abdessalam, Ibn Tofail University, Morocco

Kassmi Kamal, EST-Oujda, University Med First, Morocco

Falyouni Farid, Med First University, Morocco

A. Benami, FST-Errachidia, Morocco

Guillaume Caron, University of Picardy Jules Verne, France

Y. Khlifi, ENSA, Oujda, Morocco

T. Jarou, ENSA-Kenitra, Morocco

R. El Ghouri, ENSA-Kenitra, Morocco

D. Lara, Tecnologico Nacional de México

S. Naimi, ENSA-Oujda, Morocco

A. Azizi, EST-Oujda, Morocco

M. El Arbi, El Hafyani, ENSA-Oujda, Morocco

J. Ezzahar, ENSA-Safi, Morocco

A. Galadi, ENSA-Safi, Morocco

A. Mansouri, ENSA-FES, Morocco

S. D. Bennani, ENSA-FES, Morocco

D. Yousfi, ENSA-Oujda, Morocco

Elmiloud Chaabelasri, FSO-UMP, Morocco

A. A. Merrouni, FSO-UMP, Morocco

Hanae Azzaoui, ENSA-Oujda, Morocco

H. Qjidaa, Sidi Mohamed Ben Abdallah University, FES, Morocco

A. El Mehdi, ENSA-Oujda, Morocco

Michele Calì, University of Catania, Italy

T. Sgheir, University of Laghouat

R. El Bouayadi, ENSAK, Morocco

Fatima-zahra El Amri, FSO-UMP, Morocco

M. Benhaddou, ENA-Oujda, Morocco
A. Touhami, University of Sidi Bel-Abbes, Algeria
Abdelmalik Zorig, Université Mohamed Boudiaf M'sila, Algeria
Derouich Aziz, EST-FES, Morocco
Ghita Zaz, USMBA-FES, Morocco
Chouki Sentouh, UMR CNRS, UPHF, France
Houcine Chafouk, University of Rouen Normandy, France

Preface

Brief Synopsis About ICEERE 22 Book

The book ICEERE 22 deals with the broad discussion regarding recent advances and research issues related to the fields of electronic engineering, control and instrumentation, embedded systems, communication and renewable energy. The content of this book is focused on the new concepts and latest results on the application of machine learning methods and Internet of Things to address problems in renewable energy systems. The advance of renewable energy applications would not have been possible without the advance of electronic and information technologies.

With the successful experience of the first and second editions (in Saidia, Morocco, 2018–2020), we truly believe that the third edition of ICEERE 2022 will achieve greater success and provide a better platform for all the participants to have fruitful discussions and to share ideas of researches.

We expect that the published papers in the conference will be a trigger for further related research and technology improvements in this importance subject.

ICEERE 22 will also include presentations of contributed papers and state-of-the-art lectures by invited keynote speakers. The book has a special focus on Internet of Things for smart agriculture powered by renewable power energy systems. Special interest will also be given to the energy challenges for developing the Euro-Mediterranean regions through new renewable energy technologies in the agricultural and rural areas.

We would like to thank the program chairs, organization staff and the members of the program committees for their hard work. Special thanks go to Springer publisher. We hope that ICEERE 2022 will be successful rand enjoyable to all participants. We look forward to seeing all of you in two years at ICEERE 2024.

Oujda, Morocco	Hajji Bekkay
Jijel, Algeria	Adel Mellit
Catania, Italy	Antonio Gagliano
Amiens, France	Abdelhamid Rabhi
Oujda, Morocco	Mohammed Amine Koulali

Invited Speakers: Biography and Abstract

Smart Renewable Energy Systems and Decarbonisation

Dr. Henrik Lund
Editor-in-Chief of Journal *Energy*, Denmark

Biography

Henrik Lund (born July 2, 1960) is Danish Engineer (M.Sc. Eng. 1985) and Professor in Energy Planning at Aalborg University in Denmark. He holds a Ph.D. in Implementation of Sustainable Energy Systems (1990) and Dr.Techn. in Choice Awareness and Renewable Energy Systems (2009).

He is a highly ranked world-leading researcher. He is listed among ISI Highly Cited researchers ranking him among the top 1% researchers in the world within engineering.

He has many years of management experience as head of the department for approximately 200 staff persons (1996–2002), head of the section for approximately 50 persons (2014–2016) and head of the research group of 20–30 persons

(2002–present). During his time, the Sustainable Energy Planning Research Group at Aalborg University has now grown to approximately 30 staff members including 4 professors.

He is Editor-in-Chief of Elsevier's high-impact journal *Energy* with annual 10,000+ submissions.

He is Author of more than 400 books and articles including the book *Renewable Energy Systems.* He is the architect behind the advanced energy system analysis software EnergyPLAN, which is a freeware used worldwide that have form the basis of more than 200 peer-reviewed journal papers around the world.

Abstract: Smart Renewable Energy Systems and Decarbonisation

This presentation focuses on how societies can design and implement renewable energy and decarbonization strategies. The presentation presents and discusses a set of methods and criteria to design smart energy systems, while taking into account the context of 100% renewable energy on a national level. Countries should handle locally what concerns local demands, but acknowledge the international context when discussing resources and industrial and transport demands. To illustrate the method, it is applied to the case of Denmark within the context of a European and a global energy system.

Recently, the Danish Government supported by the Danish Parliament decided for the target of a 70% decrease in greenhouse gasses by 2030. This presentation includes a list of theoretical and methodological considerations as well as a concrete proposal on how such targets can be implemented. It is highlighted that already now one have to think beyond 2030 in order to prepare for the next step to achieve a fully decarbonization by 2040 or 2050. It is also highlighted that a country such as Denmark have to consider how to include its share of international shipping and aviation as well as how to design a solution with Denmark's share of sustainable biomass resources.

Book Publishing with Springer

Dr. Ramesh Nath Premnath
Editor Applied Sciences
Engineering, Materials Science, Energy, Water, Climate
Senior publishing editor for academic books
Springer Singapore

Biography

Ramesh Nath Premnath is Senior Publishing Editor for academic books covering the region of South East Asia and Australasia region. His academic background includes B.Sc., M.Sc., Ph.D. in Materials Science and Engineering with 4 years as Postdoctoral Researcher working in the field of ferroelectric/piezoelectric materials. During his time at Springer, he has launched/takeover several scientific journals and has developed a substantial book program from the region has published more than 400 academic books.

Abstract: Book Publishing with Springer

In this talk, he will briefly introduce Springer Nature publishing program and highlight the different academic book products and online learning platforms that Springer Nature offers for our users. In turn, he will also reach out to prospective authors who are interested to publish their next book or online resources with Springer Nature by providing the step-by-step process in their route to publication.

Internet of Things: Applications, Enablers, Security

Dr. Manu Malek
Editor-in-Chief of Elsevier's
International Journal of Computers and Electrical Engineering, USA

Biography

Manu Malek is Editor-in-Chief of Elsevier's *International Journal of Computers and Electrical Engineering.* He was Distinguished Member of Technical Staff at Lucent Bell Labs until 2001; then, he joined Stevens Institute of Technology as Professor of Computer Science and Telecom Management, from which he retired a few years ago.

He has held various academic positions in the USA and overseas, as well as technical leadership positions at Telcordia Technologies and AT&T/Lucent Bell Labs. He is the author, co-author or editor of seven books and author or co-author of numerous publications in the areas of communication networks, computer communications and network management.

An alumnus of University of California, Berkeley, he is Life Fellow of the IEEE, and an IEEE Third Millennium medalist for his contributions. He was Editor for Network Management of the IEEE Transactions on Communications (1989–92) and IEEE ComSoc Distinguished Lecturer (1999–2007). He founded and was Editor-in-Chief of Springer's *Journal of Network and Systems Management* (1993–2010).

He is a frequent invitee as keynote speaker at international conferences. His favorite topic is information security with the subtopics IoT, security forensics and defense and autonomous vehicles.

Abstract: Internet of Things: Applications, Enablers, Security

The Internet is used daily for many online services: for communicating, finding information, doing transactions and for entertainment; and this is possible independent of location due to its openness and distributed nature. However, this very openness has

posed vulnerabilities, allowing attacks such as ID theft, denial of service, industrial espionage and extortion. The Internet of Things (IoT) anticipates connecting billions of objects to the Internet in the next few years, thus potentially creating additional challenging security and privacy issues.

The objective of this talk is to describe current trends affecting Internet security, and point out the sources of security problems, as well as provide an overview of IoT and its related security issues. IoT enablers and typical IoT applications will be presented. Also, some potential security solutions, safeguards and defenses will be offered.

One-Dimensional Photonic Crystals: Fundamentals and Applications

Dr. El Houssaine El Boudouti
Mohammed First University, Morocco

Biography

Dr. El Houssaine El Boudouti is Professor at the Department of Physics, Faculty of Sciences, Mohammed First University, Oujda (Morocco). He is M.Sc. (1992), Ph.D. (1994) and Habilitation (2007) from the University of Lille (France). His research interest concerns elementary excitations (phonons, photons, electrons, plasmons and magnons) in composite materials such as phononic, photonic, electronic and magnonic crystals. He published 114 articles, 3 books, 1 chapter book and 2 review papers. He presented his research work in 100 national and international conferences. He is awarded the "Research Prize of excellence in Physics" from Mohammed First University in 2018.

Abstract: One-Dimensional Photonic Crystals: Fundamentals and Applications

Photonics is the physical science and application of electromagnetic waves such as generation, detection and manipulation through emission, transmission, modulation, switching and sensing. This science is equivalent to what electrons do in semi-conductors in terms of band gaps, transmission spectra, etc. In this talk, I will present the main results of our book entitled *Photonics* published in Elsevier (L. Dobrzynski et al. 2020) which describes the science of photonic transmission properties of the interfaces of composite materials systems and devices. I will review the general analysis methods of interface transmission and give many examples applying these methods to one-dimensional photonic systems. Applications discussed include photonic circuits, layered materials and devices.

Health Monitoring Systems for the Renewable Energy

Dr. Houcine Chafouk
ESIGELEC, University of Rouen Normandy, France

Biography

Houcine Chafouk, IEEE Member, is Professor at ESIGELEC and Researcher at IRSEEM/University of Rouen Normandy, France; he obtained a doctorate in automation at the University of Nancy, Lorraine, France, in 1990, and then, he joined the same year the engineering school ESIGELEC, Rouen. From 2000 to 2008, he held the position of the director of research and head of the research team in automation and systems. Since 2000, he has supervised around thirty doctoral, postdoctoral and HDR students who have carried out their research within IRSEEM or with international partners. He also participated in thesis juries as rapporteur (20 theses) and examiner (10 theses). He is the author and co-author of more than 150 research articles (publications and communications) in the fields of advanced control systems, diagnostics

and fault-tolerant control applied to the fields of renewable energy, automotive and energy and aerospace.

Abstract: Health Monitoring Systems for the Renewable Energy
The latest developments about the monitoring, the diagnosis and the fault-tolerant control of a windfarm are presented. A wireless sensor and actuator network are issued for the supervisory control and data acquisition. As there is a large amount of information, big data processing frameworks are needed and different computing paradigms are investigated to store and process the information in real time, which reduces cost and added security. The fault detection and identification must be done with fast and precise algorithms to minimize the number of false alarms. Also, it must consider the great multitude of defects that can appear in a wind turbine, from internal component failures to outside influences. Fault-tolerant control needs to consider not only the cases when a turbine fails completely, but also when some of them are not working properly, or when the environmental conditions, e.g., the wind, are not optimum. Perspectives on future developments, especially on how the different components can be integrated together, in the case of a windfarm, are given.

Energy Performances of a Photovoltaic Thermal Plant Using Different Coolant Nanofluid

Dr. Antonio Gagliano
University of Catania, Italy

Biography
He teaches at Engineering Faculty of University of Catania. He gives academic courses on the following subjects: Environmental Applied Physics, Acoustic and Environmental Control Techniques (since 2001). He has supervised more than 50

thesis of master and tutor of five Ph.D. thesis. He is the scientific supervisor for research contract, scholarship, research grant activated within research projects, and he is Scientific Manager of the Laboratory of Applied and Building Physics of the Department of Electric, Electronic and Computer engineering (DIEEI). He is Member of the editorial board of the following journals: *Energies*; *AIMS Energy*; *Energy Sources*; *Part A: Recovery, Utilization, and Environmental Effects*; *The Open Civil Engineering Journal*; *African Journal of Environmental Science and Technology*; *TECNICA ITALIANA-Italian Journal of Engineering Science*; *Progress in Energy and Fuels*. He is an regular reviewer for the following international journals: *African Journal of Environmental Science and Technology*, *Automation in Construction, Building and Environment, Buildings, Energy and Building, Energy, Energy Conversion and Management*, *Energy Efficiency, Engineering Science and Technology an International Journal, Energy and Environment, Environmental Science and Pollution Research, International Journal of Heat and Mass Transfer*, *International Journal of Thermal Science, International Journal of Geoinformation, International Journal of Sustainable Energy*, *Journal of Environmental and Public Health*, *Journal of Computational Environmental Sciences*, *Journal of Building Engineering Progress in Computational Fluid Dynamics*, *Renewable Energy, Sustainable Cities and Society*, *Sustainability, Transportation Research Part A: Policy and Practice*, *Urban Forestry and Urban Greening*.

Abstract: Energy Performances of a Photovoltaic Thermal Plant Using Different Coolant Nanofluid

In the last four decades, greater attention has been paid to PV/T systems due to their advantages compared with PV or solar thermal systems alone. Recently, nanofluids have been implemented as heat transfer fluids in PV/T systems. The thermal properties of several investigated water nanofluids are derived from literature data. In particular, Al_2O_3–water (1–4% particle volume fraction) and TiO_2–water (1–6% particle volume fraction) water nanofluids, as well as ZnO-EG at different volume fractions are investigated in this study.

The performance of the different nanofluid PVT plant is simulated through "TRNSYS" software. First, several limitations and constraints are needed to be resolved to develop a trustworthy simulation environment within the TRNSYS framework. In particular, the great sensitivity to particle volume fraction and temperature of the thermal conductivity of the nanofluid has to be taken into account. Thermal and electrical efficiency as well as the energy yields by changing the heat transfer fluid, considering variable weather conditions, are calculated. The achieved results showed the positive contribution of the implementation of nanofluid in PVT plant.

New Methods Developed for Precision Agriculture

Dr. Pierre Temple-Boyer
LAAS-CNRS, France

Biography

Pierre Temple-Boyer, born in 1966, received the Engineering degree in Electronics and Microelectronics from the French "Ecole Supérieure d'Electricité" (SUPELEC, France) in 1990 and his master degree in solid state physics from the University "Paul Sabatier" of Toulouse (UPS, France) in 1992. He joined the French "Laboratoire d'Analyse et d'Architecture des Systèmes" (LAAS) from the French "Centre National de la Recherche Scientifique" (CNRS) in 1992 and received the Ph.D. degree from the French "Institut National des Sciences Appliquées de Toulouse" (INSAT, France) in 1995. Since then, as a CNRS researcher at LAAS, he has been working on the integration of materials for microtechnological applications as well as on the study of physical, chemical and biological microsensors. He received the French "Habilitation à diriger des recherches" from the University "Paul Sabatier" of Toulouse (UPS, France) in 2004 and became the CNRS research director in 2011. In 2016, he became the deputy director of LAAS-CNRS, incharge of administrative affairs as well as institutional interactions with the section 08 of CNRS dedicated to micro-/nanotechnologies, micro-/nanosystems, photonics, electronics, electromagnetism and power engineering. His expertise field involves the technological integration and the study of bio-electro-chemical detection/transduction principles for the aqueous phase analysis at the microscale, aiming to health, environment and agri-business applications. His research interests concern the realization of bio-electro-chemical microsensors, the integration of pH-metry techniques, the detection of bio-electro-chemical species in real media, the monitoring of cell cultures and/or the analysis of single cells.

Abstract: New Methods Developed for Precision Agriculture
In order to deal with environmental impacts of intensive agriculture, new farming methods should be developed where fertilizers doses are determined over time at small scale according to soil variability. In this frame, the proposed work is dedicated to the development of a low-cost, robust and real-time analysis system based on electrochemical microsensors for the in-situ monitoring of soil nitrogen. pH-sensitive chemical field effect transistors (pH-ChemFET) were adapted to ion detection in liquid phase using polymer-based ion-sensitive membranes, focusing on the analysis of ammonium NH_4^+ and nitrate NO_3^- ions. The so-obtained ion-sensitive field effect transistors (ISFET) were confronted with real soil analysis. Our study showed, firstly, that pH-ChemFET microdevices are useable for soil measurement for a six-month period, dealing with soil texture and humidity, and, secondly, that ISFET microsensors are suitable for quick on-site analysis of nitrogen-based nutrients measured directly in soil. Thus, a pH/pNH4/pNO3 multi-ISFET sensor was finally fabricated and integrated in an autonomous, communicating device, including the power supply and data transmission unit. Thus, the operation of this device was evaluated through the monitoring of soil in simulated on-site conditions and real wheat fields. Overall, this work highlights the promising future of the ChemFET technology for the nitrogen cycle monitoring associated to plant metabolism as well as for the soil analysis in the frame of precision agriculture.

Image Blur Control Benefits to Visual Servo Control in Robotics

Dr. Guillaume Caron
University of Picardie Jules Verne, France

Biography

Guillaume Caron is Associate Professor since 2011 at Universite de Picardie Jules Verne (UPJV), France. He received the Ph.D. degree in Robotics and the Habilitation degree from the same university in 2010 and 2019. He was the leader of the Robotic Perception Group of the MIS Laboratory (UPJV) from 2016 to 2020. He has been CNRS (French National Scientific Research Center) delegate at the CNRS-AIST Joint Robotics Laboratory, IRL, Tsukuba, Japan, since 2019. Before, he stayed at Inria, Rennes, France (2010–2011), and at the University of Osaka, Japan (April–May 2013). His research interests include artificial vision for robotics, real-time visual tracking and servoing and digital heritage.

He has been Chair of the Technical Committee "Computer vision for cultural heritage applications" of the International Association for Pattern Recognition (IAPR) since 2018.

Abstract: Image Blur Control Benefits to Visual Servo Control in Robotics

Visual perception plays a key role toward the functional autonomy of robots. Among the several paradigms of robot vision, visual servoing defines the robot positioning by the automatic control of its degrees of freedom from visual information. The visual information of lowest level is pixel brightness that is successfully exploited by the direct visual servoing to control the robot maximizing the similarity between an image, acquired or rendered at the desired pose, and the video stream currently acquired by the camera onboard the robot.

This talk will show that considering blurred images and controlling the variation of blur permit direct visual servoing to reach accurately a distant desired pose. Investigations on optical blur and blur as a result of image processing for conventional and panoramic cameras demonstrate this assessment on robot arms, mobile robot and drone.

Lab-on-Disc for In-Situ Monitoring of Surface Water Quality by Algae Biosensors and Physicochemical Sensors

Dr. Jérôme Launay
LAAS-CNRS, France

Biography

Jérôme Launay, 42 years old, received the engineering degree in electronics and microelectronics from the French "Institut National des Sciences Appliquées de Toulouse" (INSAT, France) in 1998 and his master degree in microelectronic from the INSAT in 1998. He joined the French "Laboratoire d'Analyse et d'Architecture des Systèmes" (LAAS) from the French "Centre National de la RechercheScientifique" (CNRS) in 1998 and received the Ph.D. degree from the INSAT in 2001. Since 2006, as a lecturer at LAAS, he has been working on the integration of materials for microtechnological applications as well as on the study of physical, chemical and biological microsensors. His expertise field involves the technological integration and the study of bio-electro-chemical detection/transduction principles for the aqueous phase analysis at the microscale, aiming to health, environment and agri-business applications. His research interests concern the realization of bio-electro-chemical microsensors, the integration of pH-metry techniques, the detection of bio-electro-chemical species in real media, the monitoring of cells cultures and/or single cell analysis.

Abstract: Lab-on-Disc for In-Situ Monitoring of Surface Water Quality by Algae Biosensors and Physicochemical Sensors

To overcome the issue raised by the lack of available tools for monitoring water quality, from the source to the consumers, the BELUGA project aims to develop an easy-to-use dummy-proofed portable system allowing for a rapid in-situ diagnosis of surface water pollution. This low-cost demonstrator will benefit from the microfluidic technology in a "Lab-On-a-Disk" (LOD) format and will perform multi-parameter

analyses of complex matrices of pollutants. The LOD consists of several disks, each integrating one or more functions for the control of fluids, the detection of pollutants using microalgae-based biosensors as well as physico-chemical sensors (e.g., dissolved O_2, fluorescence, pH, temperature, nitrates, etc.). The proposed multi-sensors detection will improve the reliability of measurements in terms of qualitative response and will allow for obtaining a selective and effective test for surface water pollutants.

Fault Detection and Diagnosis Applied to Photovoltaic Power Plants

Dr. Anne Migan Dubois
Paris-Saclay University, France

Biography

She is Researcher at GeePs and Professor at U-PSay. She is the leader of the activities "Advanced characterizations in real conditions" and "soft integration of PV in smart-grid" at GeePs. She has supervised 13 Ph.D. thesis and several master thesis. She is an author/co-author of more than 100 papers and international communications and co-owner of 3 patents in PV. She is an expert for ANR evaluations, regular reviewer for *IEEE* journal (occasional for some other journals in the field of PV) and Editorial Board Member of the MDPI journal *Sustainability*. She teaches at the University of Paris-Saclay in energy conversion, electronics, optronics and instrumentation. She was responsible for two professional undergraduate degrees in Energy Efficiency in Buildings (creation) and in Maintenance.

Abstract: Fault Detection and Diagnosis Applied to Photovoltaic Power Plants

There is an increasing interest both in academic and in industry for health monitoring of photovoltaic (PV) power plants. The main reasons are safety issues and the loss

of income due to faults or failures. In a PV power plant, on the DC side, the fault can affect a single cell, a module or a string. The fault effect or signature can be detectable or not, depending on the available information, the fault severity and the fault diagnosis method. From the abundant literature, there is a diversity of approaches based on different input data (array I-V characteristic, array or string maximum power point, module-level power point, infrared images, etc.), different techniques (image processing, neural network, etc.) depending on fault types (mismatch, short-circuit, open-circuit, etc.). From the application point of view, it is not obvious to identify what would be the most efficient method to implement a condition-based maintenance that is now recognized as the most cost-effective method. Therefore, we propose in this work from the analysis of the publications in 2017 to classify the fault detection and diagnosis methods through a framework defined in four steps: modeling, pre-processing, features extraction and features analysis.

Contents

Electronics and Biomedical Applications

Renewable Energy

About the Editors

Hajji Bekkay is Professor and Director of the Renewable Energy, Embedded Systems and Information Processing Laboratory in the National School of Applied Sciences at the Mohammed First University. He received his DEA in microelectronics from the "Institut National des Sciences Appliquées of Toulouse" (INSAT, France) in 1996 and his Ph.D. from the Laboratory for Analysis and Architecture of Systems (LAAS-CNRS) Toulouse, France. He is author of two proceedings books and one handbook. His research activities focus on the development of biochemical micro-sensors and the application of artificial intelligence techniques in photovoltaic systems.

Adel Mellit is Professor at the Faculty of Sciences and Technology, Jijel University, Algeria. He received his MS and Ph.D. in electronics from the University of Sciences Technologies (USTHB), Algiers, in 2002 and 2006, respectively. His research interests focus on the application of artificial intelligence techniques in photovoltaic systems and micro-grids. Dr. Adel Mellit has authored or co-authored more than 170 papers in international peer-reviewed journals (mostly with Elsevier), papers in conference proceedings (mostly with the IEEE) mainly on photovoltaic systems, six book chapters, two proceedings books and one handbook. He is serving on the editorial board of the Renewable Energy and is Editor of the *IEEE Journal of Photovoltaic and of Energy* (Elsevier Ltd.).

Antonio Gagliano is Associate Professor of Building Physics at the University of Catania, Italy, where he is the Scientific Manager of the Laboratory of Applied and Building Physics of the Department of Electric, Electronic and Computer Engineering (DIEEI). He received his MS in 1991 and Ph.D. degree in Environmental Applied Physics in 1998. He participated in several national and international research projects. He is the author of more than 130 articles published in reputed international journals and conference proceedings and component of the editorial board of several international journals.

Abdelhamid Rabhi is Associate Professor at the Faculty of Sciences, University of Picardie Jules Verne of Amiens, France. He completed his Masters in robotics and control systems from the University of Versailles, and his Ph.D. in observation and control for nonlinear and complex systems in December 2005. His research deals with robust control, diagnosis and observation for complex systems with focus on areas such as vehicles, autonomous robots applications and renewable energy systems.

Mohammed Amine Koulali is Associate Professor at the National School of Applied Sciences, Mohammed First University. He received his Engineering Degree in Computer Science from Ecole Nationale des Sciences Appliquées d'Oujda in 2006 and Ph.D. degrees in Computer Science from the Université Mohammed V Souissi, Rabat. His research focuses on Computer Science and he is responsible of several scientific projects.

Communication, Networks and Information Technology

A Fuzzy Logic-Based Intrusion Detection System for WBAN Against Jamming Attacks

Asmae Bengag, Amina Bengag, Omar Moussaoui, and Blej Mohamed

Abstract Wireless Body Area Network (WBAN) is a set of special nodes called medical sensors. These sensors are very useful and helpful for making the user able to connect everywhere and every time. However, they suffer from many problems like the low computing capacity, energy and memory space. In terms of security, WBAN systems are threatened by various types of attacks due to the wireless communication. This technology must have a robust mechanism to detect attacks for making medical applications more reliable and safety. In our work, we have focused on identifying jamming attacks from WBAN using the fuzzy logic system (FLS) that is one of the powerful mechanisms of artificial intelligence (AI) to determine different network cases. The proposed system used one of the fuzzy inference methods named Mamdani model and based on three network parameters Packet delivery ratio (PDR), received strength signal indicator (RSSI) and energy consumption amount (ECA). Our intrusion detection system is simulated by using MATLAB 9.0 and Castalia for analyzing the output result as jamming detection index (JDI).

Keywords WBAN · Security · Intrusion detection system · Jamming · Fuzzy logic · Mamdani model · False alert

1 Introduction

Currently, WBAN becomes one of the important parts of our daily life that improves the quality of the health care and studies, including emergency medical and remote medical surveillance [1]. This technology is based on mini medical sensors that are attached in the human body in order to communicate with the medical center using wireless. In fact, the medical nodes transmit a sensitive data via wireless medium to the coordinator node or the personnel device assistant (PDA) using ZigBee (802.15.4)

A. Bengag (✉) · A. Bengag · O. Moussaoui · B. Mohamed
MATSI Research Laboratory, ESTO, Mohammed First University, Oujda, Morocco
e-mail: asmaebengag@gmail.com

or Bluetooth (802.15.1) [2]. After that, the data is transmitted to the medical center using mobile networks or internet. However, this communication is not reliable because it deployed in open radio frequencies and in various attacks [3, 4]. Jamming attack is one of the attacks that threats the availability of the network as a security aspect. This attack makes the legitimate nodes not able to send or receive any information by transmitting a high-range signal. More specifically, it disrupts the communication between the medical nodes and involves the collision between them, and it causes the energy consumption of the sensors.

The intrusion detection system (IDS) has become an essential solution security, for detecting an intrusion in a system. We can implement an IDS in two ways, the first one on specific device as host intrusion detection system and the second way as a controller for the network. Indeed, the IDS is one of the effective solutions to detect and monitor an intrusion in the network as jamming attacks. Nevertheless, there are various challenges used in detection techniques as the binary decision [5]. For instance, the IDS can classify normal activities as an intrusion that increases the false alerts. Therefore, the FLS is used to solve diverse kinds of problems and makes the IDS to take the good decisions with high detection attack and low false alert.

In this paper, we aim to develop a novel IDS based on the fuzzy logic using Mamdani inference mechanism, to detect jamming attacks in WBAN. The proposed solution uses three network parameters, namely PDR, RSSI and ECA that are implemented as crisp inputs in FLS, to identify the network state and detect jamming attacks. More specifically, the parameters values were valued by the fuzzy inference system (FIS) to calculate the level of jamming, which is called jamming detection index (JDI).

The rest of the paper is organized as follows: Sect. 2 gives an overview of the main components of the fuzzy logic system. In Sect. 3, we present briefly some previous mechanisms based on FLS that detect jamming attacks. After that, we describe the performance of our proposed technique by explaining why we used the three parameters and how we calculated the level of jamming attack via FLS. Finally, we conclude our paper and give some future works.

2 Related Work

There has been different mechanisms for studding the state of the network in the literature to ameliorate the security issues in wireless sensor network (WSN), in order to detect jamming attacks. In fact, various techniques have been proposed for WSN that could be implemented in WBAN. In this section, we give a set of previous techniques for jamming detection based on fuzzy logic.

In [6], fuzzy logic is applied for determining node's malicious level in WSN, by calculating two metrics packet delivery ratio (PDR) and packet loss ratio (PLR). The simulation results are evaluated in MATLAB 7 and NS2. The authors in [7] proposed a

system named fuzzy logic-based jamming detection algorithm (FLJDA) for detecting jamming. This mechanism applied the Mamdani model using two input parameters PDR and RSSI. Indeed, the cluster head is used to calculate these parameters to identify if the cluster member is jammed or not.

Vijayakumar et al. [8] developed two methods for detecting the presence of jamming attack in cluster-based WSN (CWSN), using two main parameters RSSI and PDR. The first technique is named fuzzy inference system-based jamming detection system, which consists to optimize the detection by applying Takagi–Sugeno fuzzy model. The second one uses learning ability named the adaptive neuro-fuzzy inference system, basing on existing dataset for the prediction of the future values to detect various types of jamming. Reyes et al. [9] proposed a mechanism to check a link loss based on fuzzy logic technique, using the following inputs: PDR, RSS, bad packet ratio (BPR) and clear channel assessment (CCA). This technique is used to identify the level of jamming index (low, medium or high) in WSN.

Angrishi et al. [10] suggested the fuzzy-based detection and prediction system mechanism that is implemented on IEEE 802.15.4 low-rate wireless personal area network (LR-WPAN). This technique used two crisp inputs signal to noise ratio (SNR) and packets dropped per terminal (PDPT), for predicting and detecting DoS that affects the availability of the network.

3 Overview of Fuzzy Logic System

In recent times, the FLS is used in different real-time applications such as intelligent personnel, medical service, temperature monitoring and digital image processing, in order to identify the superlative decision. This methodology is built by Lotfi A. Zadeh professor in 1965 [11], to handle the uncertainty and ambiguity made from human reasoning. Principally, the FLS involves four main elements as shown in Fig. 2, namely fuzzification, fuzzy rules, fuzzy inference system and defuzzification.

The crisp inputs are collected from network traffic as real values in the first part of the fuzzy system named "fuzzification", in order to transform them into fuzzy inputs set. Then, with help of fuzzy rules, the fuzzy inference consists to calculate the crisp output set from the fuzzified input via fuzzy logic [5]. Basically, the FIS involves several rules base in the form IF-THEN statement, to optimize the range by using membership function (MF) [11]. Finally, the defuzzification translates the output into crisp values. Indeed, the MF determines the membership value or the degree of truth of each input and output between 0 and 1. The most shapes used for the MF are as follow: bell curves, triangular and trapezoidal.

There are three models of FIS, namely Mamdani, Sugeno and Tsukamoto models. The fundamental difference among these three models is how fuzzy input generates the crisp output [11]. In general, Mamdani model is based on the Center of Gravity mechanism used in the defuzzfication process, to get the fuzzy output [12]. On other hand, the Sugeno and Tsukamto calculate the crisp output using the weighted average [11]. Therefore, our proposed IDS employs the Mamdani because it is the

most appropriate for our system in detecting jamming attacks in WBAN. The fuzzy inference rule of Mamdani model is written as in (*A*) [7], where *x*, *y* and *z* are fuzzy sets linguistic variables defined by fuzzy sets. The linguistic values are presented in *A* and *B*, whereas the output is defined in *C*.

$$\text{Rule} = \text{If } x \text{ is } A \text{ and } y \text{ is } B, \text{ then } z \text{ is } C(A)$$

4 Proposed Fuzzy Logic Detection Jamming

4.1 Description

The proposed intrusion detection system used the FIS with three main network parameters: PDR, RSSI and ECA as illustrated in Fig. 3. These parameters are used as the jamming attack metrics in order to evaluate the network state. The PDR is presented as the ratio of the number of packets successfully sent by the node to the total number of packets transmitted by the node [9]. The RSSI metric presents the power content of the received radio signal at the receptor [4]. Besides, the ECA is the amount energy consumed by the node in a specified time for a sensor node [13].

The main reason for choosing these parameters is they are changed depending to the normal and abnormal conditions of the medical sensor. Furthermore, these parameters are used to avoid as much as possible the cases resemble the jamming cases, called the false positive alerts. For instance, these last could be the problems related to the collision problem, low energy or imperfect connection. Therefore, The FIS uses these parameters as crisps inputs, defined with three fuzzy sets: low (L), medium (M) and high (H), whereas the MF for the output has four fuzzy sets as follows: very low (VL), low (L), medium (M) and high (H). The output is a Jamming Detection Index (JDI) that represents the probability of jamming level in the network. The JDI value varies from 0 to 100, indicating "low jamming" to "high jamming", respectively.

Actually, to fuzzify the inputs and identify the fuzzy MF, we used the trapezoidal membership functions with the Mamdani inference method for the rules base. Indeed, the trapezoid shape is chosen because it can be mathematically manipulated to be very close to the most natural function [11]. Figure 4 illustrates the combination of three trapezoidal functions for the PDR parameter.

The generation of fuzzy rules and the values of each MF are based various tests and on the instructions of expert knowledge [5]. Table 1 illustrates the values of each membership function.

Table 1 MFs of the input and output functions

Input variable	Fuzzy value/MF	A	B	C	D
PDR	Low	− 0.5	0	10	25
	Medium	15	32	38	55
	High	45	70	100	102
ECA	Low	− 0.5	0	20	30
	Medium	25	35	45	55
	High	50	60	100	102
RSSI	Low	− 0.5	0	5	10
	Medium	5	10	15	20
	High	15	20	100	102

4.2 Fuzzy Inference System

The fuzzy inference is the second step in the FL. To define the set of rules, we are using the Comb method to avoid combinatorial explosion [14]. In our case, there are three (3) linguistic variables with three (3) possible levels (high, medium and low), so to calculate the rules basing on the traditional fuzzy system, we have as a result 27 rules (3 to the power of 3), whereas, we can reduce this number to 9 rules (3 * 3) due to the Comb method. In our case, we selected the logical nine fuzzy rules logical, because there are some rules and results are unlikely to occur in a real situation.

Furthermore, it is not necessary to add all the possibilities, and we used 9 fuzzy rules as given below, in the fuzzy inference system basing on how jamming affects the network parameters in order to identify the degree of JDI. In fact, we eliminate some rules that do not have a value added to the system. For instance, in case the PDR is high (Rule 1), we can conclude that there is no jamming whatever the value of RSSI and ECA. In addition, if the RSSI is low (Rule 2), it means that the medical node is not jammed, whatever the value of PDR and ECA. The set of fuzzy rules is given as follows:

1. If PDR is high, then JDI is very low;
2. If RSSI is low, then JDI is low;
3. If PDR is low, RSSI is medium, and ECA is low, then JDI is medium;
4. If PDR is low, RSSI is high, and ECA is high, JDI is high;

The relationships obtained from the rule base are interpreted using the minimum operator "AND". The surface plot corresponding to membership functions of PDR and RSSI variables is given in Fig. 5.

The node with linguistic values low PDR, high RSSI and high energy has the highest probability to be attacked by the jamming attack, and the node with high PDR, low RSSI and low energy has the lowest probability to be under jammer node.

4.3   Test and Simulation Results

According the communication architecture of the WBAN, our work is focused on the first-level "inter-BAN communication". The values of three crisp inputs PDR, RSSI and ECA are generated through OMNET++ as network simulator using Castalia platform, which is useful for the WBAN. Indeed, we have simulated the WBAN in normal and abnormal cases for collecting the parameters values. In the normal scenario, we have simulated the WBAN with three medical nodes and coordinator node using the ZigBee as MAC protocol. While for studding the impacts of jamming attack, we used also the BypassMac that does not respect the MAC protocol mechanism, in order to simulate a jammer node as shown in Fig. 1.

After the different simulation, the parameters values of each node are studied in our proposed mechanism that was built in MATLAB. Due to the Fuzzy Logic Toolbox

Fig. 1 Communication between medical nodes and PDA under jamming attack [2]
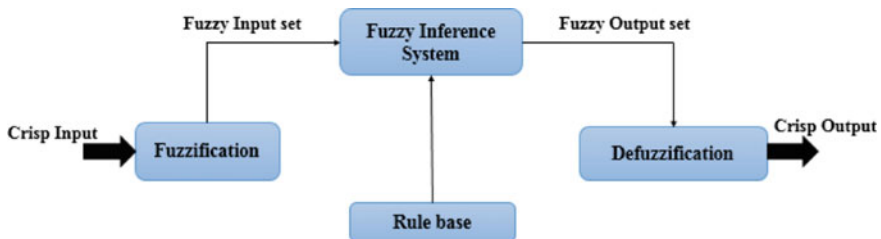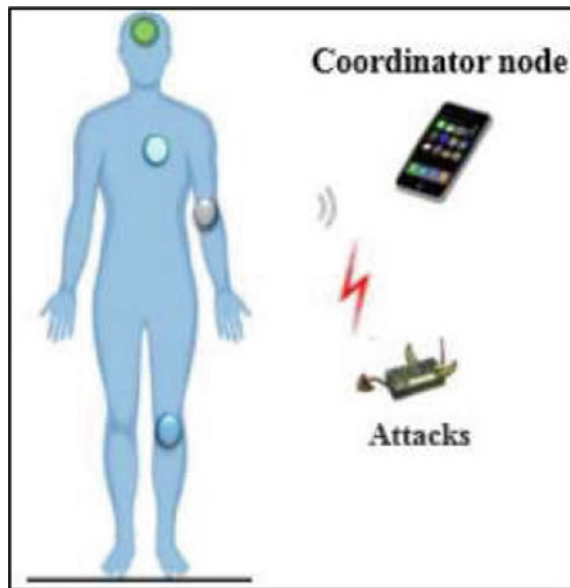
Fig. 2 General fuzzy system architecture
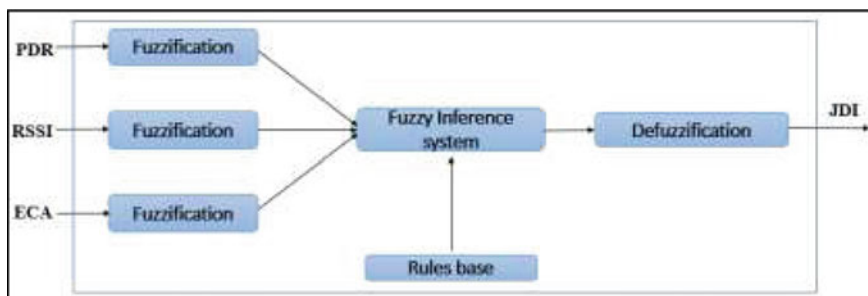
Fig. 3 FLS of our proposed system
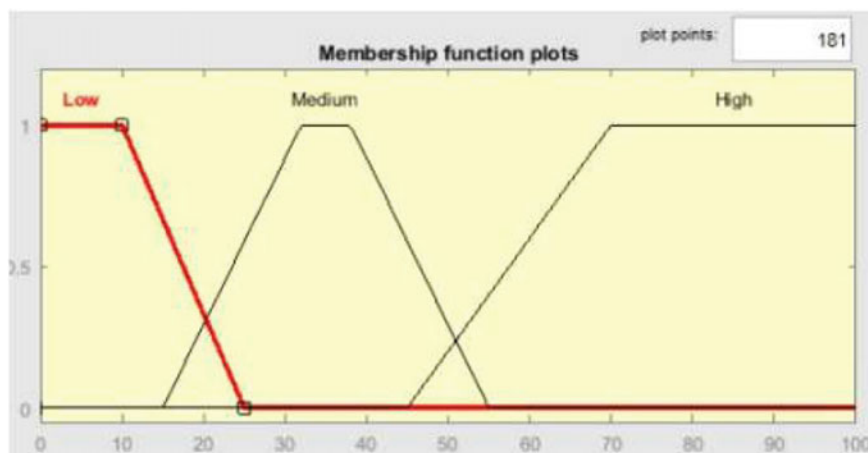
Fig. 4 Membership functions for the input PDR
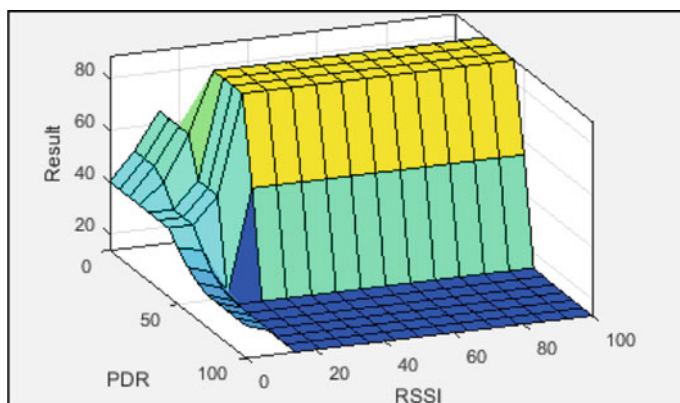
Fig. 5 Surface plots of PDR and RSSI

Table 2 Evaluated the traffic network using our proposed system

Inputs			Outputs		
PDR	RSSI	ECA	JDI value	JDI level	Decision
29.34	85	76	88.4763	High	Jammed: high-power jamming
11.438	65	57	65	High	Jammed
4.618	82	73	79.8829	High	Jammed
54.21	12.73	10	40	Low	Not jammed: possible collision case
90.033	22	5	40	Low	Not jammed
79.33	44.21	32	40	Low	Not jammed

of MATLAB, we defined the MF using the graphical interface. This Toolbox aims also to complete the fuzzification of the input variables and the fuzzy operator as AND or OR to define the rules base. According to the combination of the different input values RSSI, ECA and PDR, our system is able to control the communication system and indicates the level of Jamming Detection Index, as shown in Table 2. If the output has high as result, it means that the medical node is under jamming.

5 Conclusion

Jamming attacks are among attacks of DoS attacks. The goal of these attacks is to destroy the communication of the network by transmitting a successive signal. In this paper, we developed a new intrusion detection system that aims to control the network traffic, for detecting jamming attacks in WBAN system. Indeed, we have integrated the fuzzy logic approach into our IDS to take the good decisions with high detection attack and low false alert. The system is evaluated by using different scenarios simulated in Castalia. In our future research, we aim to implement our proposed system in coordinator node used as CH in order to calculate the parameters values of each medical sensors (cluster member).

References

1. Bengag A, Bengag A, Moussaoui O (2021) Effective and robust detection of jamming attacks for WBAN-based healthcare monitoring systems. In: Lecture notes in electrical engineering, vol 681, pp 169–174
2. Bengag A, Bengag A, Moussaoui O (2020) Attacks classification and a novel IDS for detecting jamming attack in WBAN. Adv Sci Technol Eng Syst 5(2):80–86
3. Bengag A, Bengag A, Moussaoui O (2021) Classification of security attacks in WBAN for medical healthcare. NISS 21:1–5
4. Bengag A, Moussaoui O, Moussaoui M (2019) A new IDS for detecting jamming attacks in WBAN. In: 2019 Third international conference on intelligent computing in data sciences (ICDS), pp 1–5

5. Almseidin M, Kovacs S (2018) Intrusion detection mechanism using fuzzy rule interpolation. J Theor Appl Inf Technol 96(16):5473–5488
6. Kanagasabapathy PMK (2019) Rapid jamming detection approach based on fuzzy in WSN. 1–14
7. Vijayakumar KP, Ganeshkumar P, Anandaraj M, Selvaraj K, Sivakumar P (2018) Fuzzy logic–based jamming detection algorithm for cluster-based wireless sensor network. Int J Commun Syst 31(10):1–21
8. Vijayakumar KP, Pradeep Mohan Kumar K, Kottilingam K, Karthick T, Vijayakumar P, Ganeshkumar P (2019) An adaptive neuro-fuzzy logic based jamming detection system in WSN. Soft Comput 23(8):2655–2667
9. Reyes HI, Kaabouch N (2013) Jamming and lost link detection in wireless networks with fuzzy logic. Int J Sci Eng Res 4(2):1–7
10. Angrishi K et al (2013) Fuzzy based detection and prediction of DDoS attacks in IEEE 802.15.4 low rate wireless personal area network. IEEE Netw 24(2):943–983
11. Chaudhary A, Tiwari VN, Kumar A (2014) Design an anomaly based fuzzy intrusion detection system for packet dropping attack in mobile ad hoc networks. In: Souvenir of the 2014 IEEE international advance computing conference (IACC), pp 256–261
12. Hiremath PS, Anuradha T, Pattan P (2017) Adaptive fuzzy inference system for detection and prevention of cooperative black hole attack in MANETs. In: Proceedings—2016 international conference on information science (ICIS), pp 245–251
13. Çakiroğlu M, Öozcerit AT (2011) Design and evaluation of a query-based jamming detection algorithm for wireless sensor networks. Turk J Electr Eng Comput Sci 19(1):1–19
14. Combs WE, Andrews JE (1998) Combinatorial rule explosion eliminated by a fuzzy rule configuration. IEEE Trans Fuzzy Syst 6(1):1–11

An Enhanced Approach Based on PCA and ACO Methods for Facial Features Optimization

Chaimaa Khoudda ⓘ **, El Miloud Smaili** ⓘ **, Salma Azzouzi** ⓘ **, and Moulay El Hassan Charaf** ⓘ

Abstract An automatic system for facial expression analysis consists generally of three main phases: detection, feature extraction and classification. In this study, we focus on the extraction of face characteristics (feature extraction) as well as the optimization of the obtained results. The objective is to reduce the number of facial features by removing noisy and redundant data in order to ensure an acceptable facial recognition accuracy while guaranteeing an optimal selection of distinctive facial information. For this purpose, we suggest a new approach based on the combination of principal component analysis (PCA) and ant colony algorithm (ACO). The study was conducted by exploiting the database of the Olivetti Research Laboratory (ORL).

Keywords Face recognition · Feature extraction · ACO · PCA · Optimization

1 Introduction

With the advent of the COVID-19 pandemic, facial recognition has gained popularity as a safe way to recognize people without the need of direct contact since facial image scan be acquired remotely by a camera. Therefore, this can be particularly beneficial for health, security and surveillance purposes. In this context, various methods have been proposed to achieve high accuracy in face recognition.

C. Khoudda (✉) · E. M. Smaili · S. Azzouzi · M. E. H. Charaf
Laboratory of Research in Computer Science, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco
e-mail: chaimaa.khoudda@uit.ac.ma

E. M. Smaili
e-mail: miloud.smaili@uit.ac.ma

S. Azzouzi
e-mail: salma.azzouzi@uit.ac.ma

M. E. H. Charaf
e-mail: my.charaf@uit.ac.ma

The complete process of face recognition comprises three steps: face detection, feature extraction and recognition. Basically, feature extraction maintains acceptable classification accuracy by reducing the maximum number of irrelevant features. The method is widely used in data analysis, pattern classification, multimedia information retrieval, machine learning, medical data processing and data mining applications.

Within the field of facial recognition, the extraction function remains an extremely important step that usually involves the application of an extraction algorithm on digital images to reduce redundancy and irrelevancies in images. Therefore, these algorithms transform an image (input data) into a set of features, while selecting the features containing the most relevant information from the original data.

On the other side, dimensionality has been considerably increased in recent years. As a result, many challenges arise for most of learning and optimization algorithms including storage requirements and high computational cost. In this context, feature optimization is considered as one of the most effective and efficient techniques to prepare high-dimensional data for optimal use in machine learning. The aim is to ensure accurate results by allowing the selection of an optimal subset of features from the original feature set. Therefore, feature optimization retains the effective significance of each selected feature, making it more suitable in terms of readability and interpretation. The choice of the ACO algorithm for feature optimization has many advantages. In fact, the positive feedback in dynamic applications leads to the rapid discovery of good solutions. In addition, the inherent parallelism lets avoid premature convergence as there is no central control in the system. As per the PCA method, it is very commonly used to improve visualization by transforming high-dimensional data into low-dimensional data but also to speed up the machine learning algorithm by removing correlated variables that do not contribute to the decision making. However, PCA will not be able to find the optimal principal components. Therefore, the combination of both PCA and ACO approaches is used in order to obtain more accurate and optimal results.

The paper is structured as follows: Sect. 2 introduces some preliminaries. The problematic statement is explained in Sect. 3. Then, we present in Sect. 4 some related works. Afterward, Sect. 5 explains our solution by giving an overview of the proposed architecture to optimize facial by combining PCA and ACO methods. Finally, Sect. 6 gives a conclusion and identifies future work.

2 Preliminaries

2.1 Principal Component Analysis (PCA)

Several techniques are used for extraction of face characteristics (feature extraction) such as discrete cosine transform (DCT), discrete wavelet transform (DWT) and principal component analysis (PCA). PCA is a common feature extraction method in data science. The PCA method converts a matrix of n features into a new dataset

of less than *n* features. That is, it reduces the number of features by constructing new smaller number variables which capture a significant portion of the information found in the original features [1].

2.2 Swarm Intelligence

Swarm intelligence refers to the collective behavior of decentralized, self-organized systems, natural or artificial. The main idea is to imitate the social behavior of insects (bird flocking, fish schooling, etc.). A famous example of swarm intelligence is the ant colony optimization (ACO) method working on the basis of ants' behavior looking for an optimal way between their state and a source of nourishment [2].

3 Problematic Statement

Facial recognition relies on an accurate feature extraction method that includes dimensionality reduction. The aim is to ensure optimal and accurate results but also to guarantee a reliable and efficient system.

In fact, the success of any methodology depends particularly on the choice of a powerful algorithm that can extract facial features and ensure high accurate results. For this purpose, we suggest in this paper to handle data using two approaches in order to extract distinctive facial information efficiently.

Therefore, we firstly suggest a method called principal components analysis (PCA) to perform features extraction. The PCA method is still an appropriate way that transforms the face into a small series of essential features using the eigenface technique. Then, we present an algorithm called ant colony optimization (ACO) to optimize the extracted coefficients.

4 Related Work

Many recent researches are devoted to ensure an acceptable facial recognition accuracy while guaranteeing an optimal selection of distinctive facial information. Indeed, the authors in [3] suggest an improved feature selection algorithm based on ant colony optimization whereas the work [4] presents an optimized method for extracting a selection of distinctive facial features using the ant colony optimization algorithm and its applications for pattern recognition in spatial imagery.

Another work [5] proposes a new text feature selection method that uses a wrapper approach, integrated with ant colony optimization (ACO) to guide the feature selection process. They also suggest using the k-nearest neighbor (KNN) as a classifier in order to evaluate and generate a candidate subset of optimum features.

In addition, the paper [6] aims to improve are cognition system based on Gabor features using the ant colony optimization algorithm. The evaluation of the proposed system was achieved using two image datasets; Olivetti Research Laboratory (ORL) database and African Face Image Database (AFI). Furthermore, the authors in [7] propose a system for grammatical recognition of facial expressions based on a hybrid of fuzzy rough ant colony optimization and nearest neighbor classifier.

A practical approach is presented in [8] where the authors develop a hybrid filter model for feature selection based on both principal component analysis and information gain. The model is then applied to support classification using machine learning techniques. The objective of the paper [9] is to describe a face detection and recognition system based on hybrid statistics and machine learning. The proposed method is based particularly on principal component analysis (PCA), support vector machine (SVM), K-nearest neighbor (KNN) and the ACO algorithm.

Finally, the paper enhances a previous work [10] where we suggest combining the PCA method with another optimization algorithm called particle swarm optimization to tackle the feature optimization issue.

5 Face Recognition Model

5.1 Architecture

This paper focuses on the optimization of facial features extraction in order to obtain optimal and accurate results. Figure 1 gives an overview of our facial recognition architecture.
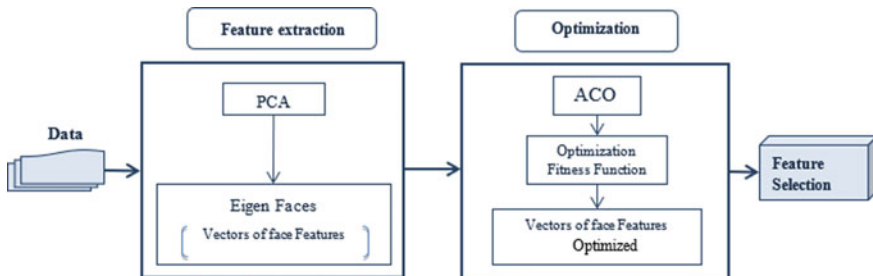
Fig. 1 Feature selection process: architecture

5.2 Model Description

Data Collection: In this study, the ORL database of faces is used. The database contains 400 face images taken at the AT&T Laboratories (by the Oliver Research Laboratory in Cambridge, UK) and corresponding to 40 distinct persons. The original size of each image is 92×112 pixels, with 256 gray levels per pixel. Each subject has ten different images taken in various sessions varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). For each person, four images out of ten are used for learning phase and the remaining are used for the testing phase.

Feature Extraction: We use features extraction to reduce the input data complexity and ensure a simple representation of data by representing each variable in the feature space as a linear combination of the original input variables. In this paper, we used a simple nonparametric method called: principal component analysis (PCA) to extract the most relevant information from a redundant or noisy dataset. The PCA algorithm is applied by considering 160 images corresponding to the learning phase as follows:

1. Convert each image into a vector of N^2 size.
2. Get a vector space (face space).
3. Standardize vector space by removing all common characteristics across images to have unique characteristics for each image.
4. Determine the common characteristics using the average face vector calculation method in order to obtain standardized vectors.
5. Calculate eigenvectors (eigenfaces) using computation of the covariance matrix.

Feature Optimization: The objective is to optimize the eigenvectors previously obtained by the application of the PCA algorithm. For this purpose, we suggest to extract through eigenvectors the most optimal features using the ACO algorithm.

The ACO algorithm requires that the problem be represented as a graph whose nodes represent the different features extracted from the face (eigenfaces). In our case, the search for the optimal subset of features corresponds to a search through the graph where a minimum number of nodes is visited while satisfying the stopping criterion. In fact, we start moving iteratively from one node to another as described below:

```
Begin
      Initialize
While stopping criterion not satisfied do
      Position each ant in a starting node
Repeat
For each ant do
    Choose next node by applying state transition rule
    Apply step by step pheromone update
```

```
End for
Until every ant has built a solution
      Update best solution
      Apply off line pheromone update
End while
End
```

The probability of a transition from node i to node j at time t is given by

$$p_{i,j}^k(t) = \frac{\left[\tau_{i,j}(t)\right]^\alpha \left[\eta_{i,j}(t)\right]^\beta}{\sum\limits_{l \in N_i^k} \left[\tau_{i,1}(t)\right]^\alpha \left[\eta_{i,1}(t)\right]^\beta} \tag{1}$$

where

- N_i^k are the feasible neighborhoods of node i for ant_k
- $\tau_{i,j}(t)$ is the pheromone value on the edge (i,j) at the time t
- α is the weight of pheromone
- $\eta_{i,j}(t)$ is a priori available heuristic information on the edge (i,j) at the time t
- β is the weight of heuristic information

The pheromone $\tau_{i,j}(t)$ is updated according to following equation:

$$\tau_{i,j}(t) = \rho.\tau_{i,j}(t-1) + \sum_{k=1}^{n} \Delta\tau_{i,j}^k(t) \tag{2}$$

$\Delta\tau_{i,j}^k(t)$ is the pheromone quantity left at the moment t by ant on arc (i,j)

$$\Delta\tau_{i,j}^k(t) = \begin{cases} \frac{Q}{L_k(t)}, & \text{if arc}(i, j) \text{ is chosen by ant}_k \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where

- ρ is the pheromone trail evaporation rate $(0 < \rho < 1)$
- n is the number of ants
- Q is a constant for pheromone updating
- $L_k(t)$ represents the length of the path chosen by the ant_k

Finally, we refer the work [11] in order to get a good convergence rate. In this case, the initial parameters for our ACO algorithm are set as follows (Table 1):

Table 1 ACO algorithm settings

Parameters	Values
The number of iterations	100
The number of ants in each iteration	50
The pheromone trail evaporation rate ρ	0.2
The weight of pheromone α	1
The weight of heuristic information β	0.1

6 Conclusion

In this paper, we suggest to optimize the facial features extraction by combining two approaches: principal component analysis (PCA) to extract facial features and ant colony optimization (ACO) to obtain optimized results.

To this end, we apply the PCA algorithm by considering 160 images corresponding to the learning phase from the ORL database. Afterward, we extract through the eigenvectors previously obtained, the most optimal features using the ACO algorithm.

As perspectives, we plan to implement these algorithms on other face databases in order to compare the results obtained with other algorithms. We also aim to improve the quality of these approaches in terms of performance and efficiency, especially for colored images.

References

1. Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. Philos Trans Royal Soc A: Math Phys Eng Sci 374(2065):20150202
2. Nayar N, Gautam S, Singh P, Mehta G (2021) Ant colony optimization: a review of literature and application in feature selection. In: Inventive computation and information technologies, pp 285–297
3. Peng H, Ying C, Tan S, Hu B, Sun Z (2018) An improved feature selection algorithm based on ant colony optimization. IEEE Access 6:69203–69209
4. Wang M, Wan Y, Ye Z, Lai X (2017) Remote sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm. Inf Sci 402:50–68
5. Ahmad SR, Bakar AA, Yaakub MR (2019) Ant colony optimization for text feature selection in sentiment analysis. Intell Data Anal 23(1):133–158
6. Aro T, Abikoye O, Oladipo I, Awotunde B (2019) Enhanced Gabor features based facial recognition using ant colony optimization algorithm. J Sustain Technol 10(1)
7. Gafar MG (2019) Grammatical facial expression recognition basing on a hybrid of fuzzy rough ant colony optimization and nearest neighbor classifier. In: 2019 International conference on innovative trends in computer engineering (ITCE). IEEE, pp 136–141
8. Omuya EO, Okeyo GO, Kimwele MW (2021) Feature selection for classification using principal component analysis and information gain. Expert Syst Appl 174:114765
9. Vinodini R, Karnan M (2022) Face detection and recognition system based on hybrid statistical, machine learning and nature-based computing. Int J Biometrics 14(1):3–19

10. Khoudda C, Smaili EM, Azzouzi S, Charaf MEH (2020) An optimized method for face recognition using PCA and PSO techniques. In: 3rd Edition of the international conference advanced intelligent systems for sustainable development AI2SD'2020
11. Kanan HR, Faez K, Hosseinzadeh M (2007) Face recognition system using ant colony optimization-based selected features. In: 2007 IEEE symposium on computational intelligence in security and defense applications. IEEE, pp 57–62

The GPSR Routing Protocol in VANETs: Improvements and Analysis

Amina Bengag, Asmae Bengag, and Mohamed Elboukhari

Abstract One of the most critical problems in VANETs is the frequent link breakage caused by the high velocity of vehicles. Due to the short connection lifetime between vehicles, the communication paths are frequently interrupted during the transmission of data packets between the source and destination vehicles, causing the search for a new route that increases the routing overhead and diminishes the PDR and the throughput. To manage those issues, several routing protocols have been proposed by considering important factors to improve the quality of service in VANETs. The GPSR (Karp and Kung in ACM MobiCom, pp. 243–254, 2000) for Greedy Perimeter Stateless Routing is the most popular position-based protocol. In this paper, we propose three new models to enhance this protocol that guide the selection of the next-hop vehicle based on some important metrics of the participating nodes. We have used a real urban scenario to evaluate the performance of our models, by varying the vehicle density and measuring the percentage of packet delivery ratio, throughput and routing overhead during the transmission of data packets.

1 Introduction

VANETs are considered as a special case of Mobile Ad-Hoc Networks (MANETs); they have many characteristics compared to other class of MANETs. Indeed, due to the high speed of vehicles the network's topology changes frequently, which affect the connectivity between vehicles that changes regularly. Due to those characteristics, designing an efficient routing protocol to route packets to their final destination is a big challenge.

A. Bengag (✉) · A. Bengag · M. Elboukhari
MATSI Laboratory, ESTO, Mohammed First University, Oujda, Morocco
e-mail: bengag.amina@gmail.com

As mentioned and detailed in [1, 2], routing protocols in VANETs are classified mainly into four types. They are position-based [3], topology-based [4], multicast-based [5] and broadcast-based routing protocols [6]. Researchers believe that position-based routing protocols are the best one in terms of PDR, routing overhead, throughput and scalability [7, 8]. Therefore, we will focus in our work on GPSR routing protocol that is a position-based routing.

In this paper, we suggest three innovative position-based protocols based on GPSR and some important mobility parameters used to improve the classical GPSR in terms of PDR, throughput and routing overhead in VANET scenarios; the proposed protocols are called Density-Velocity-Aware-GPSR (DVA-GPSR), DRL-GPSR for Direction-Route Lifetime aware GPSR and Enhanced GPSR (E-GPSR). To select the next-hop vehicle, the three proposed protocols take into consideration some important factors:

- The speed variation between the target and the next-hop candidate vehicle.
- The vehicles' direction utilized to calculate the angle direction between the destination's direction and the next-hop candidate's direction.
- The lifetime of the route, that is between the current node and its neighbors.
- The density of the next-hop candidate; this metric is utilized to identify the connectivity mode for each path (sparse, medium or dense).
- The distance between the current node and the destination.

The remaining of this paper is organized as follows: in Sect. 2, we describe the classical GPSR protocol. In Sect. 3, we clarify the approach of our proposed protocols. In Sect. 5, we give and evaluate the efficiency of the three proposed protocols compared to the classical GPSR. In the last section, we present the conclusion of this paper.

2   An Overview of the Classical GPSR

GPSR [9] is a routing protocol originally designed for MANET and rapidly adapted to VANET. In this section, we will give a general review on GPSR routing protocol that belongs to the class of position-based protocols. This GPSR protocol uses two approaches to forward packets. The greedy forwarding approach is used to forward packets to the closest neighbor to the destination. When this approach fails, the perimeter-forwarding mode will be applied that is based on the right-hand rule strategy. Those techniques get good results in MANETs. However, in the case of city scenario, the GPSR still suffers from some disadvantages that reduce its performances.

3 The Proposed Enhancements of GPSR

Our three proposed protocols are based on the classical GPSR. Each enhancement necessitates two components: a GPS implemented in all vehicles in order to give the exact vehicle's position and an On-Board Unit (OBU) equipment used to connect the vehicles. Each proposition consists of many factors, and they are discussed below.

3.1 E-GPSR

The strategy used by E-GPSR is based on three metrics: the speed variation, the distance between the transmitter vehicle and all its neighbors and the density of the current node's neighbors according to formula (1) and (2), respectively. After that, we use the Eq. (3) to calculate the link weight function for this strategy.

In the Eq. (3) the d_i is the density of the next-hop candidate i; the use of this metric reduces the bad influence of sparse connectivity problem. The node that has more neighbors increases the probability of being selected as the next hop. Therefore, we take into account the inverse of di ($1/d_i$) in LWF1.

$$S_{id} = |S_i - S_d| \tag{1}$$

where S_i and S_d signify, respectively, the speed of the neighbor vehicle called i and the speed of the destination vehicle.

$$D_{id} = \sqrt{(y_i - y_d)^2 + (x_i - x_d)^2} \tag{2}$$

where (x_i, y_i) signifies the neighbors' location called i and (x_d, y_d) denotes the location of the destination vehicle.

$$LWF1 = \alpha * D_{id} + \beta * \left(\frac{1}{d_i}\right) + \theta * S_{id} \tag{3}$$

where $\alpha + \beta + \theta = 1$.

3.2 DVA-GPSR

The strategy adopted by DVA-GPSR is based on four factors. The new metric is the angle direction between the transmitter vehicle and the destination vehicle. In fact, if two vehicles communicate with each other and have the same direction, the link is more stable. To calculate the angle direction, we use the Eq. (8) and the Eq. (9) will be used to calculate the link weight function (LWF), where $\alpha + \beta + \theta + \gamma = 1$.

In Eqs. (4), (5), (6) and (7), iV and dV are, respectively, the velocity of the next-hop candidate and the velocity of the destination.

$$a = \text{iV} \cdot x * \text{dV} \cdot x \tag{4}$$

$$b = \text{iV} \cdot x * \text{dV}.y \tag{5}$$

$$c = \text{iV} \cdot x^2 * \text{dV} \cdot x^2 \tag{6}$$

$$d = \text{iV} \cdot y^2 * \text{dV} \cdot y^2 \tag{7}$$

$$\varphi_{\text{id}} = \cos^{-1}\left(\frac{(a+b)}{\left(\sqrt{c} * \sqrt{d}\right)}\right) \tag{8}$$

$$\text{LWF}_2 = \text{LWF}_1 + \gamma * \varphi_{\text{id}} \tag{9}$$

In case of E-GPSR and DVA-GPSR, the source vehicle calculated the LWF for all its neighbors. Then, the node that has the smaller value of the LWF will be chosen as a next hop. The DVA-GPSR is already simulated under a highway scenario and gives good results in terms of PDR, throughput and routing overhead [10].

3.3 DRL-GPSR

In DRL-GPSR (Direction-Route Lifetime aware GPSR), we use a new metric called route lifetime in addition to the angle direction metric to select the next-hop node. When a vehicle wants to send a packet to a destination node, it first calculates how long each of its neighbors could communicate with the current node, then compares the results and chooses the longer route lifetime; the current vehicle also needs to calculate the angle between its neighbors and the destination node and then chooses the smaller one.

The duration that two nodes will remain neighbors can be predicted as follows: Let two nodes i and j be within the transmission range of each other. Let (x_i, y_i) and (x_j, y_j) be the coordinates of the vehicles i and j, respectively. Let v_i and v_j be the velocities and φ_i and φ_j, where $(0 \leq \varphi_i, \varphi_j < 2\pi)$, indicate the direction of the vehicles i and j, respectively. RL_{ij} is the interval of time the two nodes j and i will stay connected. To calculate RL_{ij}, we use the following equation:

$$\text{RL}_{ij} = \frac{-(ab + cd) + \sqrt{(a^2 + c^2)r^2 - (ad - bc)^2}}{a^2 + c^2} \tag{10}$$

where

$a = v_i \cos \varphi_i - v_j \cos \varphi_j$; $b = x_i - x_j$; $c = v_i \sin \varphi_i - v_j \sin \varphi_j$; and $d = y_i - y_j$.

At the last stage, we calculate the LWF for this strategy according to Eq. (11). We have calculated the route lifetime between the current node and its neighbors, and the angle direction between the destination and the neighbors of the current node.

$$\text{LWF}_3 = \alpha * 1 / \text{RL}_{ij} + \beta * \varphi_{id} \tag{11}$$

where $\alpha + \beta = 1$.

4 Simulation and Results Analysis

In the simulation, the NS3 and SUMO have been used as a network and traffic simulator. Moreover, the simulation was done based on a real map of Oujda city, taken from the Web site of Open Street Map. The simulation settings are presented in Table 1.

4.1 PDR

Figure 1 shows the PDR of our proposed protocols by varying the density of vehicles. The DVA-GPSR and DRL-GPSR protocols show better performance than GPSR and have the highest value augmented up to 31.2% by increasing the number of vehicles. However, the E-GPSR and GPSR have the lowest values decreased down to 24%.

Table 1 Simulation parameters

Parameters	Values
Routing protocols	GPSR, DVA-GPSR, E-GPSR, DRL-GPSR
Number of destination	10
Number of vehicles	30, 50, 70, 90
Vehicle speed	Max: 20 m/s
Transmission range	145 m
packet size	512 bytes
Data type	CBR
Simulation time	200 s
Mac layer type	IEEE 802.11p
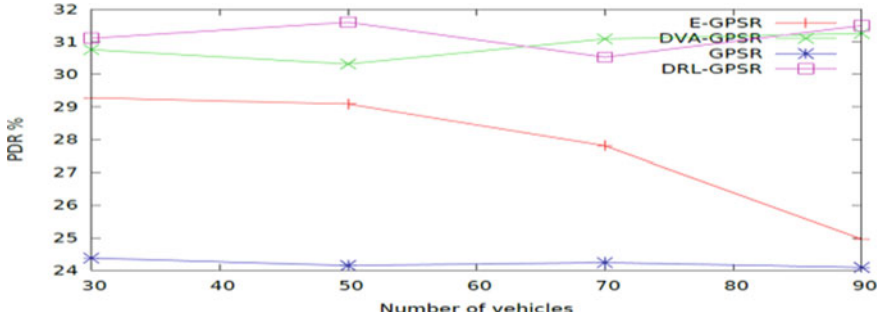
Fig. 1 Values of PDR by varying the number of nodes

4.2 Throughput

The graphs in Fig. 2 prove that our protocols have the best performance compared to the classical GPSR in terms of throughput by varying the density of vehicles in the route. Indeed, the throughput in case of DVA-GPSR and DRL-GPSR is increased up to 6.38 kbps thanks to the use of the mentioned factors in the previous section, which decrease the issue of connection loss so the lifetime of routes is increased.
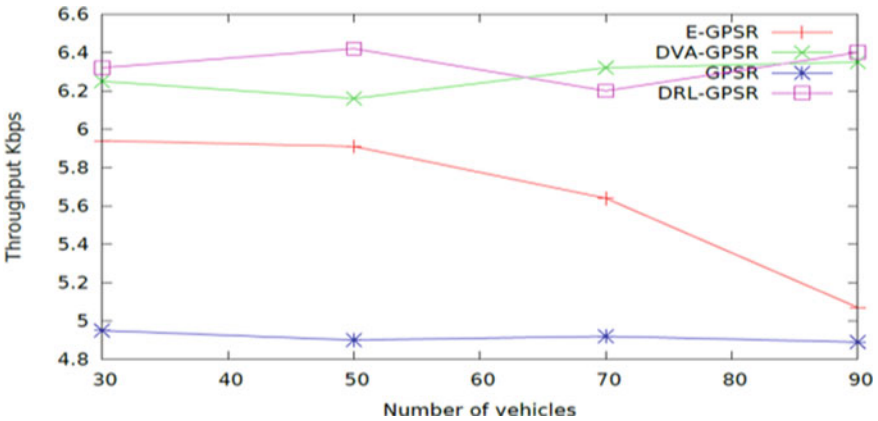
Fig. 2 Throughput by varying the number of nodes
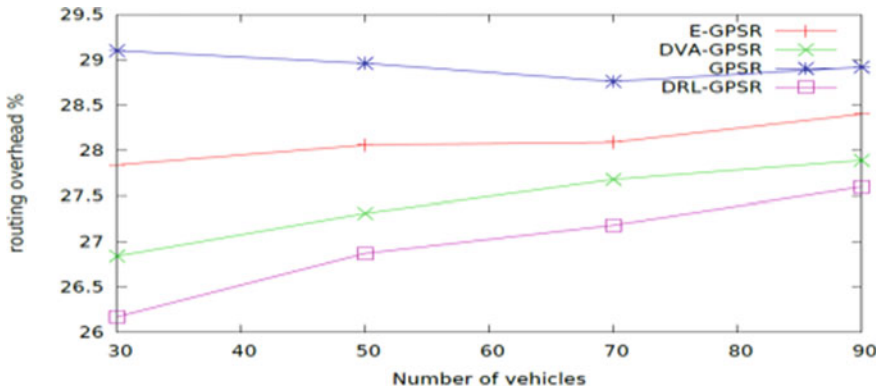
Fig. 3 Routing overhead by varying the number of nodes

4.3 Routing Overhead

The graphs in Fig. 3 present the routing overhead for the proposed routing protocols compared to the classical GPSR. Indeed, the three new protocols prove better performance and generate low routing overhead than the traditional GPSR that achieves 29%.

5 Conclusion

In this paper, we have suggested three innovative techniques to enhance the traditional GPSR protocol to be more suitable and convenient to vehicular networks. We have used SUMO and NS3 as traffic simulator and network simulator to demonstrate the quality of the proposed routing protocols, and we have used a real urban environment city that is a part of Oujda (street of El-Quds). The results of the simulation demonstrate that our protocols outperform the traditional GPSR protocol in terms of packet delivery ratio, the routing overhead and the throughput.

In E-GPSR, the strategy is based on three metrics: the distance between the target and the source vehicle, the density of the neighbors of the current vehicle that reduces the issues of void area and the speed variation between the target and the relaying candidate vehicle that reduces the effects of the high speed of vehicles that lead to the high dynamic topology and the connection damage problem. However, the connection could break quickly if two vehicles have the small variation speed with two different direction that prove the importance of including the direction in the second protocol.

In DVA-GPSR protocol, the problems caused by the frequent link breakage will be reduced compared to the standard GPSR and E-GPSR. Actually, in this protocol the angle direction metric is the new and the most important metric used for selecting a group of vehicles that move toward the target vehicle.

In DRL-GPSR, we use a new metric called route lifetime in addition to the angle direction metric to select the next-hop node. To conclude this paper and based on the discussion mentioned above, we confirm that the angle direction and the route lifetime parameters are the most important factors used to enhance the performance of the standard GPSR in urban environment in VANETs.

As a future work, we aim to simulate the three proposed strategies in other complicated urban scenarios and highway environment.

References

1. Bengag A, Bengag A, Elboukhari M (2020) Routing protocols for VANETs: a taxonomy, evaluation and analysis. Adv Sci Technol Eng Syst
2. Bengag A, El Boukhari M (2018) Classification and comparison of routing protocols in VANETs. In: 2018 International conference on intelligent systems and computer vision (ISCV), pp 1–8
3. Bilal SM, Bernardos CJ, Guerrero C (2013) Position-based routing in vehicular networks: a survey. J Netw Comput Appl
4. Bayad K, Bourhim EH, Rziza M, Oumsis M (2016) Comparative study of topology-based routing protocols in vehicular ad hoc network using IEEE 802.11p. In: 2016 International conference on electrical and information technologies (ICEIT), pp 526–530
5. Bachir A, Benslimane A (2003) A multicast protocol in ad hoc networks inter-vehicle geocast. In: IEEE semiannual vehicular technology conference, vol 4, pp 2456–2460
6. Lin Y, Chen Y, Lee S (2010) Routing protocols in vehicular ad hoc networks a survey and future perspectives. J Inf Sci 932(c):913–932
7. Amina B, Mohamed E (2018) Performance evaluation of VANETs routing protocols using SUMO and NS3. In: Colloquium in information science and technology, CIST, vol 2018-Oct, pp 525–530
8. Husain A, Raw RS, Kumar B, Doegar A (2011) Performance comparison of topology and position based routing protocols in vehicular network environments. Int J Wirel Mob Netw 3(4):289–303
9. Karp B, Kung H (2000) GPSR: Greedy perimeter stateless routing for wireless networks. In: ACM MobiCom, pp 243–254
10. Bengag A, Bengag A, Elboukhari M (2020) A novel greedy forwarding mechanism based on density, speed and direction parameters for vanets. Int J Interact Mob Technol

The Dynamics of a Population of Healthy Adults, Overweight/Obese and Diabetics With and Without Complications in Morocco

E. N. Mohamed Lamlili, Wiam Boutayeb, and Abdesslam Boutayeb

Abstract Between 1980 and 2014, the global prevalence of diabetes in adult people has nearly doubled, while the number of adult people living with diabetes rose from 108 to 422 million. In Morocco, from 2000 to 2018, obesity and diabetes increased, respectively, from 13.6 to 20% and from 6.6 to 10.6%. Half of the the adult population in Morocco are overweight. In this paper, a compartmental model is proposed to describe and analyze the evolution dynamics of an adult population from different stages (P: no overweight and no diabetes, W: overweight/obesity, D: diabetes without complications and C: diabetes with complications). Asymptotical local stability is proved, and normal foreword sensitivity index is used to guide decision makers in adopting a pragmatic strategy to control the prevalence of diabetes by acting on the flows into and out of the population living with overweight/obesity.

1 Introduction

During the last four decades, diabetes and overweight have substantially increased all over the world. According to the World Health Organization (WHO) report 2016 on diabetes, the prevalence of diabetes in adult people has nearly doubled between 1980 (4.7%) and 2014 (8.5%), while the number of adult people living with diabetes rose from 108 to 422 million during the same period. The highest prevalence of diabetes (13.7%) and the greatest relative increase (132%) were seen in the WHO Eastern Mediterranean Region (Table 1) [9].

In 2021, the International Federation Diabetes (IDF) estimated that 537 million people were living with diabetes and the number is projected to reach 643 million by

E. N. Mohamed Lamlili (✉) · A. Boutayeb
Faculty of Sciences, University Mohammed First, Oujda, Morocco
e-mail: mohamed.lamlili@gmail.com

W. Boutayeb
High School of Education and Formation (ESEF), University Mohammed First, Oujda, Morocco

Table 1 Prevalence of diabetes and number of adults (18 years or +) with diabetes by regions [9]

WHO region	Prevalence of diabetes		Number in millions	
	1980	2014	1980	2014
African Region	3.1%	7.1%	4	25
Region of the Americas	5.0%	8.3%	18	62
Eastern Mediterranean Region	5.9%	13.7%	6	43
European Region	5.3%	7.3%	33	64
Southeast Asia Region	4.1%	8.6%	17	96
Western Pacific Region	4.4%	8.4%	29	131
Total	4.7%	8.5%	108	422

2030 [4]. Delayed diagnosis and bad control of diabetes lead to complications like blindness, kidney failure, heart attacks, stroke and lower limb amputation. In 2019, diabetes caused 1.5 million deaths of which nearly half (48%) occurred before the age of 70 years. Having reached alarming levels, this disease has become a major health issue challenging health decision makers.

Similarly, around 650 million adults were suffering from obesity worldwide in 2016, representing a global prevalence of 13%. Overweight is strongly linked to diabetes and obesity is the leading risk factor for type 2 diabetes, increasing the risk of developing diabetes in women by 28 times if the body mass index (BMI) is 30kg/m^2 and by 93 times if the BMI is 35kg/m^2 [1].

In Morocco, in 2018, prevalence of overweight (BMI $> 25 \text{kg/m}^2$) and obesity (BMI $> 30 \text{kg/m}^2$) in adult population was, respectively, 53% and 20%. Women (29%) were nearly 3 times more likely to be obese than men (11%). Prevalence of diabetes was 10.6% but women (12.6%) were more affected than men (8.6%) [8]. In less than two decades (From 2000 to 2018), obesity and diabetes increased, respectively, from 13.6% to 20% and from 6.6% to 10.6% [7, 8].

Using probabilistic evolution from one stage of diabetes to another rather than the usual way of transmitted diseases, many authors have used compartmental models to describe and analyze the dynamic evolution of populations suffering from pre-diabetes, diabetes without complications and diabetes with complications [2, 3, 5, 6]. The model proposed in this paper focusses on the link between diabetes and overweight/obesity. The main objective is to show how strategies acting on the control of populations suffering from overweight/obesity can lead to the control of diabetes with and without complications.
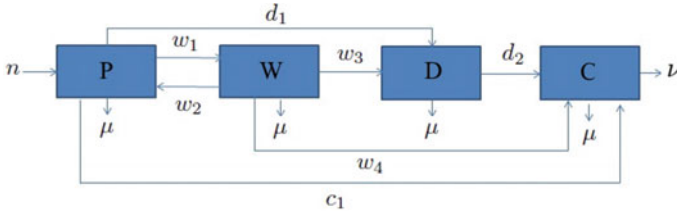
Fig. 1 Compartmental model of the dynamics of an adult population

2 Methods

The compartmental mathematical model proposed illustrates mainly the link between overweight/obesity and diabetes through the dynamics of different subpopulations. We consider an adult population of size N composed by adults without overweight and diabetes P, overweight/obese adults W, adult diabetics without complications D and with irreversible complications C. The proposed compartmental model is shown in the following figure:

Figure 1 describes the dynamics of an adult population with related parameters:

- n : denotes the incidence of adult population
- w_1 : denotes the rate of healthy persons becoming overweight/obese,
- w_2 : denotes the rate of overweight/obese persons becoming healthy,
- w_3 : denotes the rate of overweight/obese persons developing diabetes,
- w_4 : denotes the rate of overweight/obese persons developing complications,
- d_1 : denotes the rate of healthy persons becoming diabetic,
- d_2 : denotes the rate of diabetic adults developing complications,
- c_1 : denotes the rate of healthy persons developing complications,
- μ : natural mortality rate,
- ν : mortality rate related to complications,

Consequently, the proposed model can be written as:

$$
\begin{cases}
\frac{dP(t)}{dt} & = n - \alpha P + w_2 W \\
\frac{dW(t)}{dt} & = w_1 P - \beta W \\
\frac{dD(t)}{dt} & = d_1 P + w_3 W - \delta D \\
\frac{dC(t)}{dt} & = c_1 P + w_4 W + d_2 D - \varepsilon C
\end{cases}
\tag{1}
$$

with:

$$\alpha = (c_1 + d_1 + w_1 + \mu)$$
$$\beta = (w_2 + w_3 + w_4 + \mu)$$
$$\delta = (d_2 + \mu)$$
$$\varepsilon = (\mu + \nu)$$

3 Results

Theorem 1 *Assuming that initial conditions $P(0)$, $W(0)$, $D(0)$ and $C(0)$ are positive then the solution $P(t)$, $W(t)$, $D(t)$ and $C(t)$ will remain positive for all* t > 0.

Proof From the system of Eq. (1), it is easy to see that:

$$P'(t) > -\alpha P(t) \text{ or } \mathrm{d}P(t)/P(t) > -\alpha \mathrm{d}t$$
$$W'(t) > -\beta W(t) \text{ or } \mathrm{d}W(t)/W(t) > -\beta \mathrm{d}t$$
$$D'(t) > -\delta D(t) \text{ or } \mathrm{d}D(t)/D(t) > -\delta \mathrm{d}t$$
$$C'(t) > -\epsilon C(t) \text{ or } \mathrm{d}C(t)/C(t) > -\epsilon \mathrm{d}t$$

From which, we deduce after integration that

$$P(t) > P(0)\exp(-\alpha t) > 0$$
$$W(t) > W(0)\exp(-\beta t) > 0$$
$$D(t) > D(0)\exp(-\delta t) > 0$$
$$C(t) > C(0)\exp(-\epsilon t) > 0$$

Consequently, the mathematical model is epidemiologically and mathematically well posed in $E = \left\{ (P, W, D, C) \in R^{4+} \right\}$.

Lemma 1 *The system of Eq. (1) has the following equilibrium point:*

$$P^* = \frac{n\beta}{\alpha\beta - w_1 w_2}$$
$$W^* = \frac{n w_1}{\alpha\beta - w_1 w_2}$$
$$D^* = \frac{n(\beta d_1 + w_1 w_3)}{\delta(\alpha\beta - w_1 w_2)}$$
$$C^* = \frac{n(\beta c_1 \delta + \beta d_1 d_2 + d_2 w_1 w_3 + \delta w_1 w_4)}{\delta\varepsilon(\alpha\beta - w_1 w_2)}$$

Proof This equilibrium point is easily obtained by solving the following system

$$\begin{cases} \frac{\mathrm{d}P(t)}{\mathrm{d}t} & = n - \alpha P + w_2 W = 0 \\ \frac{\mathrm{d}W(t)}{\mathrm{d}t} & = w_1 P - \beta W = 0 \\ \frac{\mathrm{d}D(t)}{\mathrm{d}t} & = d_1 P + w_3 W - \delta D = 0 \\ \frac{\mathrm{d}C(t)}{\mathrm{d}t} & = c_1 P + w_4 W + d_2 D - \varepsilon C = 0 \end{cases} \tag{2}$$

Theorem 2 *The equilibrium point* (P^*, W^*, C^*, D^*) *is locally asymptotically stable*

Proof The Jacobian matrix is given by:

$$J = \begin{pmatrix} -\alpha & w_2 & 0 & 0 \\ w_1 & -\beta & 0 & 0 \\ d_1 & w_3 & -\delta & 0 \\ c_1 & w_4 & d_2 & -\varepsilon \end{pmatrix}$$

Stability analysis is based on the sign of the real part of eigenvalues λ_i obtained by solving $\det(J - \lambda I) = 0$, where I is the 4×4 identity matrix. Consequently, λ_i are roots of the following characteristic polynomial:

$$Q(\lambda) = (-\epsilon - \lambda)(-\delta - \lambda)(\lambda^2 + (\alpha + \beta)\lambda + \alpha\beta - w_1 w_2).$$

Then applying the Routh-Hurwitz stability criterion, we deduce that the equilibrium point (P^*, W^*, D^*, C^*) is locally asymptotically stable.

4 Sensitivity Analysis

The population dynamics engendered by the mathematical model depends on the values of the parameters used. In this paper, we concentrate on the control of overweight/obesity as an important risk factor of diabetes.

Hence, sensitivity analysis is used in order to determine how a relative change in parameters w_1, w_2, w_3, w_4 induces a relative change in the variable W.

In what follows, we use the normal forward sensitivity index which is given by the ratio of the relative change in the variable W to the relative change in the parameter w_i $\left(I_{w_i}^W = \frac{\frac{dW}{W}}{\frac{dw_i}{w_i}} = \frac{dW}{dw_i} \times \frac{w_i}{W} \right)$. For example, a value of $I_{w_i}^W = 1$ means that increasing w_i by 10% will induce an increase of W by 10%, while $I_{w_i}^W = -0.5$ will indicate that increasing p by 10% decreases W by 5%.

The calculus of sensitivity index is given by the following equations:

$$\begin{cases} \frac{\partial W^*}{\partial w_1} \times \frac{w_1}{W^*} &= \frac{\beta(\mu + c_1 + d_1)}{\alpha\beta - w_1 w_2} > 0 \\ \frac{\partial W^*}{\partial w_2} \times \frac{w_2}{W^*} &= -\frac{(c_1 + d_1 + \mu)w_2}{\alpha\beta - w_1 w_2} < 0 \\ \frac{\partial W^*}{\partial w_3} \times \frac{w_3}{W^*} &= -\frac{\alpha w_3}{\alpha\beta - w_1 w_2} < 0 \\ \frac{\partial W^*}{\partial w_4} \times \frac{w_4}{W^*} &= -\frac{\alpha w_4}{\alpha\beta - w_1 w_2} < 0 \end{cases} \qquad (3)$$

We can see that, an increase of the parameter w_1 will induce an increase of W, while the increase of w_2, w_3 or w_4 will decrease W.

5 Simulation, Results and Discussion

Values of the different parameters used in the mathematical model were determined from the following sources: HCP, WHO, IDF and surveys on diabetes and risk factors for NCD (Table 3).

The values of the normal foreword sensitivity index I^{W^*} are given in Table 2.

Using values of parameters as given in Table 3, we obtain curves of $P(t)$, $W(t)$, $D(t)$ and $C(t)$ showing the evolution of each of these population during a period of ten years (Fig. 2). Then using sensitivity analysis, we can see that a relative decrease of 10% in the rate w_1 will induce a relative decrease of nearly 3.2% in the size of W. Consequently, reducing the rate of becoming overweight/obese (w_1) is the most efficient strategy since it reduces simultaneously the size of W, D and C. Pragmatically, such a strategy will need to: (1) continually remind people to achieve and maintain a healthy body weight and at the same time to raise awareness on the fact that obesity is a high risk for developing diabetes and CVDs, (2) encourage people to be regularly active, (3) re-educate the whole population and particularly children to eat a healthy diet, promoting fruits and vegetables while avoiding sugar and saturated fats and (4) adopt more generally a healthy lifestyle (Fig. 3).

Table 2 Normal forward sensitivity index

	w_1	w_2	w_3	w_4
I^{W^*}	0.3204	−0.0125	−0.2803	−0.2002

Table 3 Input parameters

Parameter	Value
n	530,000
c_1	0.005
d_1	0.005
d_2	0.01
w_1	0.05
w_2	0.001
w_3	0.007
w_4	0.005
μ	0.012
ν	0.012
N (Adult population)	24×10^6
$D(0)$	$0.11 \times N$
$C(0)$	$0.5 \times D(0)$
$W(0)$	$0.5 \times N$
$P(0)$	$N - W(0) - D(0) - P(0)$
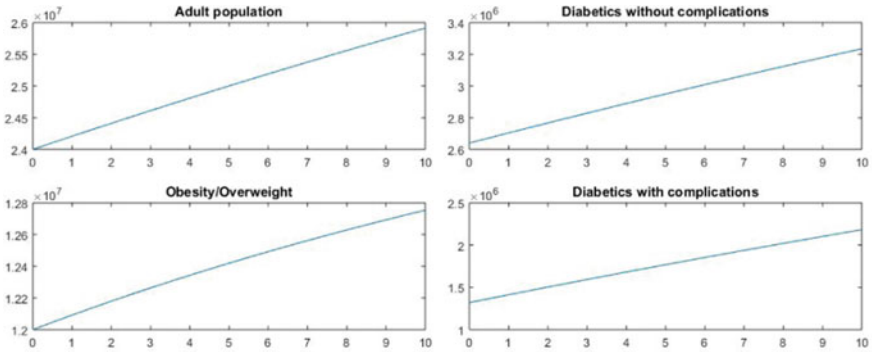
Fig. 2 Evolution of the adult population, overweight/obese adults and diabetics with and without complications using values given in Table 3 with time in years
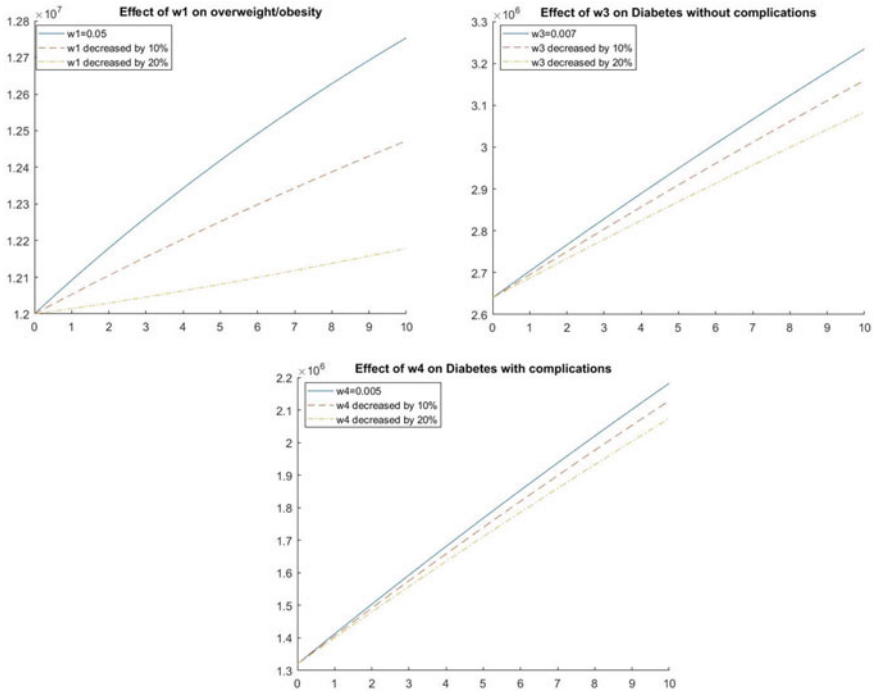
Fig. 3 Effect of parameters w_1, w_3 and w_4 on the size of populations of obesity/overweight (W), diabetes without complications (D) and diabetes with complications (C)

6 Conclusion

In this paper, a simple compartmental model was presented to illustrate the importance of the link between overweight/obesity and diabetes in adults. Our analysis showed that reducing the rate of becoming overweight/obese is the most efficient strategy that decision makers should adopt by sensitizing for a healthy lifestyle, including healthy diet, regular physical activity and healthy body weight.

References

1. Barnes AS (2011) The epidemic of obesity and diabetes: trends and treatments. Tex Heart Inst J 38(2):142
2. Boutayeb A, Lamlili MEN, Boutayeb W (2020) A review of compartmental mathematical models used in diabetology. Springer International Publishing, Cham, pp 217–250
3. Boutayeb A, Twizell E, Achouayb K, Chetouani A (2004) A mathematical model for the burden of diabetes and its complications. Biomed Eng Online 3(1):1–8
4. International Diabetes Federation (2021) Diabetes atlas 2021 (10th edn). https://diabetesatlas.org/atlas/tenth-edition/
5. Kouidere A, Balatif O, Ferjouchia H, Boutayeb A, Rachik M (2019) Optimal control strategy for a discrete time to the dynamics of a population of diabetics with highlighting the impact of living environment. Discrete Dyn Nat Soc 2019
6. Kouidere A, Labzai A, Khajji B, Ferjouchia H, Balatif O, Boutayeb A, Rachik M (2020) Optimal control strategy with multi-delay in state and control variables of a discrete mathematical modeling for the dynamics of diabetic population. Commun Math Biol Neurosci 2020
7. Ministère de la Santé: Enquête nationale sur les facteurs de risque communs des mnt. https://www.sante.gov.ma/Documents/2019/05/Rapportde%20l%20enqute%20stepwise.pdf (2017-2018)
8. Tazi MA, Abir-Khalil S, Chaouki N, Cherqaoui S, Lahmouz F, Sraïri JE, Mahjour J (2003) Prevalence of the main cardiovascular risk factors in Morocco: results of a national survey, 2000. J Hypertens 21(5):897–903
9. World Health Organization (2016) Global report on diabetes. https://www.who.int/publications/i/item/9789241565257

Dermatologist-Level Classification of Skin Cancer with Level Set Method and Isolation Forest

Khalid Bellaj, Soumaya Boujena, and Mohammed Benmir

Abstract Cancer classification (CC) by machine learning (ML), a research area that combines ML and scientific computing techniques, has experienced considerable growth in recent years. This article reviews existing and emerging approaches at CC and presents some well-known results in a unified framework. We study the combination of segmentation methods (SM) with ML techniques for solving partial differential equations (PDE). As a concrete example of SM, improved by ML, we present a method that uses isolation forest (IF) to reduce the computational cost in our SM while maintaining accuracy. Finally, experimental results show that our method significantly outperforms existing solutions for CC.

Keywords Image segmentation · Lesions · Skin · Isolation forest · Level set · Machine learning · Domain decomposition methods · Melanoma skin cancer

1 Introduction

In recent years, skin cancer or melanoma tumor has become the most common cancer in humans. Skin cancer is often divided into two groups—melanoma and non-melanoma skin cancer (see [1] and [2]). Melanoma skin cancer (MSC) is a malignant tumor of the cells that give skin its color (melanocytes), and non-melanoma skin cancer (NMSC) is a cancer of the basal and squamous cells of the epidermis. NMSC most commonly occurs on various parts of the body such as the head, neck, palms, hands, and face when the skin is exposed to direct sunlight. However, there are several other less common skin cancers, including Merkel cell tumors, cutaneous lymphomas, and sarcomas [2]. The importance of a correct melanoma diagnosis also stems from the

---

WWW home page: http://www.fsac.ac.ma/front/index.html.

K. Bellaj (✉) · S. Boujena · M. Benmir
Fundamental and Applied Mathematics Laboratory, Department of Mathematics and Computer Sciences, Ain Chock Science Faculty, Hassan II University of Casablanca, Km 8 Route El Jadida POB 5366 Maarif, Casablanca, Morocco
e-mail: KHALID.BELLAJ-ETU@etu.univh2c.ma

fact that the number of melanoma cases worldwide is increasing by 6% annually, faster than any other cancer.

1.1 Contributions

This work extends existing efforts to classify skin cancer at the dermatologist level. We proposed a combination of segmentation methods with machine learning techniques to solve partial differential equations for melanoma skin cancer abnormality detection. In addition, we explored how our SM combined with ML achieves equilibrium in images corrupted by high noise level. We also present an updated approach based on range decomposition methods and other optimizations, including steps 3 and 10 in Algorithm 1, which was developed originally for high-resolution images.

1.2 Paper Organization

The remainder of the paper is organized as follows. In Sect. 2, we first introduce the level set technique (LS), the most popular numerical method for analyzing surfaces and shapes. Then, we present a modified Perona-Malik (PM) model for image denoising. We then discuss other techniques, albeit to a lesser extent, such as IF. In Sect. 3, we present the algorithms of the proposed approaches for CC recognition. We also discuss how domain decomposition methods (DDMs) can be combined with our denoising techniques to optimize the desired results, and we give numerical examples to illustrate the approach and model predictions for CC. In Sect. 4, we present the experimental datasets and then analyze the experimental results to validate our proposed method through extensive experiments on melanoma images. Conclusions are drawn in Sect. 5.

2 Segmentation and Image Denoising Models

2.1 Level Set Model

There are many techniques for image segmentation in the literature. Some of these techniques use only grayscale histograms, others use spatial detail, and still others use fuzzy set theory. To improve the segmentation performance of region-based methods for an image with intensity inhomogeneity, LS methods have been recently proposed. As an example, we cite the model of Li et al. [3], which we improved in our previous work [4]. Namely, we used a diffusion function instead of the edge indicator function g defined in [3]. This allowed us to convert the local intensity information into an

energy function, whereas in the model of Li and others, two smooth functions are used to describe the local intensities inside and outside the contour. For an image I on a domain Ω, the energy function of our LS model [7] is given by:

$$E(\psi) = \beta\, F(\psi) + E_m(\psi) = \beta\, F(\psi) + L_g(\psi) + A_g(\psi), \tag{1}$$

with

$$F(\psi) = \frac{1}{2} \int_\Omega (|\nabla\psi| - 1)^2 \, dx\, dy,$$

$$E_m(\psi) = \rho \int_\Omega g(|\nabla U|)\delta(\psi)|\nabla\psi|\, dx\, dy + \nu \int_\Omega g(|\nabla U|)H(-\psi)\, dx\, dy,$$

where $\beta, \rho > 0$ and ν are constants, U is a restored image from the parabolic nonlinear problem (3), and the a diffusion function is:

$$g : R_0^+ \longrightarrow R^+$$

$$g(s) = \frac{1}{\sqrt{1 + \left(\frac{s}{k}\right)^2}} + \alpha,$$

The term E_m is an external energy that would direct the zero set toward the desired object edges, $L_g(\psi)$ is the weighted length of the zero set representing the region $\{(x, y)|\psi(x, y, t) = 0\}$, and $A_g(\psi)$ is the weighted area of the region $\{(x, y)|\psi(x, y, t) < 0\}$. Here, I_σ is the smooth Gaussian version of the image I. H and δ stand for the one-dimensional Heaviside function and Dirac function, respectively. The term $P(\psi)$ is the internal energy of the level function ψ, which penalizes the deviation of a signed distance function [5]. The parameter $\beta > 0$ controls the penalizing effect of this deviation. Using the continuous gradient descent method [6], we obtain the evolution PDE for the model of Li et al. [3].

$$\frac{\partial\psi}{\partial t} = \beta\,(\Delta\psi - \beta) + \rho\delta(\psi)\mathrm{div}\left[g(|\nabla U|)\frac{\nabla\psi}{|\nabla\psi|}\right] + \nu g(|\nabla U|)\delta(\psi), \tag{2}$$

with initial condition $\psi(x, y, 0) = \psi_0(x, y)$, $\psi_0(x, y)$ is the level initialization function corresponding to the initial curve $C(x, y, 0)$, $\beta = \mathrm{div}\left[\frac{\nabla(\psi(x))}{|\nabla(\psi(x))|}\right]$ corresponds to the curvature of the evolving contour, δ is the Dirac function and ν is the unit normal vector to the range boundary $\partial\Omega$.

Note that the convergence of the approximation of this PDE depends strongly on the choice of the starting contour $C(x, y, 0)$ for the segmentation process. Unfortunately, this method does not work for images corrupted by high noise, so the denoising phase is important.

2.2 Isolation Forest

As mentioned earlier, several factors influence the classification of skin cancer images:

- The melanomas have complex structure, large size differences, and complex colors in the skin.
- The thickness of the melanoma on the adjacent skin regions.
- The boundaries of melanoma are not always well defined.
- The influence of small structures, hairs, bubbles, light reflections, and other artifacts.

These factors make classification more complex. To analyze skin lesions, they must be precisely localized and isolated. The description of the concept IF, originally proposed by Liu et al. [7] in 2008, attempts to classify unsupervised anomalies for high-dimensional datasets. IF or iForest is a tree-based model that is an efficient ensemble-based method for outlier detection and is a strong and widely used algorithm that can handle large datasets. It differs from classical anomaly detection methods in its high accuracy, low memory requirements, and low linear time complexity, regardless of the size of a dataset.

2.3 Denoising Model

Most segmentation models are not suitable for noisy environments. Some work uses the Markov random field (MRF) model, which is robust in removing noise, but its implementation requires high computational cost. The use of isotropic diffusion for image denoising has a major drawback: its homogeneity, which leads to undesirable diffusion of image features such as edges. The PM equation modifies the heat equation by adding a diffusion coefficient based on the spatial activity in a given part of an image [8], measured against the norm of the local image gradient. This process for a given noisy image u_0 is described by the following mathematical model:

$$
P(u) = \begin{cases} \dfrac{\partial u}{\partial t} - \mathrm{div}(g(|\nabla u|)\nabla u) = 0, & \text{in } \Omega \times I \\ u(x, 0) = u_0(x), \forall x \in \Omega \\ \dfrac{\partial u}{\partial n} = 0, & \text{in } \partial\Omega \times I, \end{cases} \tag{3}
$$

Ω is a bounded domain of R^2 with a corresponding smooth boundary (here the image domain), n denotes the outer unit normal to Ω, and I is the scaling interval. The diffusion coefficient $g : R_0^+ \longrightarrow R^+$ is a non-increasing function of the magnitude of the local image gradient ∇u and has other properties, see [4] for more details. $u(x, t)$ is an unknown function defined in $I \times \Omega$, which is a restored image. Equation (3) is

accompanied by zero Neumann boundary conditions and the initial condition. Various diffusion coefficients have been proposed in the literature. For example, in [8] PM used:

$$g_{pm}(s) = \frac{1}{1 + \left(\frac{s}{k}\right)^2},$$ (4)

where $s = |\nabla u|$ and k is the threshold parameter of the gradient theorem that determines the extent of diffusion taking place, for more details see [8].

Note that the existence and uniqueness of the solution of (3) in Hilbert space are proved under the following hypothesis; [5]:

- $g : R^+ \longrightarrow R^+$ is a decreasing function,
- g is differentiable,
- $g(0) = 1 + \alpha$,
- $\lim_{s \to +\infty} g(s) = \alpha$,
- $2s \left| g^{'}(s) \right| \le g(s), s \ge 0$,

where $\alpha > 0$ is a positive parameter (generally very small) added to the diffusion function g_{pm} of PM to ensure monotonicity of the differential operator associated with the problem (3) without changing the numerical performances of the PM model.

2.4 Discretizations of Proposed the Problem

In the following, we present the discretization of the proposed modified PM model (3) and the partial evolution differential equation (2) for the Li et al. model using the finite element method. To approximately solve the parabolic nonlinear problem (3), we discretize it using the Galerkin finite element method and obtain the well-posed linear algebraic problem:

$$(A + \Delta t B)U^{n+1} = AU^n,$$ (5)

where

$$U^n = \begin{pmatrix} v_1^n \\ \vdots \\ v_p^n \end{pmatrix}, A_{ij} = \int_\Omega \varphi_i \varphi_j dx, B_{ij} = \int_\Omega \mu \left(|\nabla v^n| \right) \nabla \varphi_i . \nabla \varphi_j dx.$$

where the step size Δt should satisfy:

$$\Delta t < \frac{1}{\max_{i \in J} |a_{ii}(g^n)|}$$

A is a so-called mass matrix and B is the stiffness matrix. Thus, the discrete solution can be found efficiently by so-called conjugate gradient methods.

Therefore, the numerical approximation of (2) is given by the following discretization:

$$\frac{\psi_{i,j}^{n+1} - \psi_{i,j}^n}{\Delta t} = \beta \left(\Delta \psi_{i,j}^n - C_{i,j}^n \right) + \rho \delta \left(\psi_{i,j}^{n+1} \right) g \, C_{i,j}^n + v \, g \, \delta \left(\psi_{i,j}^n \right) \quad (6)$$

with

$$C_{i,j}^n = \Delta_0^x \left(\frac{\Delta_0^x \psi_{i,j}^n}{|\psi|_{i,j}^n} \right) + \Delta_0^y \left(\frac{\Delta_0^y \psi_{i,j}^n}{|\psi|_{i,j}^n} \right), |\psi|_{i,j}^n = \sqrt{\left(\Delta_0^x \psi_{i,j}^n \right)^2 + \left(\Delta_0^y \psi_{i,j}^n \right)^2}$$

In the following, we take g as follows:

$$g(s) = \frac{1}{\sqrt{1 + \left(\frac{s}{k} \right)^2}} + \alpha, \quad (7)$$

where $\alpha = 10^{-6}$ and $k > 0$ is a given parameter, for more details see [5].
.

3 Proposed Method for Cancer Classification

In order to keep the load between the denoising and segmentation models in balance, the initialization of the contour $\psi_0(x, y)$ must be chosen appropriately. For example, we used a simplex-based algorithm to set the initial contours for Eq. (2). Another important aspect of our contribution is the improvement of numerical algorithms using DDMs. As described in our previous studies [4], we use DDMs for parallel and sequential processing of super-resolution images. These improvements can be summarized in the following algorithm:

Algorithm 1

1: For each skin cancer image type.
2: Initialise: Read the value u_0 from the image of MSC.
3: Select whether to perform DDM or not?
4: For each subdomain of the skin cancer image.
5: Select whether denoising should be performed or not?
6: **if** $denoising = True$ **then**
7:　　Solve the equation.

$$u^{n+1} = \left(\Delta t A \left(g^n \right) - I \right) u^n$$

8: Initialization: $\psi_0(x, y)$ Generate the value of the initial contours using the simplex algorithm.
9: Detection and diagnosis of cancer:

$$\frac{\partial \psi}{\partial t} = \beta \left(\delta \psi - \beta \right) + \rho \delta(\psi) \mathrm{div} \left[g \frac{\nabla \psi}{|\nabla \psi|} \right] + v g \delta(\psi)$$

10: Cancer Classification with IFM.
11: For each skin cancer image type.
12: Read the image of a lesion using adaptive OM.
13: Read the FCMC image of the lesion.
14: Calculate the difference in clustering accuracy between the adaptive OM and our algorithm.
15: Calculate the accuracy between the FCMC and our algorithm.
16: Calculate the average accuracy for all lesion imaging methods for both benign and MSC images.

4　Experimental Results

In our experiment, we used a set of natural images available for download from the Harvard Dataverse or the ISIC-archive. We generated noisy images from clean training and test images by adding Gaussian noise to them. The function g used in algorithm 1 is defined by (7) and tests the same properties in [5]. The parameters of the model are $\Delta t = 0.24$, $\mathrm{d}x = \mathrm{d}y = 1$, $\alpha = 10^{-6}$ and $k = \frac{1}{4}$. DD is used to maximize parallel processing capacity by ensuring that all subdomains have the same number of pixels. We recommend dividing the number of processors into 4, 9, 16 or more subdomains.

4.1　Numerical Examples

We will now give some numerical examples illustrating the predictions of our algorithms. In Table 1, we show the efficiency and accuracy of our proposed method, especially algorithm 1, compared to the other methods, such as the FCMC method, especially in CPU. This indicates that algorithm 1 can effectively detect cancer even in highly noisy images while preserving the original image information faster than

Table 1 Result of segmentation technique for image (c) in Fig. 1

Method	Accuracy	Sensitivity	Specificity	CPU
AS-Net [9]	95.16	88.00	89.25	34.82
SLSNet [10]	91.42	90.45	91.44	47.81
OM	96.5	98.1	95.4	29.58
FCMC	92.66	96.75	94.93	30.88
Algorithm 1	97.5	96.8	95.8	20.22
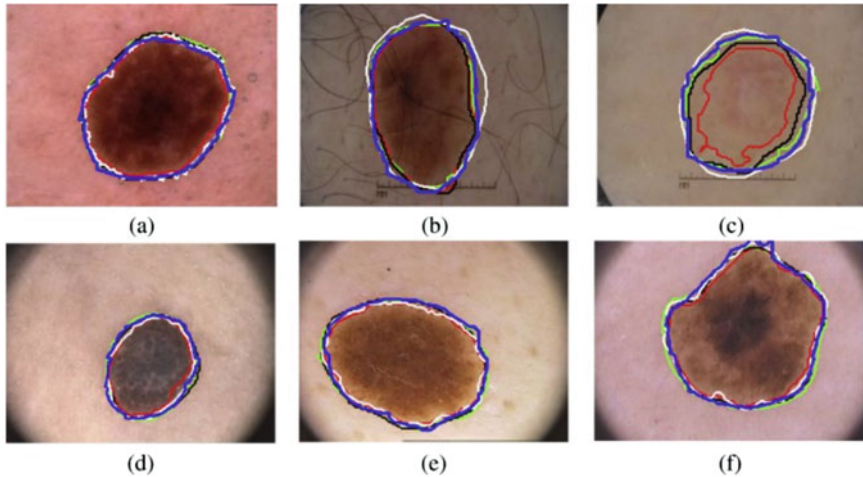
Fig. 1 The segmentation results of our algorithms compared with the AS-Net method, Garcia method, OM and FCMC for **a** benign, **b** melanoma and **c** seborrheic keratosis cases from ISBI 2017 dataset while **d** common nevi, **e** atypical nevi and **f** melanoma show the cases from PH2 dataset. The segmentation results of OM (white) and AS-Net method (blue), while FCMC (green), algorithm 1 (red) and SLSNet (black), see Table 1 for more details

other methods. This remarkable improvement of algorithm 1 is due to the optimizations used, such as steps 3 and 10 in algorithm 1, especially for images with large size.

5 Conclusions

In this paper, we present a new approach for skin cancer detection and classification using a combination of LS and ML trained with denoising auto-encoders. We propose a new training scheme for iForest algorithm to denoise and detect images in a uniform framework. The conducted experiment shows that the proposed method achieves comparable performance to OM and AS-Net. Moreover, our nonlinear approach

successfully handles the image size issue, which, to our knowledge, has not been addressed before. Our experimental results show that the quality of image denoising and recognition by our approach is better, especially when the proposed algorithm 1 is used, which can improve the performance of CC in the unsupervised feature learning tasks.

References

1. Silveira M, Nascimento JC, Marques JS, Marccal AR, Mendonca T, Yamauchi S, Maeda J, Rozeira J (2009) Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. IEEE J Sel Top Sign Process 3:35–45
2. Yang S, Xiao H, Cao L (2021) Recent advances in heat shock proteins in cancer diagnosis, prognosis, metabolism and treatment. Biomed Pharmacother 142:1124
3. Li C, Xu C, Gui C, Fox MD (2010) Distance regularized level set evolution and its application to image segmentation. IEEE Trans Image Process 19:3243–3254
4. Bellaj K, Boujena S, Guarmah EE (2020) An improved approach for image segmentation and three-dimensional reconstruction. Discontinuity Nonlinearity Complex 9:199–215
5. Boujena S, Pousin J, El Guarmah E, Gouasnouane O (2021) An improved nonlinear model for image restoration. Pure Appl Funct Anal 2:599–623
6. Bellaj K, Boujena S, El Guarmah E (2019) One approach for image denoising based on finite element method and domain decomposition technique. Int J Appl Phys Math 7:141
7. Liu FT, Ting KM, Zhou Z-H (2010) On detecting clustered anomalies using SCiForest. In: Joint European conference on machine learning and knowledge discovery in databases. Springer, pp 274–290
8. Penrona P, Malik J (2017) Scale space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell 12:629–639
9. Hu K, Lu J, Lee D, Xiong D, Chen Z (2022) AS-Net: attention synergy network for skin lesion segmentation. Expert Syst Appl 201:117112. ISSN: 0957-4174. https://doi.org/10.1016/j.eswa.2022.117112
10. Sarker MMK, Rashwan HA, Akram F, Singh VK, Banu SF, Chowdhury FUH, Choudhury KA, Chambon S, Radeva P, Puig D, Abdel-Nasser M (2021) SLSNet: skin lesion segmentation using a lightweight generative adversarial network. Expert Syst Appl 183:115433. ISSN: 0957-4174. https://doi.org/10.1016/j.eswa.2021.115433

A Generalized Freeman Chain Code for Offline Arabic Character Recognition

Mohammed Kadi, Youssef Douzi, and M'barek Nasri

Abstract This work goes through several stages: First, the Arabic character is preprocessed and thinned. And a preliminary classification of the character is done according to the number of its loops and the nature, number, and position of its complementary parts (Dots, Hamza, etc.). Then, the Freeman chain code of 8 directions is generated according to a new innovative algorithm that can represent the internal parts of the Arabic character if they exist and without redundant information. After that, a normalization process based on the generalization of the code of 8 directions to a code of 24 directions is done. Finally, and for comparison, the two algorithms, traditional and proposed, are implemented separately to recognize Arabic characters. After 1023 tests on a set of 300 printed characters, isolated and distributed over 10 fonts, the recognition rate obtained was always higher than that obtained with the traditional algorithm, from 90.7 to 100%. Also, the number of font combinations giving good results almost doubled, and the execution time when normalizing the generated code was less.

Keywords Arabic character recognition · Generalized Freeman chain code · Coding of 24 directions · Feature extraction

1 Introduction

Freeman coding is a well-known structural method for representing the contours of shapes via an ordered sequence of directions called a "chain code" [1–3]. It consists of traversing the contour of a shape from a point called the "starting pixel", in a specific direction (clockwise or anti-clockwise) and recording the different directions of its curve until arriving again at the starting point [2, 4–6]. Several works have used

M. Kadi (✉) · M. Nasri
Laboratory MATSI, EST Mohamed I University, Oujda, Morocco
e-mail: kadi.m926@gmail.com

Y. Douzi
Laboratory ACSA, FSO Mohamed I University, Oujda, Morocco
e-mail: y.douzi@ump.ac.ma

Freeman's coding for Arabic character recognition after the skeletonization process [7–10]. Despite the promising results obtained in these works, this coding has been used traditionally. Indeed, Freeman's coding originally aims to represent the shapes that do not contain other internal shapes and whose contour is continuous and closed and does not intersect with itself [6, 11]. And as described in [12], the traditional Freeman coding gives a code only for the perimeter of the Arabic character. It does not take into account the case where the Arabic character contains an internal part as the case of the two forms "ه" and "ﻪ" of the Arabic character "ه", which makes these two forms like the two forms "ﻢ" and "ﻤ" of the Arabic character "م". This can lead to confusion if the purpose is to recognize these two forms. In addition, the resulting code contains generally a repetition of the information. In 2019, a new algorithm was proposed in [12] to overcome these problems and improve Freeman's coding for Arabic characters. This algorithm takes into account all the forms on which the Arabic character can be and reduces generally and significantly the length of the resulting code. In this work, we applied the proposed algorithm in [12] and the traditional algorithm to recognize isolated and printed Arabic characters which are written with different fonts. Then, we compared the obtained results by each algorithm in terms of recognition rate and in terms of execution time. We used the same database, the same normalization process, and the same computing conditions.

2 Preprocessing

After the binarization, and the skeletonization process by the method [13], the basic preprocessing is to remove the right angles composed of three connected pixels which are a 4-connected component. This removing has no impact on the recognition rate. It is for reducing the execution time when we generate the generalized code as we will notice in the next section (see Fig. 1.).
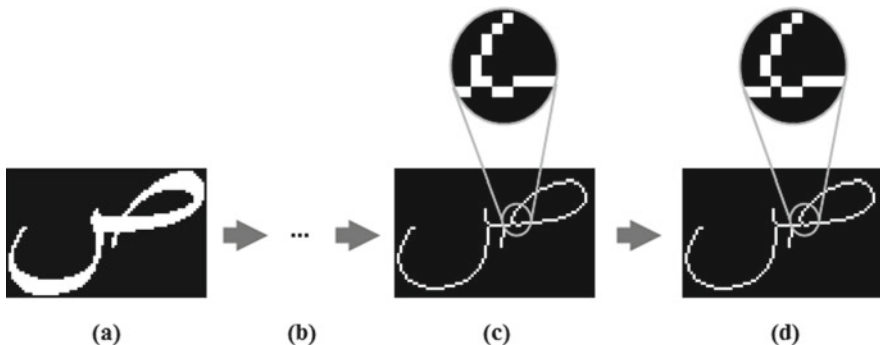
Fig. 1 a Original image of the Arabic character "ص". **b** Image after the process of enhancement and skeletonization. **c** Right angles may appear. **d** After removing the right angles

3 Feature Extraction and Classification

Although the proposed algorithm in [12] reduces significantly the length of the Freeman code for Arabic characters, this code remains long and not constant in general. Therefore, if the goal is to exploit it in the field of Arabic character recognition, it must be normalized to be short and constant to facilitate the comparison between the different codes. The common normalization used by several works close to our work as in [7–10] encounters some problems when it is implemented. Since we sometimes get a normalized code with a length shorter or longer than the desired length [8, 10]. Therefore, in our work, we propose a normalization process consisting of two basic steps: generalization and construction of a constant feature vector.

3.1 Generalization of the Freeman Coding

Freeman's coding is originally based on four or eight directions, and eight directions are used most often. We propose in this work an efficient method to generalize the Freeman coding of eight directions to a 24-direction coding. In fact, there are suggested methods to generalize the Freeman coding as in [14], but the novelty of our method is that instead of extracting the generalized code from scratch, we propose to extract it directly from the code of 8 directions. And we will set up a table to transform the code of 8 directions to a code of 24 directions. Our proposed generalization method is described as follows: In addition to the eight neighboring pixels of the current pixel, and which form a square ring around the current pixel, and which allowed us to determine eight directions (see Fig. 2a), we also considered the sixteen neighboring pixels of the next ring, which allowed us to define other 16 directions as in (Fig. 2b). The total is, therefore, 24 directions.

To generate the code, of 24 directions, we noticed that it can be extracted more quickly from the Freeman code of eight directions. For example, and as shown in
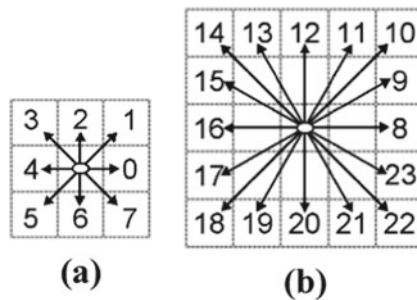
Fig. 2 **a** 8-direction Freeman coding scheme, which describes the direction of each adjacent pixel in the counterclockwise. **b** Other directions in the counterclockwise to realize a generalized coding of 24 directions
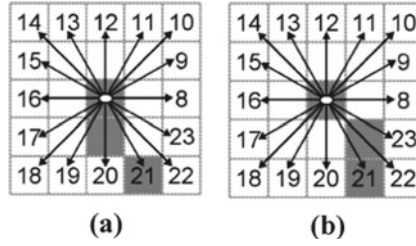
Fig. 3 To move from the current pixel to the pixel located at the direction 21, by two successive directions in the Freeman code of eight directions, there are only two possible paths (**a**) and (**b**) represented in gray

Fig. 3, to move from the current pixel (in the center) to the pixel located at direction 21, by two successive directions in the Freeman code of eight directions, there are only two possible paths (if we eliminate the right angles as indicated in Sect. 2) represented in gray. The first path (Fig. 3a) is to move from the current pixel to the pixel located at the direction 6 and then to the pixel located at the direction 7. The second (Fig. 3b) is to move from the current pixel to the pixel located at the direction 7 and then to the pixel located at the direction 6. We will express this briefly as follows:

$$(21)_{24} = (67)_8 \text{ or } (21)_{24} = (76)_8$$

Thus, we have put a table (Table 1) to convert the code of eight directions into a generalized code of 24 directions.

In the first row of the table, we have placed two consecutive digits from the code of eight directions and, in the second line, the corresponding directions in the code of 24 directions.

3.2 Construction of the Feature Vector

After extracting the generalized code of 24 directions, as indicated in the previous subsection, we will build the following feature vector F:

$$F = (a_0, a_2 \ldots, a_{23}) \tag{1}$$

where

$$a_i = \frac{f_i}{L}, \ 0 \leq i \leq 23 \tag{2}$$

f_i is the frequency of the direction i in all the code, and L is the length of the code. The division of frequencies by the length of the code aims to make the features invariant

Table 1 Transformation of a code of eight directions into a code of 24 directions

00	01	07	10	11	12	13	17	21	22	23	31	32	33	34	35	43	44	45	53	54	55	56	57	65	66	67	70	71	75	76	77
8	9	23	9	10	11	12	8	11	12	13	11	12	13	14	15	16	15	16	17	16	17	18	19	16	17	18	19	20	20	21	22

Note During the preprocessing phase, we did not allow the existence of right angles as we saw earlier. This will reduce the number of possible paths to go from the current pixel to the pixel located at one of the last 16 directions and will exclude several consecutive digits in the 8-direction Freeman code that do not match any of the 16 remaining directions in the generalized code of 24 directions
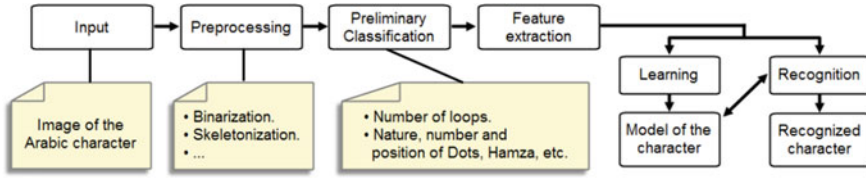
Fig. 4 Proposed system

to scaling. Also, this feature vector is invariant by its nature to translation because the chain code originally contains no information about the coordinates of the pixels that form the character. However, it remains that the method is sensitive to rotation. And to solve this, since we will be treating the character in a context of a text or word, we propose to apply the "Hough Transform" to detect the baseline of writing and then rotate it to the horizontal direction if it is tilted [15–17].

3.3 Classification

After generating the feature vector, the Euclidean distance is used to compare the different feature vectors. To reduce the execution time and increase the accuracy of recognition, we proposed a preliminary classification of the Arabic character according to the number of loops, the nature, number, and position of complementary parts (Points, Hamza, etc.), as in [15]. Finally, a schema of our system is shown in Fig. 4.

4 Results

4.1 Recognition Tests

We performed 1023 recognition tests on a set of 300 printed characters, isolated and distributed over 10 well-known fonts: {Tradition Arabic, Times New Roman, Arial, Simplified Arabic, Simplified Arabic Fixed, Arabic typesetting, Decotype Naskh, Akhbar MT, Microsoft Sans Serif, and Andalus}. Each font will be noted, in this order, by F_i, with i an integer, $1 \leq i \leq 10$. And in each test, we take a new combination of fonts, for the learning phase (without keeping any previous learning information from previous tests), and then calculate the recognition rate (RR) for each font from the whole set. Then, for each value of k, where k is an integer, $1 \leq k \leq 10$, we choose the k-combinations that give the best result. Finally, by using the traditional algorithm, we obtained the results presented in Table 2, while Table 3 presents the results obtained using the proposed algorithm in [12]. It is appropriate here to note

Table 2 Results obtained using the traditional algorithm

k	Combinations	RR (%)
1	{*F2*}, {*F3*}	86.0
2	{*F2, F8*}, {*F3, F8*}	90.0
3	{*F4, F8, F10*}	93.7
4	{*F1, F4, F6, F10*}	95.7
5	{*F1, F4, F7, F8, F10*}	97.3
6	{*F1, F4, F7, F8, F9, F10*}	99.3
7	{*F1, F4, F6, F7, F8, F9, F10*}, {*F1, F5, F6, F7, F8, F9, F10*}	99.7
8	{*F1, F4, F5, F6, F7, F8, F9, F10*}	100.0
9	{*F1, F2, F4, F5, F6, F7, F8, F9, F10*}, {*F1, F3, F4, F5, F6, F7, F8, F9, F10*}	100.0
10	{*F1, F2, F3, F4, F5, F6, F7, F8, F9, F10*}	100.0

that we used with the two algorithms the same normalization process described in Sect. 3 and the same computing conditions. The two tables show that the recognition rate obtained using the proposed algorithm is mostly higher than the rate obtained using the traditional algorithm. And the two rates are equal only when the number of learning fonts increases. But, when the number of learning fonts is less, it is better. In addition, the number of combinations giving the best results is 30 combinations using the proposed algorithm and only 14 combinations using the traditional algorithm. In other words, the proposed algorithm almost doubled the number of combinations that gave the best results. To further illustrate the obtained results, we have used curves (see Fig. 5). The curve which represents the proposed algorithm is generally above the curve which represents the traditional algorithm. And the two curves approach each other only when the number of training fonts increases. But, the fewer the training fonts, the better; this is one of the main advantages of the proposed algorithm.

4.2 Execution Time

Regarding the execution time, we calculated the total time needed to extract the generalized code of 24 directions from the code of eight directions generated in an attempt by the proposed algorithm and in another attempt by the traditional algorithm. The time taken by the proposed algorithm was less by about 31.7%. All the experiments were done with MATLAB 7.14 on an Intel Dual-Core Celeron CPU N2830, 2.16 GHz, and x64-based with 4 Go RAM.

Table 3 Results obtained using the proposed algorithm

k	Combinations	RR (%)
1	{*F1*}, {*F4*}	90.7
2	{*F2, F5*}, {*F3, F5*}, {*F4, F8*}	93.7
3	{*F4, F6, F10*}	97.7
4	{*F5, F7, F9, F10*}	99.0
5	{*F5, F6, F7, F9, F10*}	99.7
6	{*F1, F5, F6, F7, F9, F10*}	100.0
7	{*F1, F2, F5, F6, F7, F9, F10*}, {*F1, F2, F5, F6, F8, F9, F10*}, {*F1, F3, F5, F6, F7, F9, F10*}, {*F1, F3, F5, F6, F8, F9, F10*}, {*F1, F4, F5, F6, F7, F9, F10*}, {*F1, F5, F6, F7, F8, F9, F10*}	100.0
8	{*F1, F2, F3, F5, F6, F7, F9, F10*}, {*F1, F2, F3, F5, F6, F8, F9, F10*}, {*F1, F2, F4, F5, F6, F7, F9, F10*}, {*F1, F2, F4, F5, F6, F8, F9, F10*}, {*F1, F2, F5, F6, F7, F8, F9, F10*}, {*F1, F3, F4, F5, F6, F7, F9, F10*}, {*F1, F3, F4, F5, F6, F8, F9, F10*}, {*F1, F3, F5, F6, F7, F8, F9, F10*}, {*F1, F4, F5, F6, F7, F8, F9, F10*}	100.0
9	{*F1, F2, F3, F4, F5, F6, F7, F9, F10*}, {*F1, F2, F3, F4, F5, F6, F8, F9, F10*}, {*F1, F2, F3, F5, F6, F7, F8, F9, F10*}, {*F1, F2, F4, F5, F6, F7, F8, F9, F10*}, {*F1, F3, F4, F5, F6, F7, F8, F9, F10*}	100.0
10	{*F1, F2, F3, F4, F5, F6, F7, F8, F9, F10*}	100.0
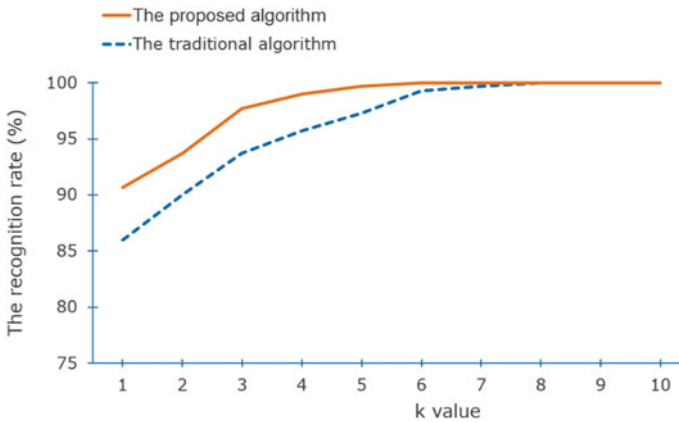
Fig. 5 Highest recognition rate for each value of *k*

5 Conclusion

The application of the proposed algorithm in [12] to recognize Arabic character shows that the recognition rate obtained was better than the rate obtained with the traditional algorithm, from 90.7 up to 100%. In particular, when the number of fonts used in the learning phase is small. Also, the number of font combinations giving good results was higher, and the execution time taken by the proposed algorithm

when exploiting the generated code was less. Further research is still needed. We will seek to expand the database and enhance the results with the latest advances in artificial intelligence and deep learning. This is a starting point for our future research.

References

1. Brinda T, Dharma D (2021) Enhancing the compression performance in medical images using a novel hex-directional chain code (Hex DCC) representation. Soft Comput 25(7):5807–5830
2. Gonzalez RC, Woods RE (2018) Digital image processing, 4th edn. University of Tennessee, Pearson
3. Arif M, Hassan H, Nasien D, Haron H (2015) A review on feature extraction and feature selection for handwritten character recognition. Int J Adv Comput Sci Appl 6(02):204–212
4. Burger W, Burge MJ (2016) Digital image processing: an algorithmic introduction using Java, 2nd edn. Springer, London
5. Haboubi S, Nakkach H, Amiri H (2016) Online Arabic character recognition using global and local features. In: Proceedings of 3rd international conference on automation, control engineering and computer science (ACECS'16), pp 46–51. At Hammamet, Tunisia
6. Jähne B (2005) Digital image processing, 6th edn. Springer, Berlin Heidelberg
7. Althobaiti H, Lu C (2017) A survey on Arabic optical character recognition and an isolated handwritten Arabic character recognition algorithm using encoded Freeman chain code. In: The 51st annual conference on information sciences and systems (CISS), Baltimore, MD, pp 1–6
8. Izakian H, Monadjemi SA, Ladani BT, Zamanifar K (2008) Multifont Farsi/Arabic isolated character recognition using chain codes. Int J Comput Inf Eng (World Academy of Science, Engineering and Technology) 2(7):67–70
9. Abed MA (2012) Freeman chain code contour processing for handwritten isolated Arabic characters recognition. AL-yarmouk J (College of Computers Sciences and Mathematics, University of Tikrit, Iraq) 1:90–105
10. Lamghari N, Charaf MEH, Raghay S (2016) Template matching for recognition of handwritten Arabic characters using structural characteristics and Freeman code. Int J Comput Netw Inf Secur 14(12):31–40
11. Jambi K (2018) A survey on the existing arabic optical character recognition and future trends. Int J Adv Res Comput Commun Eng (IJARCCE) 7(3):78–88
12. Kadi M, Nasri M (2019) A modified algorithm for improving the coding of Arabic characters by the Freeman chain code. In: The 2019 international conference on wireless technologies, embedded and intelligent systems (WITS), Fez, Morocco, pp 1–5
13. Lam L, Lee S-W, Suen CY (1992) Thinning methodologies—a comprehensive survey. In: IEEE transactions on pattern analysis and machine intelligence, vol 14, no 9, p 879
14. Benchaou S (2018) Contributions à la reconnaissance des caractères par approches statistiques, structurelles et évolutionnistes. PhD thesis, Mohamed I University, Oujda, Morocco
15. Kadi M, Nasri M (2019) Isolated Arabic characters recognition, using a robust method against translation, noise and scaling, based on the «Hough Transform». Int J Inf Sci Technol (iJIST) 3(4):34–43
16. Boudraa O, Hidouci WK, Michelucci D (2020) Using skeleton and Hough transform variant to correct skew in historical documents. Math Comput Simul 167:389–403
17. Bezmaternykh PV, Nikolaev DP (2020) A document skew detection method using fast Hough transform. In: Twelfth international conference on machine vision (ICMV 2019), vol 11433. International Society for Optics and Photonics, pp 114330J

Performing Spectrum Sensing Using a Deep Learning Algorithm for Cognitive Radio

Omar Serghini, Hayat Semlali, Asmaa Maali, Najib Boumaaz,
Abdallah Soulmani, and Abdelilah Ghammaz

Abstract Cognitive radio network is an intelligent technology, used to solve the scarcity of the radio spectrum by allowing the unlicensed users to have access to the licensed spectrum. We study spectrum sensing methods in cognitive radio networks, the problem in the process of spectrum sensing that the detection rate of the the primary user signal is low in the environment of low signal-to-noise. A spectrum sensing algorithm based on multilayer perceptron is proposed which does not require a prior knowledge of the traffic characteristics of the licensed user. The performance of the proposed model is evaluated in terms of accuracy and loss for different epoch number also in terms of detection and false alarm probabilities. Simulation results have shown that the proposed approach provides good detection compared to the classic technique based on energy detection.

Keywords Cognitive radio · Deep learning · Multilayer perceptron · Spectrum sensing

1 Introduction

The present static spectrum allocation policies have caused spectrum scarcity among the unlicensed user named secondary user (SU) systems and under utilization of spectrum among the licensed user named primary user (PU) systems according to a survey conducted by the U.S. Federal Communications Commission (FCC) [1]. Spectrum sensing is a main function of CR which is introduced to manage the use of the frequency band between PU and SU users.

Many spectrum sensing techniques have been developed: Matched Filter Detector (MFD) [2, 3], Energy Detector (ED) [4, 5], Cyclostationary Feature Detector (CFD) [6] and Eigenvalue Detector [4, 7]. The implementation of these algorithms

O. Serghini (✉) · H. Semlali · A. Maali · N. Boumaaz · A. Soulmani · A. Ghammaz
Laboratory of Electrical Systems, Energy Efficiency and Telecommunications, Department of Physics, Faculty of Sciences and Technologies, Cadi Ayyad University, Marrakesh, Morocco
e-mail: serghiniomar8@gmail.com

requires different conditions, and their detection performance is correspondingly distinguished.

Machine learning (ML) concept is adopted in the cognitive radio, which has opened up new ideas for researching [8]. Most machine learning algorithms view the spectrum sensing processing as a signal classification. Recently, deep learning (DL) has demonstrated its remarkable potential in extracting the hidden structure of different objects in various complicated tasks such as computer vision [9] and wireless communication [10]. In this paper, we propose a MLP Deep Learning algorithm which is applicable for arbitrary types of primary signals. It is worth highlighting that the proposed detector does not require any additional information of the primary signal or noise density when deployed online.

The rest of the paper is organized as follows. In Sect. 2, we present the spectrum sensing principle. In Sect. 3, we describe the Multilayer perceptron algorithm. Our proposed method is numerically demonstrated by computer simulation in Sect. 4, followed by a conclusion in Sect. 5.

2 Sensing Model

Suppose that a CRN is composed of a PU and many Secondary Users (SUs), so for any SU, whether the PU exists or not can be represented as a binary hypothesis-testing problem [11]:

$$
\begin{aligned}
H_1 &: x_n = s_n + \omega_n \\
H_0 &: x_n = \omega_n
\end{aligned}
\tag{1}
$$

where x_n is the received signal, s_n represents the PU signal and ω_n represents the channel noise; in our case, we consider Additive White Gaussian Noise (AWGN), where the mean is zero and the variance is σ_n^2. Also, H_0 and H_1 are the two hypotheses denoting the absence and presence of primary signal in a certain band, respectively.

For evaluating these spectrum sensing techniques, two criteria are considered: the probability of false alarm (P_{fa}) which is defined as the probability where the test gives a wrong declaration about the occupancy of the considered band, and the probability of detection (P_d) which is the probability of declaring that the PU is truly present on the considered band. These probabilities can be defined as follow [2, 12]:

$$
\begin{aligned}
P_{fa} &: \mathrm{Prob}\{T > \lambda / H_0\} \\
P_d &: \mathrm{Prob}\{T > \lambda / H_1\}
\end{aligned}
\tag{2}
$$

where T is the output of the detector which is compared to the threshold λ to make decision.

Fig. 1 MLP architecture [15]
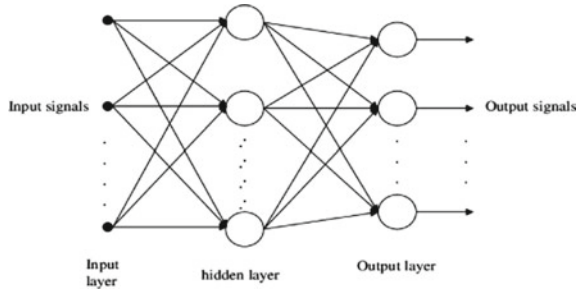
3 Deep Learning-Based Detector

A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN). It contains a series of layers, composed of neurons and their connections. An artificial neuron has the ability to calculate the weighted sum of its inputs and then applies an activation function to obtain a signal that will be transmitted to the next neuron [13]. MLP utilizes a supervised learning technique called backpropagation for training [14]. Its multiple layers and nonlinear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable (Fig. 1).

3.1 Dataset Generation and Preprocessing

For dataset, we generate 8 kinds of digitally modulated signals at SNR = − 12dB. The idea is to obtain the power spectral density of the sampled synthesized signal as the square of the modulus of the Fourier transform, divided by the spectral bandwidth. The power spectral density is [16]:

$$\Gamma_x = \frac{\mid Y_x \mid^2}{T} \tag{3}$$

where y_x is a signal and Y_x its Fourier transform. It represents the frequency distribution of the power of a signal according to the frequencies that compose it. It is used to characterize stationary and ergodic Gaussian random signals and is essential for the quantification of electronic noise. In our case, PSD allows us to characterize the white noise present on the signal and to distinguish it by estimating its power.

The entire dataset is partitioned into three different sets for training, validation and testing (Table 1).

Table 1 Dataset parameters

Modulation scheme	8PSK, BPSK, GFSK, PAM4, QAM16, QAM64, QPSK, WBFM
SNR	-20 to $0\,dB$ in $2\,dB$ increments
Training samples	60,000
Validation samples	20,000
Testing samples	20,000

Table 2 Model hyperparameters

Hyperparameter	Value
Neurone per FC layer	64, 256, 2
Optimizer	Adam
Batch size	100
Dropout rate	0.1

3.2 Network Architecture Design

The chosen layers and its disposition are strictly related to the input. In this case, the input size is 64, which describes features number and the height of the first layer. We use 256 neurons in hidden layer and two in the output layer. For activation function, the output neurons uses Softmax, while all other layers use ReLu. Dropout is also used after every layer to prevent overfitting [17]. The sequential model performs an optimization over its error function, for which the optimizer used can be selected. Adam [18] is an adaptive learning rate optimization algorithm that is been designed specifically for training deep neural networks. The hyperparameters determined through extensive cross-validation are detailed in Table 2.

3.3 MLP Training Process

The supervised training process involves presenting the training data in a patterned format. Each pattern contains an input vector and its corresponding outputs. The network then tries to capture the effects that each input exerts on the outputs by adjusting a randomly initialized weight space to minimize the error between the network prediction and the specified targets. Technically, the weights in the neural network are first initialized with random values. The data flow from the input units in a strictly forward manner to predict network outputs and compare them to measured targets. In the backpropagation process, the error between the predicted and measured outputs is propagated backward from the output layer to modify the network connection
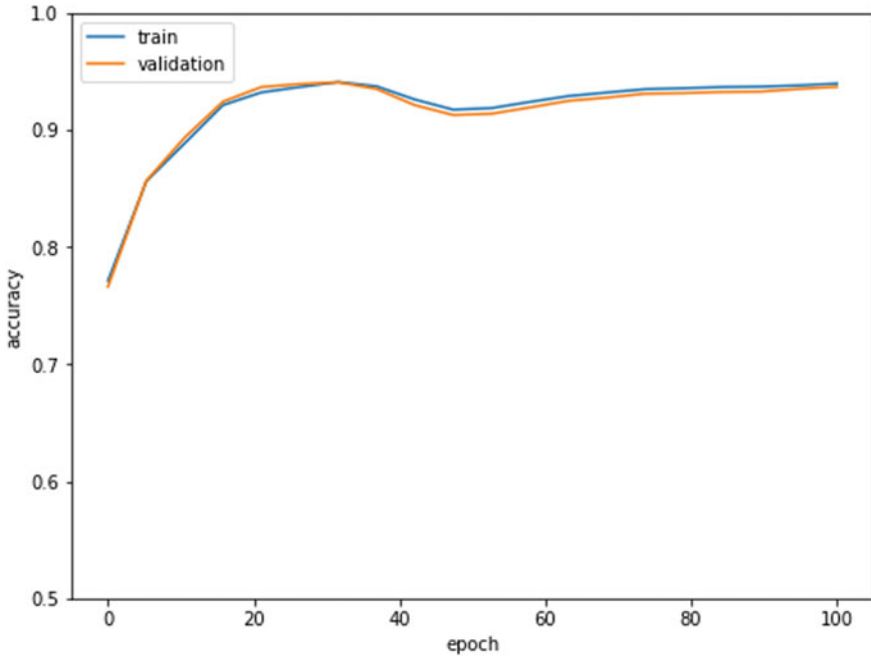
Fig. 2 MLP learning curves "Accuracy"

strengths in order to improve its performance [19]. This iterative process continues until it finds a set of weights that minimizes the system error to a desired value and leads to satisfactory performance of the network.

4 Simulation Results and Discussion

In our simulations, the benchmark algorithm is the energy detector (ED) and the studied deep learning detector MLP.

In the first step, to measure the DL-based algorithm's performance in an interpretable way, we have Accuracy-Loss metrics (Figs. 2 and 3) that show us how well our model is doing. The model's accuracy is expressed as a percentage. It is a metric for how close a model's prediction is to the actual data. Loss is a value that represents the summation of errors in our model, and it is a number indicating how bad the model's prediction was on a single example.
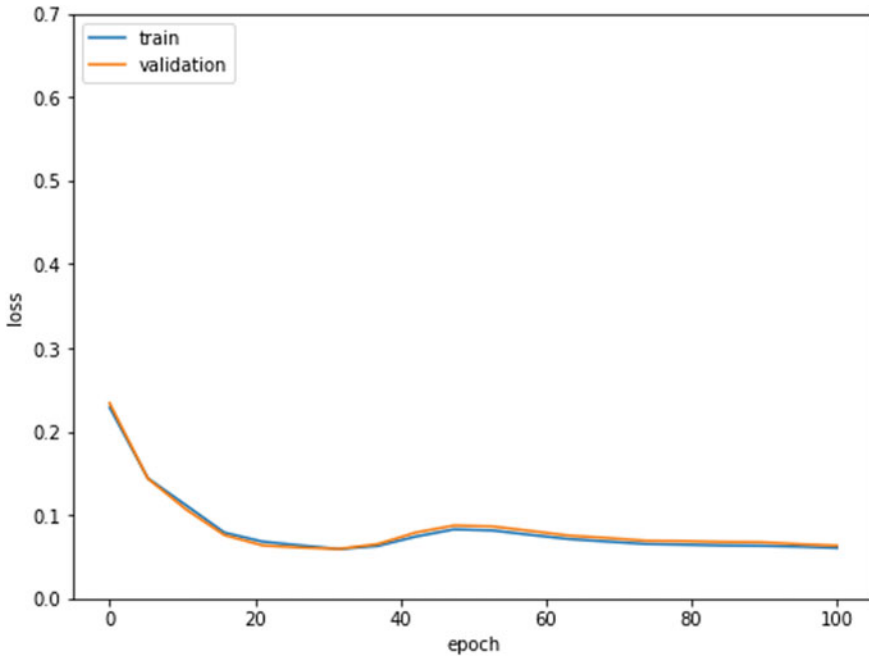
Fig. 3 MLP learning curves "Loss"

We denote that the low values for losses as well as high values for accuracies in the training process was expected, but it is the iteration where they are reached what is significant to be analyzed, as they state the rate of convergence of each of the model. For the training phase, it is also expected that these values are constantly increasing along the whole process, as the model keeps learning from the data and fits to it. For the validation part, we notice that the curves remained converged during training, this is due to the Dropout layer we added to prevent overfitting. Furthermore, the convergence rate and the final accuracy value indicate that the optimizer used is adequate for the kind of the application.

In the second step, to demonstrate the characteristics of the proposed MLP detector under different false alarm probabilities, we compare the receiver operating characteristics (ROC) curves of the proposed detector along with that of the benchmark detector in Fig. 4 and the signal-to-noise ratio (SNR) is set to -12 dB. The signal detection probability obtained by the learning-based detection method is higher compared to the energy detection method in the low SNR, and the performance gain of our method is more obvious under the condition of no prior information on the signal of interest.
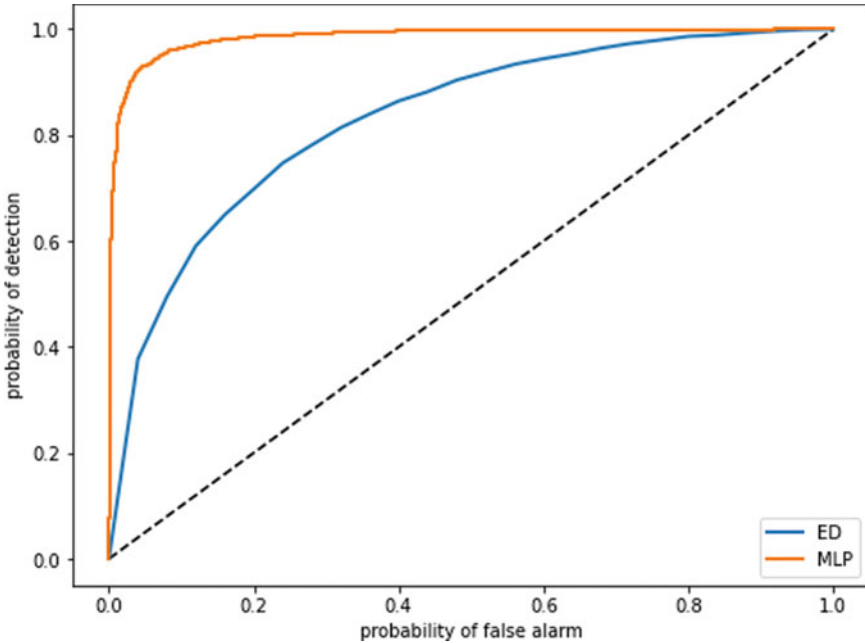
Fig. 4 ROC curves with SNR $= -12$ dB

5 Conclusion

In this contribution, we have proposed a novel DL-based signal detector. It was shown that can be achieved over the energy detector. Also, the DL-based detector is insensitive to the modulation order. From the obtained results, the multilayer perceptron detector can be considered as promising approach for spectrum sensing in CR networks. To conclude, the proposed method can be further improved by increasing the data size and adding other types of modulation. This is beneficial for the classifier to adapt to various environments and improve the generalization ability.

References

1. McHenry MA (2005) NSF spectrum occupancy measurements project summary. Shared Spectrum Company Report
2. Maali A, Semlali H, Laafar S, Boumaaz N, Soulmani A (2019) An overview of the different principal spectrum sensing techniques in cognitive radio systems. Adv Sci Eng Med 11:1–6
3. Brito A, Sebastião P, Velez FJ (2021) Hybrid matched filter detection spectrum sensing. IEEE Access 99:165504–165516

4. Semlali H, Boumaaz N, Soulmani A, Ghammaz A, Diouris J (2010) Energy detection approach for spectrum sensing in cognitive radio systems with the use of random sampling. Wireless Pers Commun 79(2):1053–1061
5. Lorincz J, Ramljak I, Begusic D (2021) Algorithm for evaluating energy detection spectrum sensing performance of cognitive radio MIMO-OFDM systems. Sensors
6. Um J, Spooner CM, Reed JH Cyclostationary approaches to signal detection and classification in cognitive radio. Virginia Tech Mobile and Potable Radio Research Group 432 Durham Hall, Mail Stop 0350, Blacksburg, VA 24061, USA
7. Zeng Y, Koh CL, Liang YC (2008) Maximum eigenvalue detection: theory and application. In: IEEE international conference on communications, ICC'08. IEEE, pp 4160–4164
8. Bkassiny M, Li Y, Jayaweera SK (2013) A survey on machine-learning techniques in cognitive radios. IEEE Commun Surv Tutorials 15(3):1136–1159 (Third Quarter)
9. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
10. Jiang P et al (2018) Artificial intelligence-aided OFDM receiver: design and experimental results. [Online]. Available: https://arxiv.org/abs/1812.06638
11. Digham FF, Alouini MS, Simon MK (2007) On the energy detection of unknown signals over fading channels. IEEE Trans Commun 55(1):21–24
12. Liang Y, Zeng Y, Peh ECY, Hoang AT (2018) Sensing-throughput tradeoff for cognitive radio networks. IEEE Trans Wireless Commun 7(4):1326–1337
13. Popescu MC, Balas VE, Perescu-Popescu L, Mastorakis N (2009) Multilayer perceptron and neural networks. WSEAS Trans Circuits Syst 8(7)
14. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Parallel distributed processing: explorations in the microstructure of cognition, volume 1: foundation. MIT Press
15. Hong Y-M (2017) Feasibility of using artificial neural networks to forecast groundwater levels in real time. Landslides 14(5):1815–1826. https://doi.org/10.1007/s10346-017-0844-5
16. Miller SL, Childers D (2004) Random processes in linear systems. In: Probability and random processes, pp 413–455
17. Sanjar K, Rehman A, Paul A, JeongHong K (2020) Weight dropout for preventing neural networks from overfitting. In: 2020 8th International conference on orange technology (ICOT)
18. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization
19. Tan HH, Lim KH (2019) Review of second-order optimization techniques in artificial neural networks backpropagation. IOP Conf Ser: Mater Sci Eng 495(1):012003

A Retrospective on OOADARE as an Automated Object-Based Approach for Requirements Engineering

Amal Khalil, Hajar Lamsellak, Zineb Bougroun, Mohammed Saber, and Mohammed Ghaouth Belkasmi

Abstract In agile methods, the communication between customer and development team is ensured by requirements, most often presented in an unstructured textual format, which frequently involves redundancy or non-precision information. This leads, in practice, to poor system quality, especially if we use classical approaches such as scenario-based approach. Yet, OOADARE approach is introduced in this way, using semi-structured text models in the form of user stories and constraint story cards (CSC), to automate the object-oriented transformation of requirements into a class diagram. However, the approach failed to capture all the elements needed to construct a correct and complete class diagram. This paper, at that point, proposes templates in natural language, which are part of the same perspectives of CSCs proposed by OOADARE, namely semi-structured text, to fill these gaps and ensure the completeness of the class diagrams thus generated.

Keywords Requirements engineering · Object-oriented application · User stories · Constraint story card (CSC) · Transforming requirements · Class diagram

A. Khalil (✉) · H. Lamsellak · Z. Bougroun · M. Saber · M. G. Belkasmi
SmartICT Laboratory, Mohammed First University Oujda, ENSAO, Oujda, Morocco
e-mail: amal.khalil@ump.ac.ma

H. Lamsellak
e-mail: lamsellak.hajar@ump.ac.ma

Z. Bougroun
e-mail: z.bougroun@ump.ac.ma

M. Saber
e-mail: m.saber@ump.ac.ma

M. G. Belkasmi
e-mail: m.belkasmi@ump.ac.ma

65

1   Introduction

Requirements engineering is the process of eliciting, analyzing, specifying, validating, and managing software requirements [1]. It is not an isolated front-end activity to a software lifecycle process; rather, it is an integral part of the larger process connected to other parts through continuous feedback loops [2]. The scenario-based approach to expressing requirements is the approach most used by developers [3, 4]. After eliciting requirements, they are structured in the form of user stories, use cases, sequence diagram, etc. Those forms are not automatically mapped to the object model [5]. object-oriented analysis and design approach for the requirements engineering (OOADARE) approach, which is part of the design approach trend aimed at automating the generation of analysis diagrams [6–8], fits to overcome these limits, so it took user stories and constraint story cards (CSC) to automatically generate the objects, their actions, and the rest of the software artifacts (MVC, DAO, Test…) [9].

The purpose of this study is to follow the same approach to ensure its completeness and therefore determine the remaining gaps, which should be filled in order to advance the current research concretely.

2   OOADARE Approach

2.1   Presentation

A sequence of transformations is performed to build a software system, starting from requirements and ending with implementation. However, requirements are mostly in the form of an unstructured text, but not a model that can be easily understood by computers [10]. OOADARE is an approach based on the OOAD approach evolved by Booch [11], which requires, among its constraints, the use of a semi-structured text to extract structured models [12]. OOADARE contributes to the automatic transformation of requirements into a class diagram and offers technical concepts and practices that facilitate the development flow and improve the quality of the software product [5, 9]. It introduced a template that is based on the well-known user story template for solving the disorganization issue [9].

2.2   Overview of the "User Story" Template

After retrieving the requirements, the OOADARE approach requires preparing them in user stories to move on to identifying objects and methods [5].

User stories are a popular method for representing requirements using a simple template [13]. Their adoption is growing [14] and is massive especially in the context of agile software development [15].
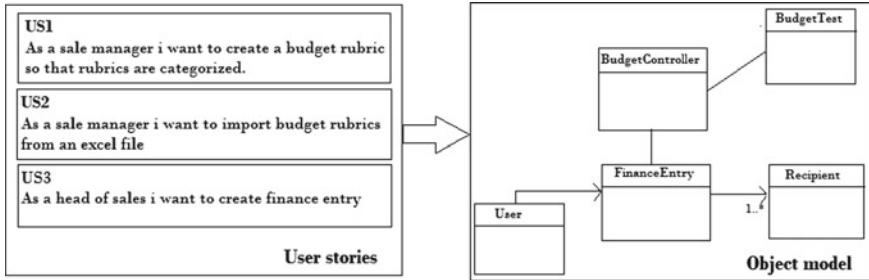
Fig. 1 Using "user stories" to generate an "object model" [9]

The OOADARE approach takes "user stories" as an entry point to automate the identification of objects and actions of the object model, which implies respecting a very precise syntax [9]:

$$As\,a\ <\ role\ >\ I\ can\ <\ \boldsymbol{action}\ >\ <\ \boldsymbol{object}\ >\ so\,that\ <\ business\,value\ >$$

The diagram thus generated specifies the classes of objects and the methods (Fig. 1).

The "user stories" only allow to generate the classes, and the methods of a class diagram, the "constraint story card" (CSC), are then used to manage the associations.

2.3　Using CSC to Generate Associations

Constraint story card (CSC) is a story card in which we can write the requirement's constraints in a human-like object constraint language (OCL) [16], that represents a precise text language that provides constraint and object query expressions on any meta-object facility model or meta model with a formal specification language [17].

The OOADARE approach suggests four templates to report the different associations within the UML class diagram.

The four templates use a generic formula as shown below [18]:

$$[Role]\ <\ \boldsymbol{Source\ Object}\ >\ <\ \boldsymbol{Association\ expression\,type}\ >\ [\boldsymbol{Cardinality}]\,[Role]$$
$$<\ \boldsymbol{Target\ Object}\ >$$

The "Association expression type" argument corresponds to four different values depending on the type of the association [18]:

- [**Has or a verb**]: in this case, it is a simple association
- [**Is a**] is the proposed model for inheritance
- [**Contains**]: to illustrate aggregation between classes
- [**Is composed of**]: for the composition model.
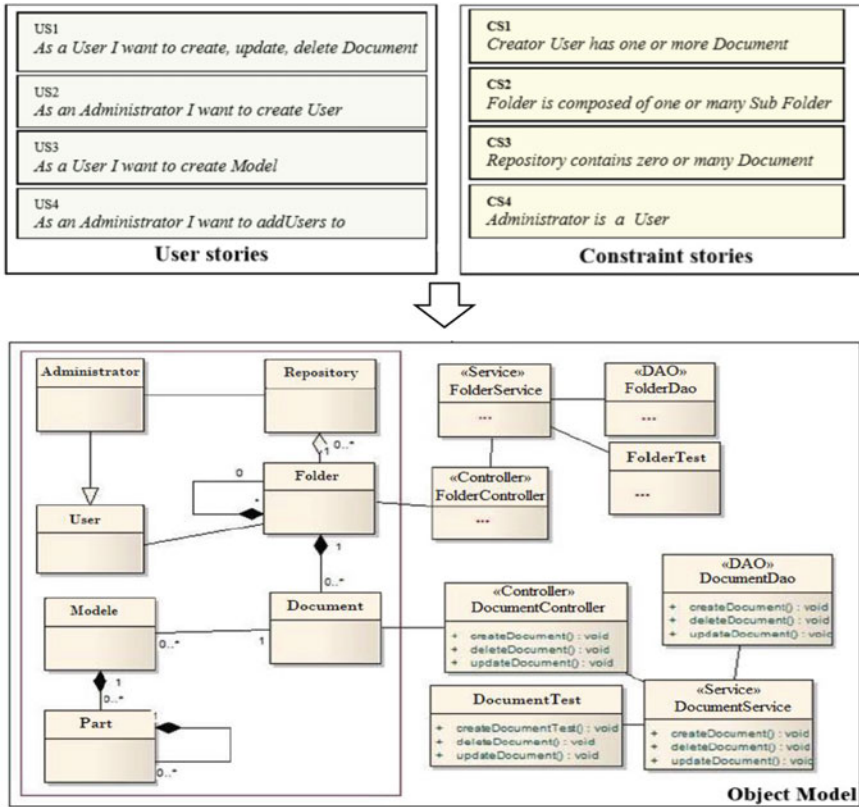
Fig. 2 Class diagram generated from user stories and CSC using the OOADARE approach. *Other artifacts (DAO, Controller …) generated were omitted for a clarity purpose.* [18]

The "Cardinality" can be expressed in ***letters*** (from zero to ten in English), ***numbers*** or by the word "***many***". When the minimum and the maximum are not the same, the two sides are separated by the word "***or***" (Fig. 2) [18].

Nevertheless, the mapping model proposed by OOADARE, in compliance with the constraints and templates, remains incomplete since it does not take into consideration all the elements that can be found in specifications and which are essential for software development.

3   Filling the Gap in the OOADARE Approach

As mentioned above, important elements are missing in the class diagram generated using OOADARE, which take their place in the center of the requirements and without which we cannot speak of a complete class diagram. It lacks, first,

attributes that represent the variables storing state information characterizing the object. Secondly, we notice the absence of association classes in which "attributes" are directly dependent on two other classes and could not be placed in either one or the other. We propose in the following part solutions to fill this gap and ensure the completeness of the method with indispensable features.

3.1 Generating Attributes in a Class Diagram

In domain modeling class diagrams, an attribute represents a data definition for an instance of a classifier. An attribute describes a range of values for that data definition. A classifier can have any number of attributes or none at all. Attributes describe the structure and value of an instance of a class [19].

As a part of CSC, and taking into account textual expressions often used to present the different properties related to an object, we consider the following model to extract the different attributes of a class:

[Each] < Object > has| is identified by| is represented with| is described by < Attribute 1 > , < Attribute 2 > , …and < Attribute n >

where

- **Object** represents the entity for which the properties are identified
- **Attribute *n***: the nth single, named fragment of the persistent state.

3.2 Consider Association Classes in Class Diagram Generation

In UML diagrams, an association class is a class that is part of an association relationship between two other classes. It is identical to other classes and can contain operations, attributes, as well as other associations [20]. We can attach an association class to an association relationship to provide additional information about the relationship [17].

This case occurs in a "many to many" association, so we must take into account this type of association before looking for the association attribute(s) through the template that we propose as CSC (Table 1):

Table 1 Formula parts description

Formula argument	Description
Source object	The source object of an association
Attribute	The single, named fragment of the persistent state
Target object	The target object of an association

*[**Each**] < Source Object > **has| have** < Attribute > (**for| in) each** < Target Object >*

3.3 Example of Application to a Case Study

Consider the following set of requirements for a university information system that is used to keep track of student's transcripts: *"…Each department is described by a name, a code, an office number, an office phone and a college. Each course has a course name, description, course number, credit hours, level, and offering department. Each student is identified by a number, a name, a birthdate, a gender, a social security number, current address and phone number. To ensure that the system generates a transcript for students, each student must have a grade for each course.…"*.

To perform the extraction task of UML class diagram concepts, each sentence in the input text will be matched against the different lists of patterns that have been defined by the above models.

If we analyze the sentence "*Each department is described by a name, a code, an office number, an office phone and a college.*', we notice that it matches very well with the attributes pattern (Fig. 3); it begins with a "number of characters" that mentions the concerned "object," followed by the key words "is identified by," then followed by "a number of characters" that list the specified attributes, which gives us the resulting class "Department" with its correspondent attributes.

Then, considering the case of the last sentence presented by the example: '*To ensure that the system generates a transcript for students, each student must have a grade for each course.*', we note its correspondence to the second pattern proposed for the generation of association classes (Fig. 4).

It should be noted that in this case, the association between the two classes (student and course) has on both sides a maximum multiplicity equal to "many" represented by the symbol "*", which means that a "student" studies one or more "courses" and that a "course" is taught for one or more "students".
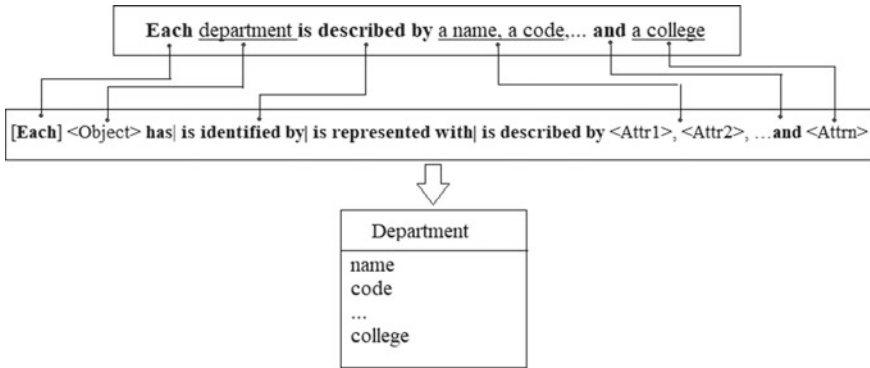
Fig. 3 Diagram class illustrating an entity with attributes as mapped from a CSC template
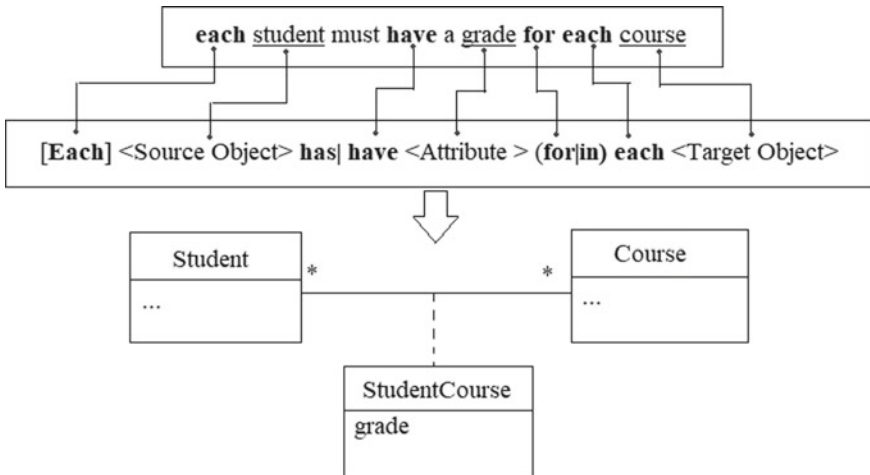
Fig. 4 Class association extraction pattern example

4 Conclusion

More efficient and more flexible technologies are accelerating the growth of fully automated production facilities. This is the context of our study which aims to improve the OOADARE, an approach targeting the automatic generation of class diagrams from requirements presented as semi-structured text templates.

In this paper, we started by presenting the general context of the study, then explaining the approach used by OOADARE which offers an intermediate representation of the requirements in the form of user stories and CSC, thus allowing an automatic generation of class diagrams. Although we discussed the failures of the approach to finally propose solutions that follow the same rules and constraints.

With the aim of eliminating ambiguity and redundancy, the templates proposed by OOADARE in addition to the models proposed in this document are relatively simple to build and to deduce from the requirements expressed either in textual or oral form.

References

1. Lucassen G et al (2016) The use and effectiveness of user stories in practice. In: International working conference on requirements engineering: foundation for software quality. Springer, Cham
2. Macaulay LA (2012) Requirements engineering. Springer Science & Business Media
3. Sutcliffe A (2003) Scenario-based requirements engineering. In: Proceedings of the 11th IEEE international requirements engineering conference, 2003. IEEE
4. Kim M et al (2007) Managing requirements conflicts in software product lines: a goal and scenario based approach. Data Knowl Eng 61(3):417–432
5. Zeaaraoui A, Bougroun Z, Belkasmi MG, Bouchentouf T (2012) Object-oriented analysis and design approach for the requirements engineering. J Electr Syst 2(4):147–153
6. Elallaoui M, Nafil K, Touahni R (2018) Automatic transformation of user stories into UML use case diagrams using NLP techniques. Procedia Comput Sci 130:42–49
7. Fischbach J, Frattini J, Vogelsang A (2021) Cira: a tool for the automatic detection of causal relationships in requirements artifacts. arXiv Preprint. arXiv:2103.06768
8. Yang Y et al (2019) RM2PT: a tool for automated prototype generation from requirements model. In: 2019 IEEE/ACM 41st international conference on software engineering: companion proceedings (ICSE-Companion). IEEE
9. Zeaaraoui A, Bougroun Z, Belkasmi MG, Bouchentouf T (2013) User stories template for object-oriented applications. In: Third international conference on innovative computing technology (INTECH 2013)
10. Yue T, Briand LC, Labiche Y (2011) A systematic review of transformation approaches between user requirements and analysis models. Requirements Eng 16:75–99
11. Booch G et al (2008) Object-oriented analysis and design with applications. In: ACM SIGSOFT software engineering notes, vol 33, no 5, pp 29–29
12. Sanyal R, Ghoshal B (2018) Automatic extraction of structural model from semi structured software requirement specification. In: 2018 IEEE/ACIS 17th international conference on computer and information science (ICIS). IEEE
13. Cohn M (2004) User story applied: for agile software development. Addison-Wesley, Boston, MA
14. Kassab M (2015) The changing landscape of requirements engineering practices over the past decade. In: Proceedings of EmpiRE. IEEE, pp 1–8
15. Wang X, Zhao L, Wang Y, Sun J (2014) The role of requirements engineering practices in agile development: an empirical study. In: Zowghi D, Jin Z (eds) APRES 2014, vol 432. CCIS. Springer, Heidelberg, pp 195–209
16. Warmer J, Kleppe A (2003) The object constraint language—getting your models ready for MDA, 2nd edn. Addison Wesley
17. OMG (2017) Unified modeling language specification version 2.5.1
18. Dahhane W et al (2014) An automated object-based approach to transforming requirements to class diagrams. In: 2014 Second world conference on complex systems (WCCS). IEEE
19. IBM Documentation. Rational software architect standard edition (Version 7.5.0)
20. IBM Documentation. Rational software modeler (Version 7.5.0)

Computational Analysis of Human Navigation Trajectories in the VR Magic Carpet ™ Using *K*-Means

Ihababdelbasset Annaki , **Mohammed Rahmoune, and Mohammed Bourhaleb**

Abstract In this research, we use unsupervised machine learning clustering techniques, notably *K*-means (Jain in Pattern Recogn Lett 31:651–666, 2010 [1]), to explore human navigation using the VR Magic Carpet (Berthoz and Zaoui in Dev Med Child Neurol 57:15–20, 2015 [2]). This is a variant of the Corsi Block Tapping task (CBT) (Corsi in Human memory and the medial temporal region of the brain. McGill University, 1972 [3]) that was carried out within the experimental framework of virtual reality. The participant's trajectory was captured as raw spatial data and afterward kinematically evaluated. Our previous research (Annaki et al. in Digital technologies and applications. ICDTA 2021. Lecture notes in networks and systems, vol 211. Springer, Cham, 2021 [4]) found three distinct groups. However, the classification remained unclear, suggesting that they include both types of people (ordinary and patients with cognitive spatial impairments). Based on this premise, we used *K*-means to distinguish patients' navigation behavior from that of healthy people, highlighting the most significant differences and validating the feature on which our previous analysis was based.

Keywords Artificial intelligence · Machine learning · Virtual reality · Neuropsychological assessments · Cognitive impairments

I. Annaki (✉) · M. Rahmoune · M. Bourhaleb
Embedded Systems, Electronics, and IT Team, Research Laboratory in Applied Sciences, National School of Applied Sciences, P.B. 669, 60000 Oujda, Morocco
e-mail: i.annaki@ump.ac.ma

M. Rahmoune
e-mail: m.rahmoune@ump.ac.ma

M. Bourhaleb
e-mail: m.bourhaleb@ump.ac.ma

1 Introduction

Neuropsychological assessments (NPAs) are frequently used to diagnose cognitive deficits (CIs) [5]. In the pursuit of identifying a stereotypical model of human navigational behavior, researchers from a wide range of disciplines, including mathematics, robotics, neuroscience, and computer science, are interested in analyzing computational locomotor NPAs [6–10].

In conjunction with a plethora of navigational data, the development of digital hardware has allowed neuroscientists to analyze evaluation findings using statistical methods that, although helpful, are still sufficient for identifying classifiers that define human navigation behavior [11].

The advent of recent artificial intelligence technology has provided a compelling and diverse set of tools for producing more dependable results. A significant contribution has been made by clustering algorithms such as K-means, which have allowed for the development of insightful clusters based on an analysis of NPAs from their kinematic relationship to locomotion outputs [12].

The primary emphasis of this study is an examination of data obtained from the VR Magic Carpet [2], a virtual reality-based variant of the Corsi block tapping test. Our early data analysis[4], which included statistical and kinematic characteristics such as average speed and timestamps, created the classes.

We then used the K-means machine learning algorithm, a data partitioning technique and a combinatorial optimization problem [1] to gain additional insights into the classes discovered in previous work and to delve deeper into the middle classes that remained ambiguous, implying that they contain both types of participants (ordinary and patients with cognitive spatial impairment), because we'd like to be able to distinguish patients from normal individuals by displaying the main differences based on our K-means clustering results.

2 Materials and Methods

2.1 The Virtual Carpet ™

Pr. Alain Berthoz and Eng. Mohamed Zaoui developed and used a virtual reality adaptation of the Magic Carpet [2]. It is a Corsi Block Tapping task (CBT) [3] variant that adds locomotion and displays the sequence via tile illumination rather than supervisor interactivity in a virtual reality environment developed in unity, like in the Walking Corsi test [13], where an experimenter displays the sequence or the targets that the subject should visit and where the participant is expected to replicate the sequence in the same order.

2.2 Experimental Data

The data was gathered from a pilot test by Dr. Bernard Cohen with the help of unity c scripts and the HTC Vive VR Kit and SteamVR tracking system. The following items are included in the provided:

- The calibration data files contain the locations of the different tiles.
- The session data, which includes participant coordinates collected at the time of the clinical study session, are saved in a folder labeled with the participant's id, including at least ten session data files.

2.3 Participants

To prevent any prior bias in the research, all data were collected anonymously, from gender and age to the state of the participants. The protocol had no side effects on the participants and was approved by the ethical committee.

2.4 Data Preprocessing

We created an algorithm to keep track of the targets that the users visited. This was accomplished by calculating the walking speed and waist positions during the session (tangential velocity). The position data were compared to a target calibration file that included the position of the centers of nine targets in determining whether the patient had reached a target. As a result, we could compute the participant's time spent on the target in seconds. Each dataset in the user folder corresponds to a particular set of tiles.

Each file is stored as a data frame using the "Pandas" data analysis and manipulation tool package to avoid file conflicts. The technique works by comparing the participant's position on the trajectory to the target's location in the calibration file. The software also monitors the participant's kinematic behavior until the participant approaches the goal or the session terminates.

2.5 Visual Replication

Based on our theoretical assumptions, we utilized Matplotlib, a robust framework for generating static, interactive, and immersive Python graphics. We double-checked our findings using a visual representation of the sequence upon which they were based. The use of animation to reproduce the trajectory and targets allowed us to
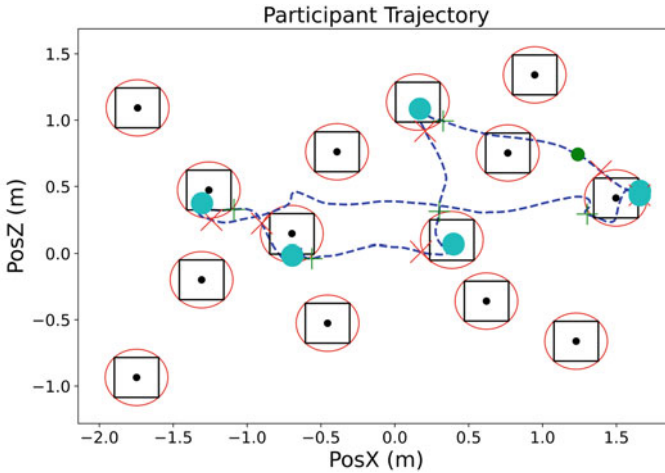
Fig. 1 Visual replication of a session using Python

better understand the participants' kinematic behavior during each session of the experiment (see Fig. 1).

The participant's trajectory is in blue, the confidence area is in red and the starting point is in green. When a tile is visited, it is marked with a cyan dot as the closest point from the trajectory to the center of the target, a red mark when the participant leaves the target's confidence area and a green plus mark when the target enters the target's confidence area.

2.6 K-Means: Unsupervised Machine Learning

After completing the data preparation, we find that our dataset lacks categorical data. As a result, using an unsupervised machine learning clustering technique, such as K-means clustering, is a viable option. For the algorithms in this section's implementation, we mostly utilized the Python machine learning software "Scikit-learn". **Algorithm** The main steps of the K-means algorithm are as follows [1]:

1. Select an initial partition with K clusters; repeat steps 2 and 3 until the cluster membership stabilizes.
2. Generate a new partition by assigning each pattern to its closest cluster center.
3. Compute new cluster centers.

K-means parameters The user needs to provide three parameters for the K-means algorithm: the number of clusters K, cluster initiation and distance metric.
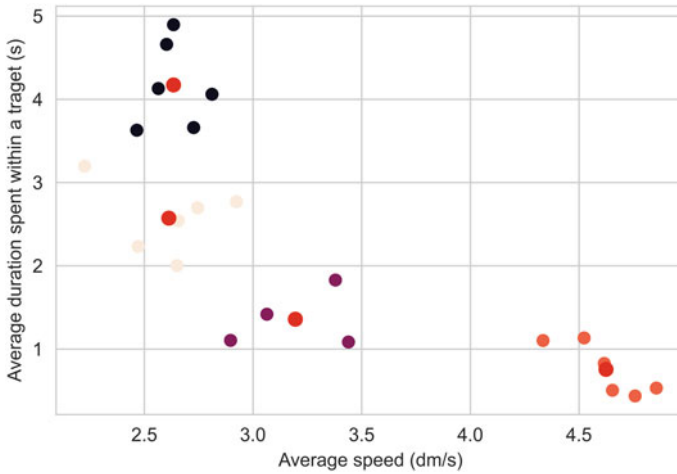
Fig. 2 *K*-means clusters (i.e., *K* = 4 clusters), indicating the difference between different groups of participants

3 Results

Several clusters were discovered in this study. We emphasize that the formed clusters are based on variables that have been analyzed and aggregated. Based on a neuroscientist's examination of the NPA, the primary goal of this clustering is to gain insight into potentially interesting profiles or clusters that can be associated with clinical categories. These findings pique the interest of neuroscientists, and average speed and time spent within the target could be validated as early classifiers that require further refinement. These findings are part of the process of identifying important features that may aid in the development of a stereotypical model of human navigation (Fig. 2).

4 Discussion

In our previous work [4], we were able to use a kinematic approach to construct groups or categories and characterize the different actions of individuals who went through the VR Magic Carpet. However, those categories were entirely based on average speed, and we discovered that by including the average time spent, we can obtain some anomalies that rendered our initial analysis insufficient (based on expert validation). We chose *K*-means as an unsupervised learning algorithm because it is simple to implement and guarantees convergence. In the perspective of comparing it to DBSCAN [14] and HAC.

5 Conclusion

The current study is an ongoing effort to describe human behavior in a complex visuospatial assessment intended to assess memory and navigation skills. The goal is to include a new analytics method into neuropsychological assessment interpretation so that neuropsychologists may use machine learning applications. Our research will focus more on identifying precisely the K number of clusters associated with the primary clinical categories and evaluate the results of other unsupervised machine learning techniques.

References

1. Jain AK (2010) Data clustering: 50 years beyond K-means. Pattern Recogn Lett 31(8):651–666. https://doi.org/10.1016/j.patrec.2009.09.011
2. Berthoz A, Zaoui M (2015) New paradigms and tests for evaluating and remediating visuospatial deficits in children. Dev Med Child Neurol 57(Suppl 2):15–20. https://doi.org/10.1111/dmcn.12690
3. Corsi PM (1972) Human memory and the medial temporal region of the brain. McGill University
4. Annaki I et al (2021) Computational analysis of human navigation trajectories in a spatial memory locomotor Task. In: Motahhir S, Bossoufi B (eds) Digital technologies and applications. ICDTA 2021. Lecture notes in networks and systems, vol 211. Springer, Cham. https://doi.org/10.1007/978-3-030-73882-2_22
5. Tosi G, Borsani C, Castiglioni S, Daini R, Franceschi M, Romano D (2020) Complexity in neuropsychological assessments of cognitive impairment: a network analysis approach. Cortex J Dev Study Nerv Syst Behav 124:85–96. https://doi.org/10.1016/j.cortex.2019.11.004
6. Bennequin D, Berthoz A (2017) Several geometries for movement generation. In: Laumond J-P, Mansard N, Lasserre J-B (eds) Geometric and numerical foundations of movements. In: Springer tracts in advanced robotics, vol 117. Springer, pp 13–42
7. Taylor M, Bashkirov S, Rico JF, Toriyama I, Miyada N, Yanagisawa H, Ishizuka K (2021) Learning bipedal robot locomotion from human movement. In: 2021 IEEE international conference on robotics and automation (ICRA). https://doi.org/10.1109/icra48506.2021.9561591
8. Hicheur H, Pham QC, Arechavaleta G, Laumond JP, Berthoz A (2007) The formation of trajectories during goal-oriented locomotion in humans. I. A stereotyped behavior. Eur J Neurosci 26(8):2376–2390. https://doi.org/10.1111/j.1460-9568.2007.05836.x
9. Bucklin MA, Wu M, Brown G, Gordon KE (2019) American Society of Biomechanics Journal of Biomechanics award 2018: adaptive motor planning of center-of-mass trajectory during goal-directed walking in novel environments. J Biomech 94:5–12. https://doi.org/10.1016/j.jbiomech.2019.07.030
10. Ribeiro FS, Santos FH, Albuquerque PB (2019) How does allocation of emotional stimuli impact working memory tasks? An overview. Adv Cogn Psychol 15(2):155–168. https://doi.org/10.5709/acp-0265-y
11. Huang R, Nikooyan AA, Xu B et al (2021) Machine learning classifies predictive kinematic features in a mouse model of neurodegeneration. Sci Rep 11:3950. https://doi.org/10.1038/s41598-021-82694-3
12. Nowakowski K, Carvalho P, Six JB, Maillet Y, Nguyen AT, Seghiri I, M'Pemba L, Marcille T, Ngo ST, Dao TT (2021) Human locomotion with reinforcement learning using bioinspired reward reshaping strategies. Med Biol Eng Comput 59(1):243–256. https://doi.org/10.1007/s11517-020-02309-3

13. Bartonek Å, Piccardi L, Guariglia C (2020) Topographical working memory in children with cerebral palsy. J Mot Behav 53(2):200–208. https://doi.org/10.1080/00222895.2020.1748861
14. Annaki I et al (2022) Computational analysis of human navigation in a VR spatial memory locomotor assessment using density-based clustering algorithm of applications with noise DBSCAN. In: Digital technologies and applications. ICDTA 2022. Lecture notes in networks and systems, vol 455. Springer, Cham.https://doi.org/10.1007/978-3-031-02447-4_20

The Management of Approaches in the Decentralized Architecture of the PBM

Essaid Ammar

Abstract The quality of service contract between the various stakeholders mainly aims to facilitate the presentation of information, and this interpretation allows on the one hand an easy to understand human reading, and on the other hand a possible processing by machines. After the development of these two input models, we deduce our output ontology which makes it possible to describe the agreements of the SLA, agreements make it possible to guarantee a certain quality of service to customer requests. In this paper, this artificial intelligence has required more sophisticated information management, hence the need for a global management and supervision approach that supports all these criteria (diversity of equipment, multimedia flow, complex system, user mobility, Terminal mobility, etc.). Network managers (operator, access provider, etc.) have understood that good management means mastering all the software and hardware resources of the network, hence the possibility of guaranteeing a certain quality of service for each customer range, which subsequently results in more agreement retention and customer gain.

Keywords SLA · Modeling of the GDP base · PBM · Ontologies · Semantics · QoS

1 Introduction

SLA or a quality of service commitment by a contract which represents all the clauses and rules that the supplier must respect and apply concerning a specific customer, this contract also makes it possible to define the supplier penalties in the event of non-respect of the n one of these contract rules, customer data flows with a certain quality of service are measured based on predefined parameters, such as response time or service duration. These criteria are applicable to both customer and operator stakeholders [1].

E. Ammar (✉)
Faculty of Law, Economic and Social Science Aîn Sebaâ Casablanca, University of Hassan II Casablanca, Casablanca, Morocco
e-mail: essaid.ammar@gmail.com

The SLA is not a concept reserved for a specific domain or any product, the establishment of this type of contract passes in the majority of the cases by a manual signature, but the implementation of this quality of service agreements generally follows a language like WSLA, WSOL, and WS Agreement.

We then represent an overview of these different models and existing works for modeling an SLA, and each model of which presents a certain semantic description.

After this study, we found that the concept of quality of service defines a broad field of research, and we estimate that the guarantee of a better quality of service for a customer passes mainly by the satisfaction of its requests according to the contract negotiated with the provider of this service [2]. We notice at the same time that the existing approaches for the definition of agreements between the supplier and the customer prove a certain difficulty (use of simple variables, limitation on variable clauses defined previously, use of a different expression language) [3].

For the management of the computer network, the PBM offers a centralized structure through the policies, with an informational model in the form ASN.1 for the memorization of these policies, characteristics which prove a limitation in the face of the development of the Internet, a limitation which comes from the abstract form of information, which is not flexible enough to be adaptable to all NET applications, indeed the classes used have a limited scope because of the predefined types as well as heavy load circulation due to client and central server negotiation [4]. This work has allowed the development of a new dynamic, controllable, and smarter architecture, and we can subdivide the contributions according to three axes:

- The dynamism of this approach is guaranteed thanks to the decentralization of tasks toward the client, and the deployment of the semantic web for the definition of the service contract.
- Client access control in the network is guaranteed by the logical entity used to contain the policy rules.
- An intelligence through an informational model based on the ontology for the memorization of management policies.

2   General Objective of an Ontology

The notion of ontology has a strategic purpose, it comes to represent knowledge and data according to a form different from the standard, and it is heavily used in information retrieval, the semantic web, and e-commerce, and everything another area that requires some information engineering according to a specific context.

The first definition of the word ontology comes from the philosophical field. It is specified in his books that it is the science of being, with time and the evolution of the fields, and it has not ceased to take many definitions, for example, defines ontology as an explicit specification of a conceptualization. We retain by this that the notion of ontology allows the superposition of a domain according to a model in the form of concepts and constraints (Fig. 1).
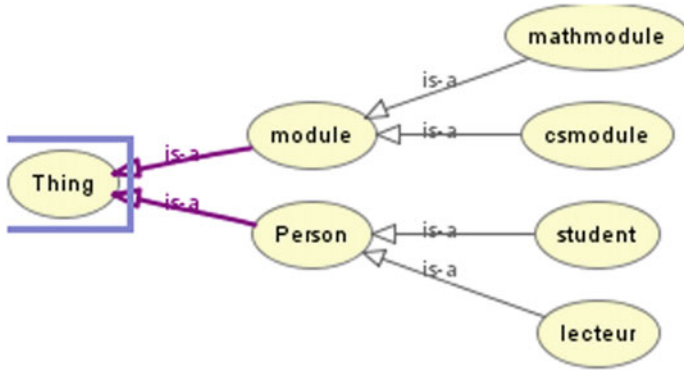
Fig. 1 Graphical presentation of the study ontology

3 Travaux existants pour la gestion des accords

The life cycle of an SLA mainly goes through 5 phases: development of an SLA model, negotiation, implementation, execution, and termination, and we notice that the SLA model takes an important place for the good management of the quality of service customer.

Several main models relating to the SLA have emerged, and the modeling of the QoS takes a significant margin in the description of the contracts. For these reasons, we present in the following a definition of the models which are based on ontologies for the determination of the quality of service.

3.1 QoSOnt

The QoSOnt model is an ontology defined by the OWL language which takes into consideration the notion of QoS according to a structure in three layers (Fig. 2).

This model offers richness for the definition of metrics and requirements related to quality of service, with possible interoperability with other ontologies and models. However, metrics use XML, a form of presentation that proves a great limitation for semantic interpretation to deduce information.

Fig. 2 Structure of the QoSOnt model
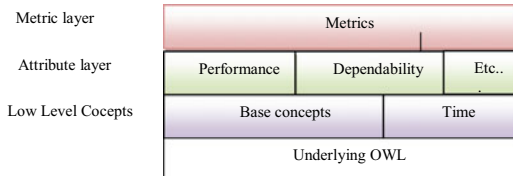
3.2   Ws-Qos

The WS-QoS framework is another ontology-based model that allows the dynamic selection of web services while respecting the QoS requirements as well as the capacity of the network, to achieve its goal, and it is based on three elements:

- QoSInfo describes data on QoS protocol server capabilities and performance.
- WSQoSOntology represents QoS parameter definitions.
- QoSOfferDefinition groups one or more elements of QoSInfo.

The major drawback of this model is the loss of the capabilities offered by the OWL language due to the use of XML language for the presentation of metrics.

3.3   Fipa QoS

FIPA QoS offers using an ontology a method-rich vocabulary for QoS, and it is intended for the community of agents for queries and information retrieval through a request, or in a direct way, or regular using a subscription by an interaction protocol. This model remains specific to the lower layers of the OSI standard.

3.4   Sl-Ontology

SL-Ontology is another model that adopts QoS specifications through a pyramid structure (Fig. 3).

This model defines a semantic presentation which groups together in its structure: supplier offers, customer requests, and stakeholder agreements [5]. This solution as we have defined it brings together supply and demand in the same form, which limits the scope for customer expression.

All the models we have developed before and others offer advantages and prove limitations, some use XML for the definition of methods and quality of service parameters to establish SLA agreements, and this form of presentation is unsuitable

Fig. 3 Structure of the SL-ontology model
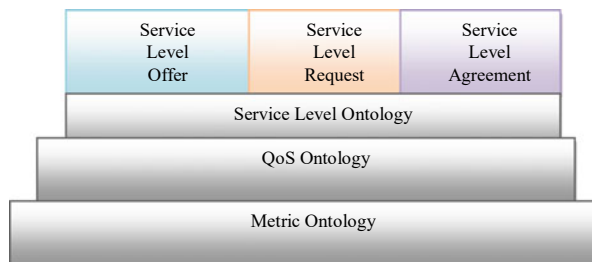
for semantic management of these agreements. Other models have a limited scope of work on specific layers (e.g., FIBA) or do not include logical definitions for quality of service.

We find that a semantic definition of operator offers and another that represents customer requests in a separate way is more practical, if not optimal for building an output ontology for SLA agreements. Each of the speakers will have the opportunity to describe these objectives in their own terms and in a language understandable to everyone. The ontology of these agreements provides automation to facilitate the analysis of information.

4　Alignment of the Two Ontologies

We now develop two ontologies, the first describes the customer intentional model, and another presents the supplier's offers. Each of these ontologies determines its own objects and relationships in its own way and uses different vocabulary from each other.

The passage from one ontology to another, seeing them merge together which is the case in our approach, reflects a certain semantic and linguistic complexity, knowing that the two ontologies are created separately and reflect the thoughts of each participant; therefore, the characteristics of the two ontologies are more or less heterogeneous.

To solve this problem, it becomes necessary to manage the correspondences between the two ontologies, and this task of extreme difficulty has given rise to several approaches and techniques based on algorithms with different visions [6] (Fig. 4).

During our research, we were faced with a choice, which is to follow one of two paths. The first is to choose among existing techniques and works that are based on

Fig. 4 Principle of alignment of the two ontologies
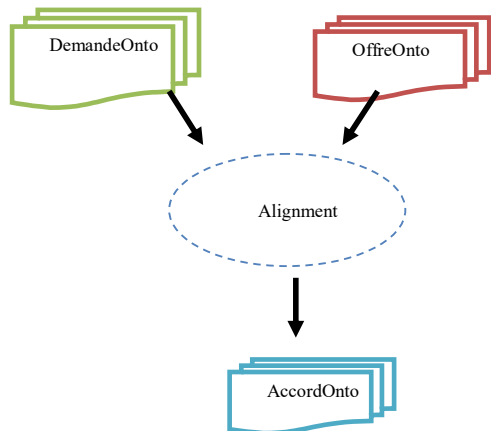
different criteria to manage this heterogeneity or to develop our own method and alignment algorithm.

Among the methods based on character strings which are based on the comparison and the use of synonyms. Linguistic techniques look for potential correspondences, or techniques take internal structures into consideration, and other techniques (extensional, semantic, relational structures, etc.) vary according to their effectiveness.

Regarding the second choice, we find that the development of a new approach is not an optimal solution in our case. We were able to establish two ontologies with very significant structures, and in the same community, the concepts used belonging to the same domain.

It should be noted that with all the satisfactory progress, alignment remains an impossible task in the face of the varieties of developers in a wide field.

5 Semantic Description of the Service Contract

After the development of the two input models, customer intentional model (DemandeOnto) and the supplier ontology model (OffreOnto), we were able to establish an AccordOnto output model based on all of these two models while respecting our first objective, which is the respect of quality of service agreements. This proposal offers different properties and functions allowing the design of a contract respecting the customer requirements.

AccordOnto is based on the semantic definition offered by WSLA. The concept "Parties" defines the participants in a service contract. The supplier and the customer represent the signatory parties, while the third party brings together the parties responsible for monitoring and measuring the quality of service in a contract, the guarantor of the reliability of this agreement can be sponsored by one or two members of the contract (supplier and customer) [7]. Both classes inherit from the Party class which encompasses a set of common components for each participant in the contract.

One of the main concepts necessary for the control of service contracts is control, and it is an essential class which will have a vision on all the obligations, as well as the policy concept which contains all the predicates and rules which manage the agreement. The control entity uses the results of the "Measuring operation" class which returns values to a catalog of functions and operations, and it makes it possible to make the control of these agreements more dynamic and practical.

We find that the obligation part represents all the constraints to be respected in order to have a quality of service agreement, and the contract established between the parties requires the respect of certain obligations (threshold, mobility, price, guarantee, etc.). The threshold concept represents a limitation to offer concerning a parameter, so the provider must strip these resources to respect this threshold. As for the price concept, it describes a very important component in each contract with or without QoS between a customer and the supplier.

The concept of "guarantee" describes the service provider's commitment to the customer's request. In the case of abuse, or non-compliance with one of the clauses established in the contract, a penalty is defined using the penalty concept. In this case, each penalty is recognized as such. The service can be mobile in relation to a user or the terminal, in which case it is signaled with the mobility concept.

6 Decentralized Architecture of the PBM, and a Semantic Definition of the Information Table

The elements of the PBM (management entity (PDP), execution entity (PEP), communication protocol (COPS)) soliciting our interest intervene at the level of the automation of the management process, in a way to manage the available resources, in order to be able to offer the customer the service levels or rather the web service negotiated during a contract. On the other hand, we have been able to make the representation of all the databases linked to the PBM more semantic, thanks to the definition of ontologies.

This form of presentation of information helped us later in our work to define a tool that allows us to intelligently process this information, so we were able to make the client more active in participating in management by policy. As a result, another gain is clear in the decentralization of tasks [8].

This control is based on the definition of an adequate device, which represents the result of research on three levels: management by PBM policy, web service, and semantic definition:

- The first PBM level consists of the change of roles between the client and the PEP, the decentralization of tasks from the PDP central point to the PEP and the client, and intelligent interpretation at the client level using an appropriate tool.
- After this task definition, we answer the following question: how to deploy the services between the PBM entities, and what approach to use for the presentation of this service.
- We finalize by representing all the information (database) using a semantic definition that can be used for software.
- The main communication protocol COPS, and more precisely its extension COPS-PR dedicated to the transfer of policies, is our choice for the exchange of information. This information is necessarily all the rules and parameters of a policy, and this is what we are trying to control, for this we have presented the part of the PIB table (policy database) dedicated to storing this information in an OWL ontology; the latter will make it possible to generate the variables of the database in a form usable by software.

7   Conclusion

We define at the beginning, the different participants in our global architecture, and the actors responsible for the creation of the contract; we first distinguish the customer requesting the service with a containment mechanism and the supplier who proposes offers controlled by an access. Access that is under centralized management (management by policy), as well as the provider's manager expert who supports the translation of customer intentions, is not clear to lead to the service contract [9].

After establishing the two phases, we have a local PEP at the customer, with the rules recorded in its base, which correspond to the customer contract negotiated with the service provider, during the customer connection, the router or any other Provider network equipment will no longer need to contact the PDP policy server to recognize the rules to be executed for the client flow, and the local PEP determines these rules and offers decentralization of standard PBM tasks [10].

Standard management that uses methods that are obsolete these days, such as manually configuring network equipment through a console, has proven its limits, compared to another form of management that automates the entire circuit, from client request to equipment configuration, and management by policy (or PBM) has given another meaning to management, an important adaptability for several domains, among them the quality of service which constitutes the objective of our thesis. However, even with a so-called fairly positive track record, PBM with its centralized decision-making method in the policy server faces some challenges (centralized model of policy rules). This is where we believe our approach can change the image, and we assume that the customer can have the technical specifications that correspond to their service contract installed at their place, through an entity logic that will support the entire communication interface with the provider network. Task migration from network equipment to the client saves time and provides easy, mobile client access to that network [11].

References

1. Information technology—Abstract Syntax Notation One (ASN.1) (2009) Specification of basic notation. International Telecommunication Union
2. IETF RFC 3198 (2010) Terminology for policy-based management
3. Fine M, McCloghrie K, Seligson J, Chan K, Hahn S, Sahita R, Smith A, Reichmeyer F (2009) Framework policy information base, draft-ietf-framework-pib-07.txt
4. Pujolle G (2007) Management, control and evolution of IP networks. British Library Cataloguing-in-Publication Data
5. McGuinness DL, van Harmelen F (Knowledge Systems Laboratory, Stanford University), (Vrije Universiteit, Amsterdam) (2004) OWL web ontology language overview. W3C Recommendation
6. Dobson G, Lock R, Sommerville I (2005) QoSOnt: a QoS ontology for service-centric systems, Lancaster Univ UK. IEEE. ISBN: (0-7695-2431-1)

7. Maedche A The text-to-onto ontology extraction and maintenance environment. In: Proceedings of the ICDM Workshop on integrating data mining and knowledge management Jose, California, USA
8. Subranamian M (2000) Network management: principles and practice. Reading, Addison-Wesley, MA
9. Moraru V, Nguyen Manh T (2008) Les protocoles pour la gestion des réseaux Informatiques. Institut de la Francophonie pour l'Informatique, Hanoi. Juillet
10. Song Y-Q (2009) Garantir la Qualité de Service Temps Réel: Ordonnancement et Gestion de Files d'Attente. LORIA—Nancy Université
11. Atelin P (2011) Réseaux informatiques. Editions ENI, p 407

COVID-19 SEIAR Model with Sensitivity Analysis

Mohamed Derouich, E. N. Mohamed Lamlili, and Abdesslam Boutayeb

Abstract Beside the unexpected toll of mortality and morbidity caused by COVID-19 worldwide, low- and middle-income countries are more suffering from the devastating issues on economic and social life. This disease has fostered mathematical modelling. In this paper, a SEIAR mathematical model is presented to illustrate how policymakers may apply efficient strategies to end or at least to control the devastating wide spread of COVID-19.

Keywords COVID-19 · Mathematical model · Basic reproduction number · Equilibrium · Sensitivity analysis

1 Introduction

The coronavirus (COVID-19) which emerged in December 2019 has been widely spreading and affecting all countries around the world. According to the World Health Organization (WHO) statistics, the number of confirmed cases is around 400 million cases (by 10th February 2022) while the number of deaths caused by COVID-19 is approaching 6 million deaths. The disease has also caused devastating economic and social problems by imposing confinements, isolation, closing borders, hospitalisations and others. The International Monetary Fund (IMF) forecasts that the COVID-19 pandemic will cost the global economy $12.5 trillion through 2024. Despite the success in producing vaccines in a relatively short time, procurement of vaccine has been very inequitable between high-income countries which started vaccination by December 2020 and low- and middle-income countries where barely 10% of the population was vaccinated by December 2021 [2]. The coronavirus epidemics has fostered mathematical modelling during the

M. Derouich (✉)
National School of applied Sciences, Mohamed I University, Oujda, Morocco
e-mail: m.derouich@ump.ac.ma

E. N. Mohamed Lamlili · A. Boutayeb
Faculty of Sciences, Mohamed I University, Oujda, Morocco

last two years. Consequently, a large number of papers have been published on this theme as indicated by the 186 references cited in two recent reviews [4, 5]. Following our previous SIAR mathematical model [3], we present in this paper an SEIAR mathematical model, focusing on sensitivity analysis to illustrate the role of prevention in ending the epidemics or at least to control its devastating spreading.

2 Formulation of the Model and Sensitivity Analysis

2.1 Parameters of the Model

Let N denote the population size. In this model, death is proportional to the population size with rate constant μ and we assume a constant Λ due to births and immigrations. So $\dfrac{\mathrm{d}N}{\mathrm{d}t} = \Lambda - \mu N$. This population of size N is formed of Susceptible S, Exposed (but not infectious) E, symptomatic Infective I , Asymptomatic Infective A and Removed R.

$\dfrac{(\beta_1 I + \beta_2 A)S}{N}$ is the incidence, i.e. the rate at which susceptible individuals become Exposed . If the time unit is days, then the incidence is the number of new infection per day. The daily contact rate β_1 is the average number of adequate contacts of susceptible with symptomatic infected person per day, The daily contact rate β_2 is the average number of adequate contacts of susceptible with asymptomatic infected person per day, $\frac{I}{N}$ is the symptomatic infected person and $\frac{A}{N}$ is asymptomatic infected person fraction of the population. Time units of weeks, months or years could also be used. The life span $(1/\mu)$ is taken equal to 25,000 d (68.5 years). The other parameters used in the model are: $\frac{1}{\mu+\sigma_1}$ infection duration among symptomatic infected person, $\frac{1}{\mu+\sigma_2}$ infection duration among symptomatic infected person, α the death rate from COVID-19 and γ the rate at which exposed individuals become infectious.

2.2 Equations of the Model

A schematic representation of the model is shown in Fig. 1.

We consider SEIAR compartmental model that is to say that susceptible individuals become Exposed with a $\dfrac{(\beta_1 I + \beta_2 A)S}{N}$ rate, Exposed individuals become symptomatic infected person with a $p\gamma$ rate, or asymptomatic infected person with a $(1 - p)\gamma$ rate, then removed with immunity after recovery from infection.

The dynamics of this disease described by the following differential equations:
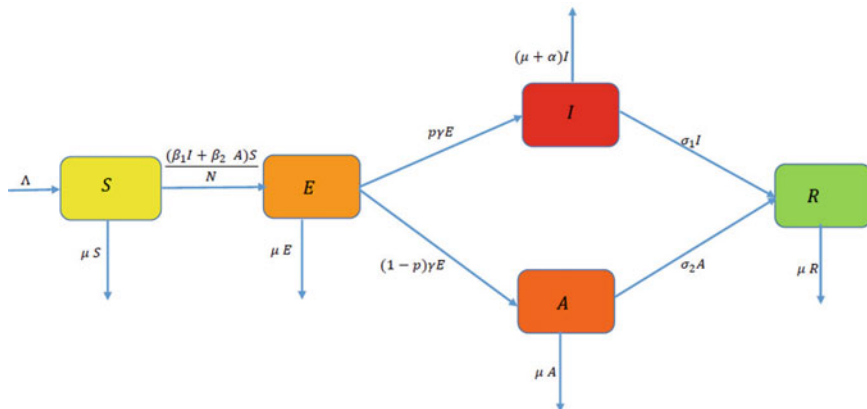
Fig. 1 Compartments population

$$
\begin{cases}
\frac{dS(t)}{dt} = \Lambda - \dfrac{S(t)\,(\beta_1 I(t) + \beta_2 A(t))}{N(t)} - \mu\, S(t) \\
\frac{dE(t)}{dt} = \dfrac{S(t)\,(\beta_1 I(t) + \beta_2 A(t))}{N(t)} - (\mu + \gamma)E(t) \\
\frac{dI(t)}{dt} = p\gamma\, E(t) - (\sigma_1 + \mu + \alpha)I(t) \\
\frac{dA(t)}{dt} = (1 - p)\gamma\, E(t) - (\sigma_2 + \mu)A(t) \\
\frac{dR(t)}{dt} = \sigma_1 I(t) + \sigma_2 A - \mu R(t) \\
\frac{dN(t)}{dt} = \Lambda - \mu\, N(t) - \alpha\, I(t)
\end{cases}
$$

With the condition $S(t) + E(t) + I(t) + A(t) + R(t) = N(t)$.
So: $R(t) = N(t) - S(t) - E(t) - I(t) - A(t)$
the previous system becomes:

$$
\begin{cases}
\frac{dS(t)}{dt} = \Lambda - \dfrac{S(t)\,(\beta_1 I(t) + \beta_2 A(t))}{N(t)} - \mu\, S(t) \\
\frac{dE(t)}{dt} = \dfrac{S(t)\,(\beta_1 I(t) + \beta_2 A(t))}{N(t)} - (\mu + \gamma)E(t) \\
\frac{dI(t)}{dt} = p\gamma\, E(t) - (\sigma_1 + \mu + \alpha)I(t) \\
\frac{dA(t)}{dt} = (1 - p)\gamma\, E(t) - (\sigma_2 + \mu)A(t) \\
\frac{dN(t)}{dt} = \Lambda - \mu\, N(t) - \alpha\, I(t)
\end{cases}
\tag{1}
$$

2.3 Positivity of Solutions

Theorem 1 *The set $\Omega = \left\{ (S, E, I, A, N) \in \mathbb{R}^5 / 0 \leq S, E, I, A, N \leq \frac{\Lambda}{\mu} \right\}$ is positively invariant under system (1).*

proof. From : $\dfrac{dN(t)}{dt} = \Lambda - \mu N(t) - \alpha I(t)$

$$\frac{dN(t)}{dt} = \Lambda - \mu N(t) - \alpha I(t)$$

$$\geq -(\mu + \alpha)N(t) \quad \text{(because } I(t) \leq N(t) \text{ and } \alpha \geq 0)$$

Let T be a fixed terminal time.

Then using Gronwall's inequality $N(t) \geq N(0)e^{-\int_0^T (\mu+\alpha)dt} \implies N(t) > 0$

On the other hand, we have: $\frac{dE(t)}{dt} = \dfrac{S(t)\,(\beta_1 I(t) + \beta_2 A(t))}{N(t)} - (\mu + \gamma)E(t)$. Assume by that there exists some time $t_* > 0$ such that $E(t_*) = 0$, other variables (S, N, I, A) are positive and $E(t) > 0$ for $t \in [0, t_*[$.

So we have

$$\frac{dE(t)e^{(\mu+\gamma)t}}{dt} = (\mu + \gamma)e^{(\mu+\gamma)t}E(t) + e^{(\mu+\gamma)t}\left[\frac{S(t)(\beta_1 I(t) + \beta_2 A(t))}{N(t)} - (\mu + \gamma)E(t)\right]$$

$$= e^{(\mu+\gamma)t}\left[\frac{S(t)(\beta_1 I(t) + \beta_2 A(t))}{N(t)}\right]$$

Integrating this Equation from 0 to t_* we have

$$\int_0^{t_*} \frac{dE(t)e^{(\mu+\gamma)t}}{dt}dt = \int_0^{t_*} \left[\frac{S(t)((\mu + \gamma)I(t) + \beta_2 A(t))}{N(t)}\right] e^{(\mu+\gamma)t}dt$$

then

$$E(t_*) = e^{-(\mu+\gamma)t_*}E(0) + e^{-(\mu+\gamma)t_*}\int_0^{t_*}\left[\frac{S(t)(\beta_1 I(t) + \beta_2 A(t))}{N(t)}\right]e^{(\mu+\gamma)t}dt > 0$$

which contradicts $E(t_*) = 0$. Consequently, $E(t) > 0 \;\; \forall t \in [0, T]$.
In the same way, we proof that: S, I, A are positive.

2.4   Equilibrium Points and R_0

For the model above, equilibrium points are defined such that there is no variations in S, E, I, A, N with respect to t:

$$\Lambda - \frac{S(t)\,(\beta_1 I(t) + \beta_2 A(t))}{N(t)} - \mu S(t) = 0 \tag{2}$$

$$\frac{S(t)\,(\beta_1 I(t) + \beta_2 A(t))}{N(t)} - (\mu + \gamma)E(t) = 0 \tag{3}$$

$$p\gamma E - (\sigma_1 + \mu + \alpha)I(t) = 0 \tag{4}$$

$$(1 - p)\,\gamma E - (\sigma_2 + \mu)A(t) = 0 \tag{5}$$

$$\Lambda - \mu N(t) - \alpha I(t) = 0 \tag{6}$$

Let $\delta_1 = \sigma_1 + \mu + \alpha$ and $\delta_2 = \sigma_2 + \mu$.
We have the following theorem:

Theorem 2 *The previous system admits two equilibrium points:*

1. *If $R < 1$, the system admits a trivial equilibrium $E_0 = (\frac{\Lambda}{\mu}, 0, 0, 0, \frac{\Lambda}{\mu})$*
2. *If $R > 1$, then there exists an endemic equilibrium $E_1 = (S^*, E^*, I^*, A^*, N^*)$.*

where $S^* = \frac{\Lambda}{\mu}\left(\frac{(\mu + \gamma)\delta_1 - \alpha p\gamma}{(\mu + \gamma)\delta_1 R - \alpha p\gamma}\right)$, $E^* = \frac{\Lambda\delta_1(R - 1)}{(\mu + \gamma)\delta_1 R - \alpha p\gamma}$,

$I^* = \frac{\Lambda p\gamma(R - 1)}{(\mu + \gamma)\delta_1 R - \alpha p\gamma}$, $A^* = \frac{\Lambda(1 - p)\gamma\delta_1(R - 1)}{(\mu + \gamma)\delta_1\delta_2 R - \alpha p\gamma\delta_2}$,

$N^* = \frac{\Lambda}{\mu}\frac{R[(\mu + \gamma)\delta_1 - \alpha p\gamma]}{(\mu + \gamma)\delta_1 R - \alpha p\gamma}$ *and* $R = \frac{p\beta_1\gamma\delta_1 + (1 - p)\beta_2\gamma\delta_2}{(\mu + \gamma)\delta_1\delta_2}$.

Reproduction number: The basic reproduction number R_0, which is defined as the average number of secondary infections produced when one infected individual is introduced into a host population where everyone is susceptible [1].

Theorem 3 *The basic reproduction number of the system (1) is given by*

$$R_0 = \frac{p\gamma\beta_1}{(\mu + \gamma)(\sigma_1 + \mu + \alpha)} + \frac{(1 - p)\gamma\beta_2}{(\mu + \gamma)(\sigma_2 + \mu)}$$

$$= R_{\mathrm{I}} + R_{\mathrm{A}}$$

R_{I} : the contribution of symptomatic infectious individuals
R_{A} : the contribution of Asymptomatic infectious individuals.

proof. Using notations in [6], the matrices F and V and their Jacobian matrices for the new infection terms and the remaining transfer term evaluated at the disease free equilibrium are respectively given by ;

$$F = \begin{pmatrix} \dfrac{S\,(\beta_1 I + \beta_2 A)}{N} \\ 0 \\ 0 \end{pmatrix} ; V = \begin{pmatrix} (\gamma + \mu)E \\ -p\gamma E + \delta_1 I \\ -(1 - p)\gamma E + \delta_2 A \end{pmatrix}$$

$$J_F = \begin{pmatrix} 0 & \dfrac{S^*\beta_1}{N^*} & \dfrac{S^*\beta_2}{N^*} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

For trivial equilibrium E_0, we have $S^* = N^*$ so $J_F = \begin{pmatrix} 0 & \beta_1 & \beta_2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

$J_V = \begin{pmatrix} \gamma + \mu & 0 & 0 \\ -p\gamma & \delta_1 & 0 \\ -(1-p)\gamma & 0 & \delta_2 \end{pmatrix} \Rightarrow J_V^{-1} = \begin{pmatrix} \frac{1}{\gamma+\mu} & 0 & 0 \\ \frac{p\gamma}{(\gamma+\mu)\delta_1} & \frac{1}{\delta_1} & 0 \\ \frac{(1-p)\gamma}{(\gamma+\mu)\delta_2} & 0 & \frac{1}{\delta_2} \end{pmatrix}$ Thus

$J_F J_V^{-1} = \begin{pmatrix} \frac{p\gamma\beta_1}{(\gamma+\mu)\delta_1} + \frac{(1-p)\gamma\beta_2}{(\gamma+\mu)\delta_2} & \frac{\beta_1}{\delta_1} & \frac{\beta_2}{\delta_2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

It follows that the basic reproduction, denoted by R_0, is given by

$$
\begin{aligned}
R_0 &= \rho(J_F J_V^{-1}) \\
&= \frac{p\beta_1}{(\gamma+\mu)\delta_1} + \frac{(1-p)\beta_2}{(\gamma+\mu)\delta_2} \\
&= \frac{p\beta_1}{(\gamma+\mu)(\sigma_1+\mu+\alpha)} + \frac{(1-p)\beta_2}{(\gamma+\mu)(\sigma_2+\mu)} \\
&= R_I + R_A
\end{aligned}
$$

where: $R_I = \dfrac{p\gamma\beta_1}{(\gamma+\mu)(\sigma_1+\mu+\alpha)}$ and $R_A = \dfrac{(1-p)\beta_2}{(\gamma+\mu)(\sigma_2+\mu)}$

Remark 1 $R = R_0$.

3 Sensitivity Analysis

Sensitivity analysis shows how a relative change in one of the parameters α, β_1, β_2, γ, σ_1, σ_2, μ or p induces a relative change in R_0. In this paper, we use the normal forward sensitivity index given by the ratio of the relative change in one of the parameters to the relative change in R_0.

For example, $\Psi_\alpha^{R_0} = \frac{dR_0}{d\alpha} \times \frac{\alpha}{R_0}$. A value of $\Psi_\alpha^{R_0}$ close to 1 would indicate that increasing (or decreasing) α by 10% will induce an increase (or decrease) of R_0 by nearly 10% while $\Psi_\alpha^{R_0} = -0.1$ will indicate that increasing (or decreasing) α by 10% will decrease (or increase) $\Psi_\alpha^{R_0}$ by only 1%. Sensitivity analysis is important for policymakers seeking an optimal strategy aiming to end (or at least to control) the epidemics by maintaining $R_0 < 1$. Computing the equilibrium points of the model allows to obtain the reproductive number as a function of the parameters used in the model: $R_0 = f(\alpha, \beta_1, \beta_2, \gamma, \sigma_1, \sigma_2, \mu, p)$. Then pragmatic strategies will seek to control R_0 by acting on the parameters which have the largest influence on R_0.

We have $R_0 = \frac{p\,\gamma\,\beta_1}{(\mu+\gamma)(\sigma_1+\mu+\alpha)} + \frac{(1-p)\,\gamma\,\beta_2}{(\mu+\gamma)(\sigma_2+\mu)}$ so

$$\Psi_{\beta_1}^{R_0} = \frac{\mathrm{d}R_0}{\mathrm{d}\beta_1} \times \frac{\beta_1}{R_0} = \frac{p\gamma}{(\mu+\gamma)(\sigma_1+\mu+\alpha)} \times \frac{\beta_1}{R_0}$$

$$\Psi_{\beta_2}^{R_0} = \frac{\mathrm{d}R_0}{\mathrm{d}\beta_2} \times \frac{\beta_2}{R_0} = \frac{(1-p)\gamma}{(\mu+\gamma)(\sigma_2+\mu)} \times \frac{\beta_2}{R_0}$$

$$\Psi_{\sigma_1}^{R_0} = \frac{\mathrm{d}R_0}{\mathrm{d}\sigma_1} \times \frac{\sigma_1}{R_0} = -\frac{p\beta_1\gamma}{(\mu+\gamma)(\sigma_1+\mu+\alpha)^2} \times \frac{\sigma_1}{R_0}$$

$$\Psi_{\sigma_2}^{R_0} = \frac{\mathrm{d}R_0}{\mathrm{d}\sigma_2} \times \frac{\sigma_2}{R_0} = -\frac{(1-p)\beta_2\gamma}{(\mu+\gamma)(\sigma_2+\mu)^2} \times \frac{\sigma_2}{R_0}$$

$$\Psi_{p}^{R_0} = \frac{\mathrm{d}R_0}{\mathrm{d}p} \times \frac{p}{R_0} = \left[\frac{\beta_1\gamma}{(\mu+\gamma)(\sigma_1+\mu+\alpha)} - \frac{\beta_2\gamma}{(\mu+\gamma)(\sigma_2+\mu)}\right] \times \frac{p}{R_0}$$

$$\Psi_{\mu}^{R_0} = \frac{\mathrm{d}R_0}{\mathrm{d}\mu} \times \frac{\mu}{R_0} = \frac{-\gamma}{(\mu+\gamma)^2}\left[\frac{p\beta_1\,(\sigma_1+2\mu+\alpha+\gamma)}{(\sigma_1+\mu+\alpha)^2} + \frac{(1-p)\beta_2\,(\sigma_2+2\mu+\gamma)}{(\sigma_2+\mu)^2}\right] \times \frac{\mu}{R_0}$$

$$\Psi_{\alpha}^{R_0} = \frac{\mathrm{d}R_0}{\mathrm{d}\alpha} \times \frac{\alpha}{R_0} = -\frac{p\gamma\beta_1}{(\mu+\gamma)(\sigma_1+\mu+\alpha)^2} \times \frac{\alpha}{R_0}$$

Theoretically, our mathematical model and sensitivity analysis indicate that acting on β_1 and σ_1 are the best ways to control R_0. Indeed, Table 1 shows that $\Psi_{\beta_1}^{R_0} = 0.82$ and hence, reducing β_1 by 10% should reduce R_0 by more than 8%. Similarly, $\Psi_{\sigma_1}^{R_0} = -0.70$, indicating that increasing σ_1 by 10% is expected to decrease R_0 by 7%. Pragmatically, however, policymakers will need to know how to apply this strategy? The first suggestion is to reduce the flow from the susceptible population to the exposed one and hence to the infected group. This can be done through prevention measures, including isolating infected individuals and confining those who are non infected, tracing every contact, physical distancing, self-protection, testing and obviously vaccinating susceptible individuals (see Fig. 2). The second important measure is more linked to the health system and its performance since it is based on screening, treating and hospitalizing infected individuals to avoid evolution towards complications and death.

The mathematical model presented in this paper will be extended to a larger model using more mathematical tools (stability analysis, optimal control and simulation) and discussing in details the effect of action on different parameters supposed to end (or at least to efficiently control) the epidemics (see Fig. 2).

Table 1 Sensitivity index of R_0

Parameter	β_1	β_2	$\sigma 1$	$\sigma 2$	γ	μ	α	p
Value	0.6	0.4	0.118	0.125	0.125	0.013	0.006	0.75
Sensitivity index	0.819	0.180	−0.704	−0.163	0.095	−0.191	−0.00047	0.277
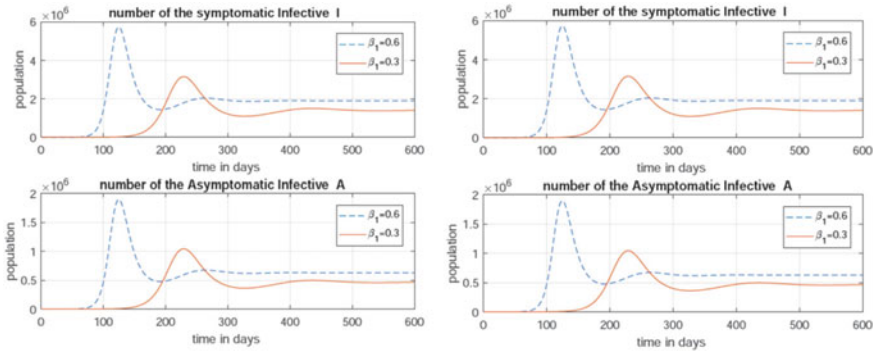
Fig. 2 Number of the symptomatic infective I and the asymptomatic infective A

4 Notes and Comments

Mathematical modelling and specially sensitivity analysis are an interesting tool that helps decision makers to adopt efficient and cost-effective strategies. In this direction, our simple mathematical model illustrates theoretically and pragmatically how policymakers should optimally act to end or at least to control the devastating spread of COVID-19.

References

1. Hethcote HW (2000) The mathematics of infectious diseases. SIAM Rev 42(4):599–653
2. Hyder AA, Hyder MA, Nasir K, Ndebele P (2021) Inequitable covid-19 vaccine distribution and its effects. Bull World Health Organ 99(6):406
3. Lamlili ME, Derouich M, Boutayeb A (2022) A SIAR model of covid-19 with control. Eng Lett 30(1)
4. Shankar S, Mohakuda SS, Kumar A, Nazneen P, Yadav AK, Chatterjee K, Chatterjee K (2021) Systematic review of predictive mathematical models of covid-19 epidemic. Med J Armed Forces India 77:S385–S392
5. Rahman A, Kuddus MA, Ip RH, Bewong M (2021) A review of covid-19 modelling strategies in three countries to develop a research framework for regional areas. Viruses 13(11):2185
6. Van den Driessche P, Watmough J (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math Biosci 180(1–2):29–48

Toward Multi-label Attribute Estimation on Multiple Faces Using CNN

Mohammed Berrahal and Mostafa Azizi

Abstract In the past few years, there has been a dramatic increase in the number of intelligent or cognitive applications. These applications can understand, learn, and interact with people in a more natural way than traditional applications. Many of these new applications are based on artificial intelligence (AI) technology, namely the ability of computers to learn and work on their own. One of the earliest examples of an intelligent application is the facial recognition system, based on deep learning models; this system can detect, estimate, and classify human faces based on facial attributes. Our approach is focusing on the estimation of attributes in images or videos. We train the first component, a YOLOv5 network, on Wider Face datasets, to detect faces, crop them, then pass the resulted data to the second component where we train a CNN multi-label network on CelebA datasets. The obtained model is capable of estimating two levels of face images, colored ones, and sketches. In addition, we make a comparison between YOLOv5 and SSD-ResNet (Single-Shot MultiBox Detector) face detectors.

Keywords Convolution neural networks · Facial attribute estimation · YOLOv5 · ResNet · Multi-label network · Single-shot MultiBox detector

1 Introduction

In recent years, there was an increasing demand for deep learning technologies [1, 2], which is mainly attributable to the explosive development of big data and artificial intelligence [3]. Against this backdrop, deep learning technology has become the focus of research in the field of artificial intelligence. Among the technologies that

M. Berrahal (✉) · M. Azizi
MATSI Research Lab, ESTO, Mohammed First University, Oujda, Morocco
e-mail: m.berrahal@ump.ac.ma

M. Azizi
e-mail: azizi.mos@ump.ac.ma

deep learning has revolutionized facial recognition technologies [4, 5], through developing neural networks that can classify, estimate, and manipulate facial attributes on any given face [6–8].

Facial attribute classification is an important task in computer vision [9], among its benefits include the ability to study the differences between faces and to identify any patterns that may exist. This can be used to improve facial recognition technology, help with the identification of criminals, and study the effects of aging on the face.

In our work, we try to implement existing deep learning models to extract human faces from giving images; to do so, we are going to train two networks; the first one is a YOLOv5 network to detect and crop faces of an input image. The second one is in CNN multi-label network that is going to estimate any existing attribute on all the faces given from the first network.

The rest of this work is structured, in the following order: Related works are shown in the second section. In the third section, we describe our implementation in further detail. Before concluding, the fourth section summarizes our findings.

2 Related Works

2.1 Facial Attributes Classification

The task of facial attribute classification is to identify and label facial features in an image. This is a common task in many fields, including computer vision, image processing, and machine learning. In many cases, the goal is to identify specific facial features, such as the position and orientation of the eyes, nose, and mouth [10]. In other cases, the goal may be to simply identify the presence or absence of certain facial features [11]. There are many different approaches to facial attribute classification. One common approach is to use a machine learning algorithm to learn a representation of the face, and then use this representation to identify facial features. Another approach is to use a set of predefined facial features, and then use these features to identify the location of the face in an image; others try to improve the old model by using the transfer learning approach [12]. There are many different facial recognition algorithms, each with its strengths and weaknesses. Some algorithms are better at identifying specific facial features, while others are better at identifying the overall shape of the face [13]. Some algorithms are better at dealing with variations in lighting, and others are better at dealing with variations in facial expression [14–16]. There are many applications for facial recognition algorithms. Some applications, such as security systems, require a high degree of accuracy [17], while others, such as social networking applications, can tolerate more errors. In addition, some applications require a large amount of training data, while others can be trained using a small amount of data.

2.2 YOLOv5 Object Detection

As an object detection method, YOLOv5 employs deep learning [18]. To recognize objects in a picture, it employs a convolutional neural network that uses only one type of neural network. The method has a high degree of accuracy and may be used in real time to identify things. In comparison to previous object detection algorithms, YOLOv5 has a number of advantages [19]. To begin, it is capable of accurately detecting a large variety of items [20]. It can also accomplish this in real time, making it ideal for use in self-driving automobiles, for example [21].

2.3 Single-Shot MultiBox Detector (SSD)-ResNet for Object Detection

A deep learning model called ResNet is a convolutional neural network, composed of multiple layers of neurons. The first layer is the input layer, and the last layer is the output layer. The neurons in the input layer receive input from the previous layer, and the neurons in the output layer send output to the next layer. The neurons in the hidden layers process the input and generate the output [22].

The power of this model resides in the use of multi-scale convolutional bounding box outputs attached to multiple feature maps at the top of the network. By this method, the network can efficiently model the space of possible box shapes [23].

3 Implementation of the Method

In this work, we are going to implement two deep learning models to estimate any existing facial attribute on giving images or videos; the challenge is to detect as many faces exist in the input file, then estimate attributes for every face. To do so, we train a YOLOv5 network on Wider datasets, to detect and crop faces, then feed them to CNN models, trained on CelebA with 40 attributes, as shown in Fig. 1.
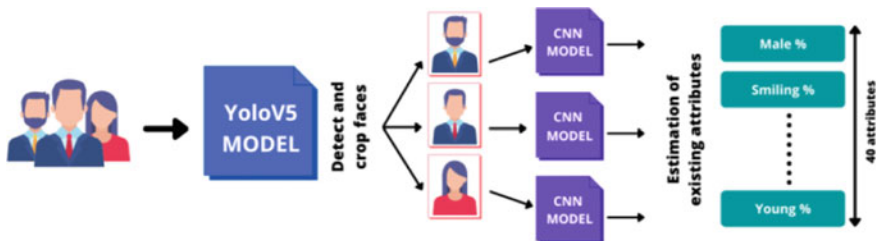
Fig. 1 Process of our work implementation

3.1 Hardware Characteristics

For the test of our models, we use high-performance computing (HPC) infrastructure cluster HPC-MARWAN:

- Compute Nodes: 2 * Intel Xeon Gold 6148 (2.4 GHz/20-core)/192 GB RAM
- GPU Node: 2 * NVIDIA P100/192 GB RAM
- Storage Node: 2 * Intel Xeon Silver 4114 (2.2 GHz/20-core)/18 * SATA 6 TB.

3.2 Datasets Used in This Work

In our work, we use two types of a dataset, the first is the Wider Face dataset that contains over 32,203 images, the image containing more than one face is labeled with a high degree of variability in scale, the total labeled face is 393,703, as the second dataset is CelebA face attributes dataset with more than 200 K celebrity images each with 40 attribute annotations.

3.3 Models' Architecture

The architecture used in our CNN model is organized into five main compound layers (ConvLayer1.0.4 and Dense1..*N*), as shown in Fig. 2. The first three ones (ConvLayer1.0.3) consist of each of the three following three layers: Conv2d layer, MaxPool2d layer, and ReLU layer, whereas the fourth one includes, in addition to these three aforementioned, the dropout layer. The last compound contains several dense layers. The input layer has 180 * 180 * 3 neurons and 40 output neurons to predict 40 attributes; the batch size is 64. Our model was trained over 20 epochs. We use ReLU as an activation function, while the optimizer function is "Adam".
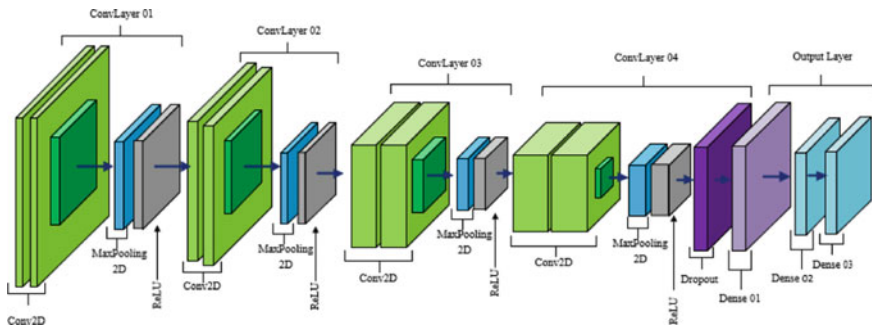
Fig. 2 CNN model layers

4 Result and Discussion

We start our experience by training two networks; we evaluate each network on two levels: the first by their efficiency using metrics of performance and second by judging the predicting results of the models.

The CNN multi-label model: Trained on CelebA datasets with over 200 k faces images, each is labeled on 40 binary attributes annotations [24]. We experimented with 60% of our dataset for training, 20% for testing, and the same rate for validation, a batch size of 64, and over 20 epochs. For all attributes, we have almost 104.5 million trainable parameters. Both training and validation accuracy reach 82%, while the loss is under 40%, the training process took over 49 min, and with 154.88 s on every iteration, the final model is capable to predict 40 predefined attributes.

The YOLOv5 model: Is trained on Wider Face datasets with over 32,203 images and 393,703 labeled faces [25]. We experimented with 80% of our dataset for training and 20% for testing, running under 150 epochs. As shown in Fig. 3, the training process gives a good result, concerning the training and validation losses descend above 1.2%. The other metrics like the mean average precision reaching 93%, the recall reaching 90%, the precision at 88%, and the inference speed is 61 FPS.

We wanted to push further our experiment, by comparing our YOLOv5 face detection model, with the SSD-ResNet pretrained model on real data, SSD-ResNet reach a Mean Average precision reaches of 97% and inference speed is 14 FPS, the Fig. 4 shows the tested models on different images. With most images, the SSD-ResNet model outruns our YOLOv5 model in terms of detection rate with a high percentage like in the image (2), but as we see in the image (1), the SSD-ResNet fails to detect the men on the left of the image, as ours succeeds to detect it. Also, the experiment
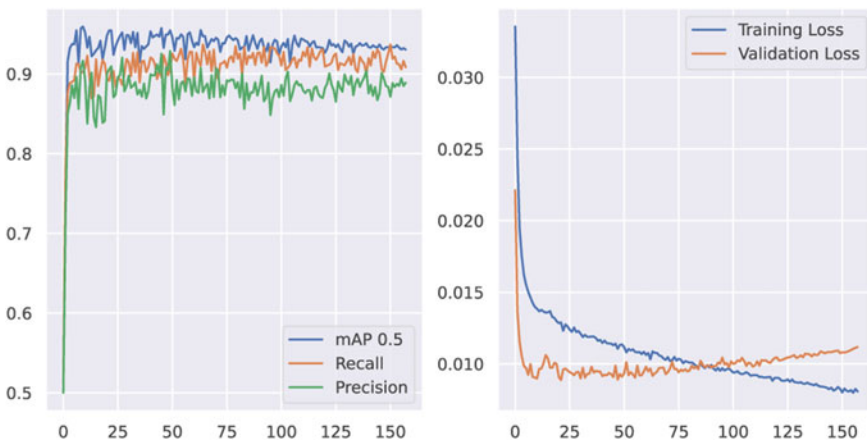
Fig. 3 Results of YOLOv5 face detection model training and validation losses and three other metrics: average precision, recall, and precision, in 150 epochs

Image from Wider dataset :

Image from external source :
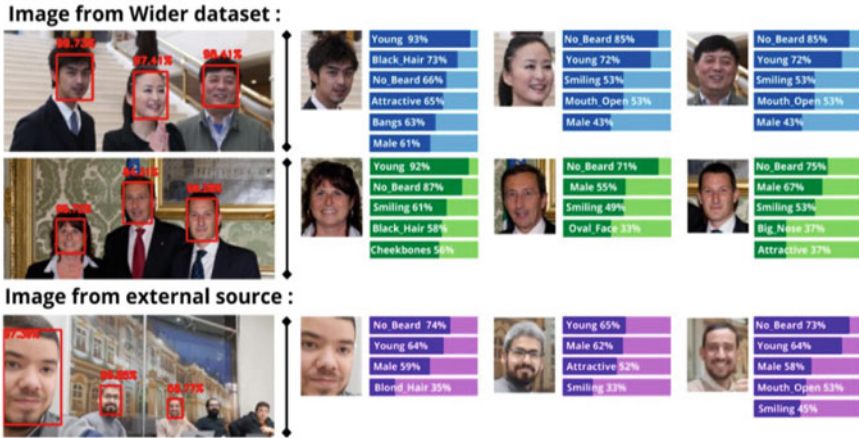
Fig. 4 Some examples of the result gotten by our implementation (SSD-ResNet, CNN multi-label) from wider datasets and external sources

shows that YOLOv5 is capable of detecting faces in videos and live and reaching 61 FPS of inference speed, while the SSD-ResNet can be fed only by one image with very low inference speed, so to pass an entire video, we have to divide it into frames and loop all, which takes more time and resources compared to YOLOv5 models. So, we can deduce that the YOLOv5 model (resp. SSD-ResNet) is more appropriate for video face detection (resp. for a single image).

Now we have two possible implementations, the first one uses YOLOv5 on the video to detect and crop every face and send them to the CNN to predict attributes, and the second uses SSD-ResNet on a single image and repeat the process. We use the second method to test some images, as shown in Fig. 5. The result succeeds not only to detect faces and estimate attributes on Wider images but also can, on external ones, of course, there will always be exceptions like the failure of detecting more than three faces in the external image, but sometimes the quality of input images has a huge impact on the detection rate.

5 Conclusion

In this paper, we combine two network models, to detect all faces existing in an image or a video, then estimate all possible attributes in each face. We train a YOLOv5 network on the Wider dataset; reaching over 93% on average precision, 90% in recall, and 88% in precision. Afterward, we compare it to an existing pretrained model SSD-ResNet by real data. We found that our YOLOv5 model is more accurate with videos; on the other hand, SSD-ResNet is more accurate with images. We also train a CNN multi-label network on CelebA datasets; we reach an accuracy of 82%, while the loss
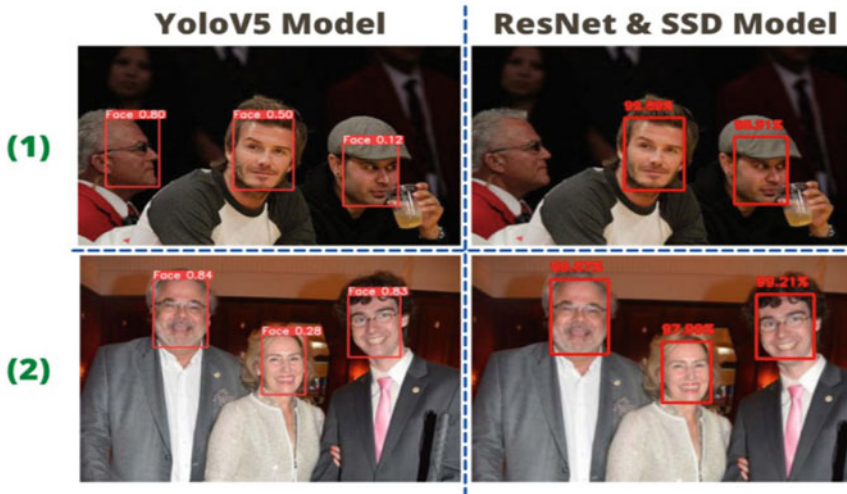
Fig. 5 Comparison between YOLOv5 (our pretrained model) and SSD-ResNet on face detection using real images

is under 40%. Finally, we use YOLOv5 to detect faces from videos and live videos, then feed the crop faces to CNN multi-label to predict and estimate attributes; the same work is done on the SSD-ResNet pretrained model and CNN multi-label but single images.

References

1. Idrissi I, Azizi M, Moussaoui O (2022) A stratified IoT deep learning based intrusion detection system. In: 2022 2nd international conference on innovative research in applied science, engineering and technology (IRASET)
2. Idrissi I, Azizi M, Moussaoui O (2021) Accelerating the update of a DL-based IDS for IoT using deep transfer learning. Indones J Electr Eng Comput Sci 23(2):1059–1067. https://doi.org/10.11591/IJEECS.V23.I2.PP1059-1067
3. Bouali H, Hammoudi Y, Idrissi I, Boukabous M, Zerguit Y (2022) Review on maintenance of photovoltaic systems based on deep learning and internet of things. Indones J Electr Eng Comput Sci Int J Comput Digit Syst 26(2):1060. https://doi.org/10.11591/ijeecs.v26.i2.pp1060-1072.
4. Guo G, Zhang N (2019) A survey on deep learning based face recognition. Comput Vis Image Underst 189:102805. https://doi.org/10.1016/j.cviu.2019.102805
5. Kortli Y, Jridi M, Al Falou A, Atri M (2020) Face recognition systems: a survey. Sensors 20(2):342. https://doi.org/10.3390/S20020342

6. Berrahal M, Azizi M (2020) Review of DL-based generation techniques of augmented images using portraits specification. In: 4th international conference on intelligent computing in data sciences, ICDS 2020. https://doi.org/10.1109/ICDS50568.2020.9268710

7. Berrahal M, Azizi M (2021) Augmented binary multi-labeled CNN for practical facial attribute classification. Indones J Electr Eng Comput Sci 23(2):973–979. https://doi.org/10.11591/IJEECS.V23.I2.PP973-979

8. Berrahal M, Azizi M (2022) Optimal text-to-image synthesis model for generating portrait images using generative adversarial network techniques. Indones J Electr Eng Comput Sci 25(2):972–979. https://doi.org/10.11591/IJEECS.V25.I2.PP972-979

9. Zheng X, Guo Y, Huang H, Li Y, He R (2020) A survey of deep facial attribute analysis. Int J Comput Vis 128(8–9):2002–2034. https://doi.org/10.1007/s11263-020-01308-z

10. Zhang N, Paluri M, Ranzato M, Darrell T, Bourdev L (2013) PANDA: pose aligned networks for deep attribute modeling. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, Nov 2013, pp 1637–1644

11. Mao L, Yan Y, Xue J-H, Wang H (2020) Deep multi-task multi-label CNN for effective facial attribute classification. arXiv, Feb 2020

12. Berrahal M, Azizi M (2022) Improvement of facial attributes' estimation using transfer learning, pp 1–7. https://doi.org/10.1109/iraset52964.2022.9737845

13. Marinescu AI, Ileni TA, Darabant AS (2019) A versatile 3D face reconstruction from multiple images for face shape classification. In: 2019 27th international conference on software, telecommunications and computer networks, SoftCOM 2019, Sept 2019. https://doi.org/10.23919/SOFTCOM.2019.8903745

14. Boukabous M, Azizi M (2021) A comparative study of deep learning based language representation learning models. Indones J Electr Eng Comput Sci 22(2):1032–1040. https://doi.org/10.11591/IJEECS.V22.I2.PP1032-1040

15. Boukabous M, Azizi M (2022) Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers. Indones J Electr Eng Comput Sci 25(2):1131–1139. https://doi.org/10.11591/IJEECS.V25.I2.PP1131-1139

16. Boukabous M, Azizi M (2021) Review of learning-based techniques of sentiment analysis for security purposes. In: Innovations in smart cities applications, vol 4, pp 1–14.https://doi.org/10.1007/978-3-030-66840-2_8

17. Boukabous M, Azizi M (2022) Multimodal sentiment analysis using audio and text for crime detection, pp 1–5. https://doi.org/10.1109/IRASET52964.2022.9738175

18. Zhou T, Zheng L, Peng Y, Jiang R (2022) A survey of research on crowd abnormal behavior detection algorithm based on YOLO network. In: 2022 2nd international conference on consumer electronics and computer engineering, Jan 2022, pp 783–786. https://doi.org/10.1109/ICCECE54139.2022.9712684

19. Qi D, Tan W, Yao Q, Liu J (2021) YOLO5Face: why reinventing a face detector, May 2021

20. Kherraki A, El Ouazzani R (2022) Deep convolutional neural networks architecture for an efficient emergency vehicle classification in real-time traffic monitoring. IAES Int J Artif Intell 11(1):110–120. https://doi.org/10.11591/IJAI.V11.I1.PP110-120

21. Kherraki A, Maqbool M, El Ouazzani R (2021) Traffic scene semantic segmentation by using several deep convolutional neural networks. In: 2021 3rd IEEE Middle East North Africa communications conference (MENACOMM 2021), pp 1–6. https://doi.org/10.1109/MENACOMM50742.2021.9678270

22. Lu X, Kang X, Nishide S, Ren F (2019) Object detection based on SSD-ResNet. In: Proceedings of the 2019 6th IEEE international conference on cloud computing and intelligence systems (CCIS), Dec 2019, pp 89–92. https://doi.org/10.1109/CCIS48116.2019.9073753

23. Liu W et al (2015) SSD: single shot multibox detector. In: Lecture notes in computer science (including subseries Lecture notes artificial intelligence and lecture notes in bioinformatics. LNCS, vol 9905, Dec 2015, pp 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

24. Large-scale CelebFaces attributes (CelebA) dataset. [Online]. Available: http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html. Accessed: 07 May 2020
25. WIDER FACE: a face detection benchmark. [Online]. Available: http://shuoyang1213.me/WIDERFACE/. Accessed: 28 Jan 2022

New Design of an X-Band 2 × 2 Patch Array Antenna with Circular Slots for Nanosatellites

Nabil EL Hassainate, Ahmed Oulad Said, and Zouhair Guennoun

Abstract In order to meet the requirements in terms of size, weight, high gain, wide bandwidth and circular polarization of antennas used for nanosatellites, a 2 × 2 circular polarization X-band patch array antenna with circular slots is designed for a 1U CubeSat. This article describes the antenna and presents its simulated performance. The antenna is formed by two layers of substrate, a planar array 2 × 2 of patch truncated at the corner and having a cut circular slot located on the top, a ground plane encapsulated between the two layers and a feed network on the bottom. The overall performance of the antenna is improved by the introduction of cut circular slots on four antenna elements. The presented antenna is excited by a suitable feed network. It offers a bandwidth of 740 MHz in the frequency range of 8.60–9.34 GHz with a maximum gain of 13.30 dB. It has a return loss of −27.70 dB and an axial ratio of 1.693 dB.

Keywords CubeSat · Antenna design · Antenna arrays · X-Band · Circular polarization · Simulation

N. EL Hassainate (✉) · Z. Guennoun
Smart Communications Team (ERSC), Mohammadia School of Engineers, Mohammed V University in Rabat, Rabat, Morocco
e-mail: elhassainate123@gmail.com

Z. Guennoun
e-mail: zouhair@emi.ac.ma

N. EL Hassainate · A. O. Said
Electronics and Telecommunications Department, Royal Air Academy, Marrakech, Morocco

N. EL Hassainate · Z. Guennoun
University Center for Research in Space Technologies (CURTS), Mohammed V University in Rabat, Rabat, Morocco

1   Introduction

Nanosatellites have revolutionized the space industry following the success of the two missions Raincube "Radar In a CubeSat" [1] and Marco "Mars Cube One" [2]. CubeSats are extremely small and light. A 1U form factor (FF) cubeSat has an overall mass less than 1.33 kg and a fixed size of 10 cm × 10 cm × 10 cm. Other FF cubeSats exist 2U, 3U, 6U and 12U. CubeSats are used in many applications such as telecommunications, earth observation and imaging.

Like other satellites, CubeSats must provide a communication link with a ground station based on earth. This communication link should be reliable by allowing the transmission of a sufficient speed. One of the important components that may affect the performance of the nanosatellite communications system is the antenna. It should be light, small in size with high gain and sufficient bandwidth. Several research works on patch array antennas design seeking an optimal operation for nanosatellite applications have been published [3–5]. For optimal operation, these antennas should have high gain, wide bandwidth and circular polarization.

This article firstly describes the design steps of an X-band 2×2 patch array antenna with circular slots and a suitable feed network, then analyzes the results obtained before concluding.

2   X-Band 2×2 Patch Array Antenna Design and Structure

To meet the requirements imposed by the CubsSat standard [6], the choice of a 2×2 patch array antenna with circular polarization is justified by the surface area available at the 1U CubeSat. The procedure adopted for the design of an X-band 2×2 patch array antenna is as follows:

- First, start the design and optimization of single patch antenna to achieve circular polarization.
- Second, carry on the design and optimization of an X-band 2×2 patch array antenna which offers good directivity and high gain.
- Last, conclude with the design of the feeding network of an X-band 2×2 patch array antenna which allows a better adaptation.

2.1   Single Patch Antenna Design

The single patch antenna design starts first by the choice of substrate and its thickness. The conducted study of several publications on patch antennas proposed and used for nanosatellites [7–9] led to the choice of the ROGGER RT5880 substrate having a relative permittivity $\epsilon_r = 2.2$. with a thickness of 1.575 mm which offers better performance in terms of gain. Fig. 1 shows the proposed patch antenna element.
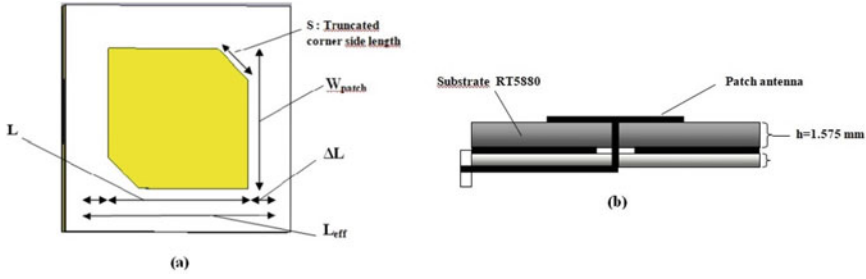
Fig. 1 **a** Top view of proposed patch antenna element, **b** proposed antenna element side view

The computation of the initial parameters (ω_{patch}, ϵ_{reff}, L_{patch}) comes afterwards [10]. In the case of a square pattern, the width of the line is half wave:

$$\omega_{\text{patch}} = \frac{c}{2 \times f \times \sqrt{\epsilon_{\text{reff}}}} \tag{1}$$

Where c is the speed of light (3×10^8 m/s), f is the resonance frequency of the 9 GHz antenna and ϵ_{reff} is the effective dielectric constant given by the expression given in Eq. 2. The computation of the effective dielectric constant of a microstrip line is based on the width of the line and the height of the substrate:

$$\epsilon_{\text{reff}} = \frac{\epsilon_r + 1}{2} + (\frac{\epsilon_r - 1}{2}) \times \frac{1}{\sqrt{1 + 12 \times \frac{h}{\omega_{\text{patch}}}}} \tag{2}$$

where ϵ_r is the relative permittivity of the substrate, h is the height of the substrate (1.575 mm) and ω_{patch} is the width of the patch. Equations 1 and 2 are related, meaning that ω_{patch} and ϵ_{reff} are correlated to each other. Therefore, it becomes necessary to perform recurrences until the result converges. The first iteration is completed from equation (Eq. 1) in which $\omega_{\text{patch}} = 0$, and $\epsilon_{\text{reff}} = \frac{\epsilon_r + 1}{2}$. The iterations are stopped when the difference between two consecutive lengths is less than the resolution that can be obtained. By the final iteration, we obtain: $\omega_{\text{patch}} = 11.87$ mm, and $\epsilon_{\text{reff}} = 1.972$. From where :

$$L_{\text{eff}} = \frac{c}{2 \times f \times \sqrt{\epsilon_{\text{reff}}}} \tag{3}$$

$$\delta L = 0.412 \times h \times \frac{(\epsilon_{\text{reff}} + 0.3) \times (\frac{\omega_{\text{patch}}}{h} + 0.254)}{(\epsilon_{\text{reff}} - 0.258) \times (\frac{\omega_{\text{patch}}}{h} + 0.8)} \tag{4}$$

where L_{eff} is the effective length of the patch antenna and δL is the extension of the length of the patch antenna. The antenna patch length is given by the following equation: $L_{\text{patch}} = L_{\text{eff}} - 2 \times \delta L$.
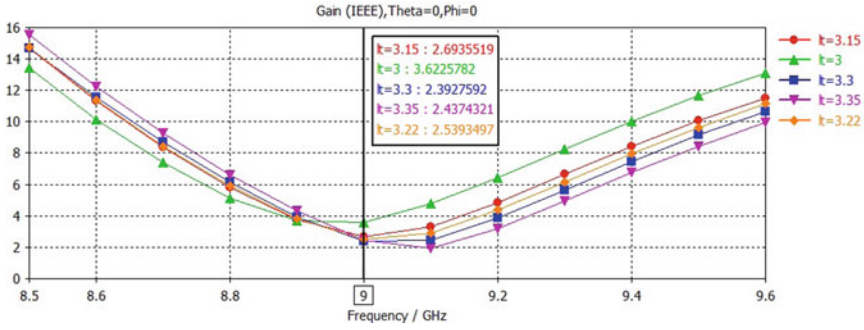
Fig. 2 Antenna element axial ratio for different values of *S*

Finally, to ensure the circular polarization of antenna element, the printed pattern is designed as a square [11] and the corners of the antenna are cut off. The initial S value of the truncated corners is 3 mm. The excitation point is chosen by setting the resonant frequency at 9 GHz and a return loss S_{11} less than -10 dB. To achieve this objective, several positions of the excitation point from coordinate point ($X_f = 0$ mm and $Y_f = -5.14$ mm) [12] were simulated. The excitation point meeting this goal is located at $X_f = 1.06$ mm and $Y_f = 2.91$ mm. The antenna element was configured to achieve an axial ratio less than 3 dB. The simulation results are presented in Fig. 2.

To improve the circular polarization of the patch, the introduction of a circular slot in the patch center has an even symmetry with respect to the axis of resonance of the antenna (the *X* axis) and improve the axial ratio of antenna element. A parametric study on the value of the radius of the circular slot R_s, which offers the best results, leads to $R_s = 0.7$ mm. The proposed antenna element with cut circular slot resonates at the frequency of 9 GHz for $S = 3.22$ which corresponds to an axial ratio of 2.43 dB and a gain of 5.539 dB with a return loss of 40.39 dB.

2.2 X-Band 2×2 Patch Array Antenna Design and Optimization

The choice of the 2×2 array architecture of the antenna aims to increase the gain of the antenna. The antenna elements are arranged in a 2×2 planar array with an initial spacing of $\lambda_g/2$ (where $\lambda_g = \lambda/\sqrt{\epsilon_{\text{reff}}}$) . Then, to avoid coupling between the antennas and eliminate side lobes and to ensure the directivity of the antenna, the spacing between the antenna elements must be well adjusted to the area available at CubeSat (10×10 cm^2). After optimization, the spacing between antenna elements, which meets our objective, is $0.541\lambda_g$ or (dx =dy =12.86 mm).

Figure 3 shows the two proposed antennas: the first is formed by two antenna elements with two cut circular slots and two without cut circular slots and the second

Fig. 3 a X-Band 2×2 patch array antenna with two cut circular slots, **b** X-Band 2×2 patch array antenna with four cut circular slots

(a)   (b)

2×R$_s$= 1.6 mm
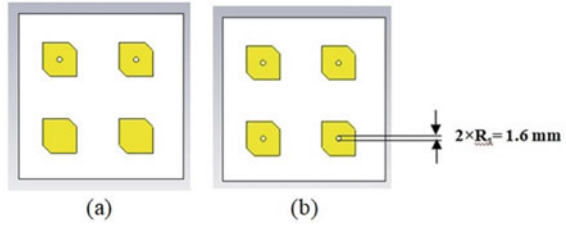
Table 1 Comparison of the performance of the two proposed antennas

Proposed array antenna	Axial ratio (dB)	S11 (dB)	Gain (dB)	Directivity (dB)
With two cut circular slots	1.658	− 27.44	13.39	13.86
With four cut circular slots	1.693	− 27.70	13.30	13.76

Fig. 4 Feed network of the X-Band 2×2 patch array antenna

50 Ω
70.71 Ω
70.71 Ω
50 Ω
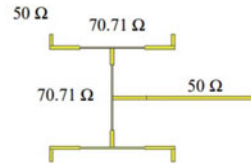
is formed by four antenna elements which all have cut circular slots. The simulation results of the two proposed antennas are presented in Table 1.

2.3 Feed Network Structure

Feed network is designed to provide a uniformly distributed power to all four antenna elements as shown in Fig. 4. A quarter-wave transformer is also used wherever an impedance matching is needed. The feed network consists of ribbon lines and T-junction power dividers with quarter-wave transformers. In order to have an impedance of 50 Ω in the input and outputs of each power divider, we use the method of Wilkinson which gives the value of the impedance Z by the following equation : $Z = Z_0 \times \sqrt{2}$ Where 2 is the number of branching points and Z0 (50 Ω) is the input impedance.

Furthermore, to minimize the effects of radiation and coupling of the microstrip lines, the feed network is fabricated on an RO3010 substrate with a dielectric constant of $\epsilon_r = 10.2$ [4]. The antenna element is powered by a probe connected to the micro ribbon line of the proposed antenna's feed network. The following Fig. 5 shows a side view of the proposed antenna.
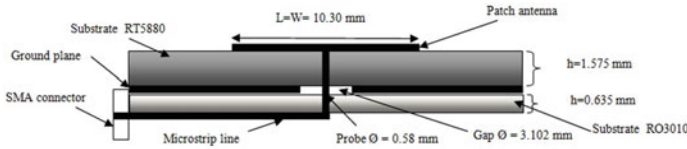
Fig. 5 Side view of the proposed antenna

Table 2 Dimensions of the proposed antenna

Symbol	Parameter	Value (mm)
L	Untrimmed patch Length	10.28
ω	Untrimmed patch Width	10.28
S	Truncated corner side length	3.39
R_s	Radius of circular slot	0.8
dx, dy	Spacing between elements	12.86

3 Simulation Results

The antenna is designed and simulated using CST Microwave Studio electromagnetic analysis and simulation software. The dimensions of the proposed antenna are given in Table 2. Figure 6 shows the simulated S11 parameter of the proposed antenna. It has a bandwidth of 740 MHz in the frequency range 8.61–9.34 GHz.

The proposed antenna offers a circular polarization gain of 13.30 dB at the frequency of 9 GHz. Figure 7 shows the radiation pattern of the proposed antenna.

The axial ratio of the proposed antenna at the resonant frequency of 9 GHz indicates an axial ratio of 1.693 dB as shown in Fig. 8.

Table 3 compares the performances of the proposed antenna with those reported in [3, 9, 14]. The proposed antenna gives a better gain at the 9 GHz frequency compared to those reported in [3, 9, 14]. It has a good impedance matching with a medium bandwidth compared to other referenced antennas.
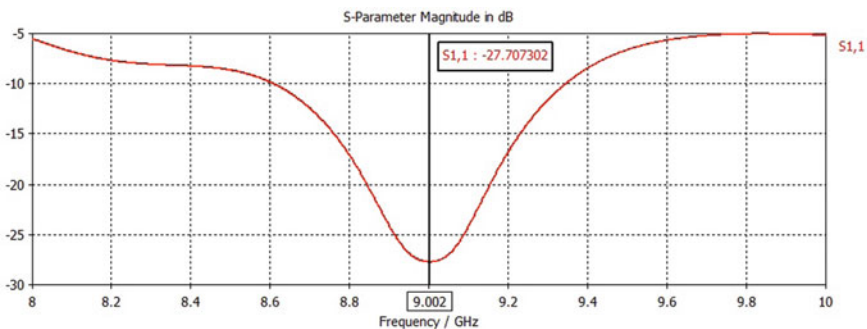
Fig. 6 Reflection coefficient S11 of the proposed antenna
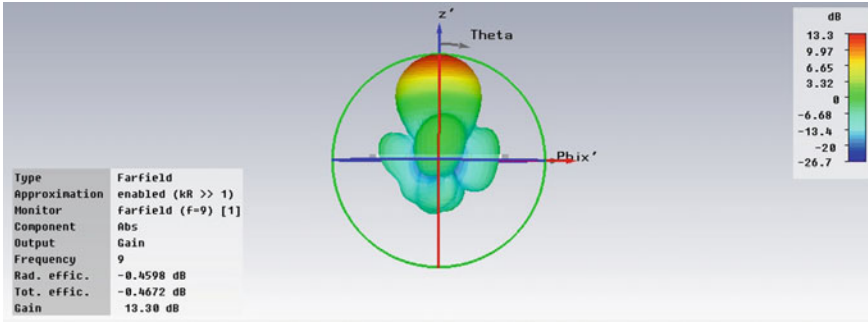
Fig. 7 Gain radiation pattern of the proposed antenna

Fig. 8 Axial Ration of the proposed antenna

Table 3 A comparison between the proposed antenna and other 2 × 2 array antennas

2 * 2 Arrays	[3]	[9]	[14]	Proposed
Gain (dB)	12.1	7.6	11.5	13.3
Frequency (GHz)	7.75–8.75	8.2	8	9
10-dB impedance bandwidth (GHz)	6.75–9.15	8–8.4	7.63–8.57	8.60–9.34
S11(dB)	≤ −10	≤ −15	≤ −10	−27.70

4 Conclusion

In this paper, the simulation results of the designed X-band 2 × 2 patch array antenna operating at 9 GHz are presented showing higher performance than the referenced ones. The designed antenna meets the requirements imposed by nanosatellites in terms of weight, volume, power and circular polarization. To validate these simulation results, an antenna prototype would be fabricated and measurement performed on the prototype.

The proposed antenna is 50×50 mm with a total height of 2.21 mm with four circular slots cut at the center of the antennas element. This antenna design has reduced the weight of the antenna and improved its performance as well as its use in nanosatellites. The antenna presented is excited by a suitable feed network. It offers a bandwidth of 740 MHz in the frequency range of 8.60–9.34 GHz with a maximum gain of 13.30 dB. It has a return loss of -27.70 dB and an axial ration of 1.693 dB.

Our future work concerns the manufacturing and testing a prototype 4×4 patch array antenna providing 20 dB higher gain for possible use in the university 3U CubeSat project.

References

1. Peral E, Tanelli S, Haddad Z, Sy O, Stephens G, Im E (2015) 'Raincube: a proposed constellation of precipitation profiling radars in CubeSat. In: 2015 IEEE international geoscience and remote sensing symposium (IGARSS). IEEE, pp 1261–1264
2. Chahat NE (2018) A mighty antenna from a tiny CubeSat Grows. IEEE Spectr 55(2):32–37 February
3. Serup DE, Williams RJ, Zhang S, Pedersen GF (2020) Shared aperture dual S- and X-Band antenna for nano-satellite applications. In: 2020 14th European conference on antennas and propagation (EuCAP), pp 1–4
4. Hashim ISM, Al-Hourani A, Wayne Rowe ST, Scott JR (2019) Adaptive X-Band satellite antenna for Internet-of-Things (IoT) over satellite applications. In: 2019 13th International conference on signal processing and communication systems (ICSPCS), pp 1–7
5. Sushma, S, Rao VS (2019) Transceiver antenna for telemetry and telecommand applications. In: 2019 International conference on communication and signal processing (ICCSP), pp 0511–13
6. Tubbal FE, Raad R, Chin KW (2015) A survey and study of planar antennas for pico-satellites. IEEE Access 3:2590–2612
7. Alam, Touhidul, Mandeep Singh Jit Singh, Mengu Cho, and Mohammad Tariqul Islam. 'Performance Analysis of an X-Band Circular Polarized Antenna for Nanosatellite Communication'. In Space Science and Communication for Sustainability. Springer, Singapore, pp 275–283
8. Bansode P (2019) 2X8 Multilayer microstrip patch array antenna for C-Band radar micro-SAR. ArXiv:1903.00224 [Physics], 1 Mar 2019
9. Khac KN, Phong ND, Manh LH, Le Trong TA, Le Huu H, Hien BT, Ngoc DN (2018) A design of circularly polarized array antenna for X-Band CubeSat satellite communication. In: 2018 International conference on advanced technologies for communications (ATC), pp 53–56
10. Stutzman WL, Thiele GA (2012) Antenna theory and design. Wiley
11. Balanis CA (2005) Antenna theory: analysis and design, 3rd edn. Wiley, Hoboken, NJ
12. Gao SS, Luo Q, Zhu F (2014) Circularly polarized antennas University of Kent, UK 1 ed. Wiley
13. Kurniawan F, Sumantyo JTS, Ito K, Kuze H, Gao S (2017) Patch antenna using rectangular centre slot and circular ground slot for circularly polarized synthetic aperture radar (CP-SAR) application. Prog Electromagn Res 160:51–61
14. Mathur P, Kumar G (2017) Dual frequency microstrip antenna at S- and X-Bands with HigherOrder mode suppression technique. IET Microwaves, Antennas Propagation, 12–13 Nov 2017

Materials and Devices Applications

Improvement of Silicon Nanowire-Based Photovoltaic Solar Cell with the Integration of CIGS Quantum Wells

Meriem Safi, Abdelkader Aissat, Houcine Guesmi, and Jean Pierre Vilcot

Abstract The present work focuses on the modeling and simulation of a new photovoltaic solar cell based on a CIGS/Si nanowire. The study of the effect of inserting quantum wells in a nanowire structure for different characteristic parameters has been done to evaluate the potential of the confinement phenomenon. The best results of the solar cell are obtained with a number of 14 quantum wells integrated in a Si nanowire-based solar cell with an efficiency of 3.16%, FF of 73.50%, V_{oc} equal to 0.41 V and J_{sc} equal to 10.50 mA/cm^2.

Keywords Nanowires · Quantum wells · Solar cell · Photovoltaics

1 Introduction

During the last decade, a new technology has seen the world, a technology characterized by miniaturization, that is to say, the manufacture of systems at the nanoscopic scale, and the latter tends to show its skills and remarkable properties that meet the requirements of mobility and ease of use [1, 2]. This new revelation has affected several disciplines, among which we find renewable energies. Since the industrial revolution, global energy consumption has been steadily increasing, a soaring consumption has been noticed on several energy resources that is 81% of fossil fuels, 31% come from oil, 29% come from coal, and 21% come from natural gas according to a census made in 2015 [3]. For this reason, many researchers are working hard to find energy resources that allow a long life span. Solar photovoltaic energy is the right solution, because of its abundance that allows us to meet the global need [4]. The

M. Safi · A. Aissat (✉) · H. Guesmi
Laboratory LATSI, Faculty Technology, University Saad Dahlab Blida 1, Blida, Algeria
e-mail: sakre23@yahoo.fr; aa_aissat@univ-blida.dz

A. Aissat · J. P. Vilcot
Institute of Microelectronics, Electronics and Nanotechnology (IEMN), UMR CNRS 8520, University of Sciences and Technologies of Lille 1, Poincare Avenue, B.P. 60069, 59652 Villeneuve-d'Ascq, France
e-mail: jean-pierre.vilcot@univ-lille.fr

photovoltaic technology is rapidly and effectively developed in the context of micro-electronics; however, a large amount of material is needed which leads to excessively high prices to benefit from this kind of system. For this reason, several studies have been carried out in order to develop photovoltaic solar cells based on nanostructures with a high efficiency compromise modeling and simulation of a new solar photo-voltaic cell based on CIGS/Si nanowire. The study of the effect of inserting quantum wells into a nanowire structure for different characteristic parameters was done to evaluate the potential of the confinement phenomenon. The best results of the cell are obtained with a number of 14 quantum wells integrated in Si nanowire-based solar cell with an efficiency of 3.16%, FF of 73.46%, V_{oc} equal to 0.41 V and a J_{sc} equal to 10.50 (mA/cm^2). The nanowires are the revolution of these studies; indeed, they allow the trapping of the light, and also, the geometry of the latter allows an effective collection of charge carriers photogenerated with optimal dimensions for a low cost.

The present study focuses on the modeling of a silicon-based nanowire with integrated CIGS wells; this semiconductor is a very promising mixture, which also presents one of the best absorption coefficients. Also, the best efficiency achieved by a rigid CIGS cell is 23.4%, and no other low-cost thin structure has done better. The present study focuses on the modeling, simulation and characterization of a photovoltaic cell, for which we studied the influence of quantum well insertion in a nanowire in different characteristic parameters of the cell. Figure 1 shows the structure of the Si nanowire-based solar cell with integration of four layers of CIGS/Si QWs grown on a Si substrate. The CIGS QWs in red are inserted in the intrinsic region, and they are characterized by a thickness of 3 nm and a Si barrier of 8 nm, both with a radius of 190 nm.

2 Theoretical Models

The carrier-dependent concentration mobility described by the Masetti model and the carrier recombination modeled by SRH recombination using a concentration-dependent lifetime model are given in detail in Ref. [4]. We note here that we have neglected radiative recombination and this is due to the fact that our simulations are done at room temperature [5]. The drift–diffusion model is used to calculate the current density of electrons and hole in the bulk and barrier regions and is given by the continuity equations [6]:

$$\frac{\partial n}{\partial t} = G - U + \frac{1}{q}(\nabla . J_n) \tag{1}$$

$$\frac{\partial p}{\partial t} = G - U - \frac{1}{q}(\nabla . J_p) \tag{2}$$
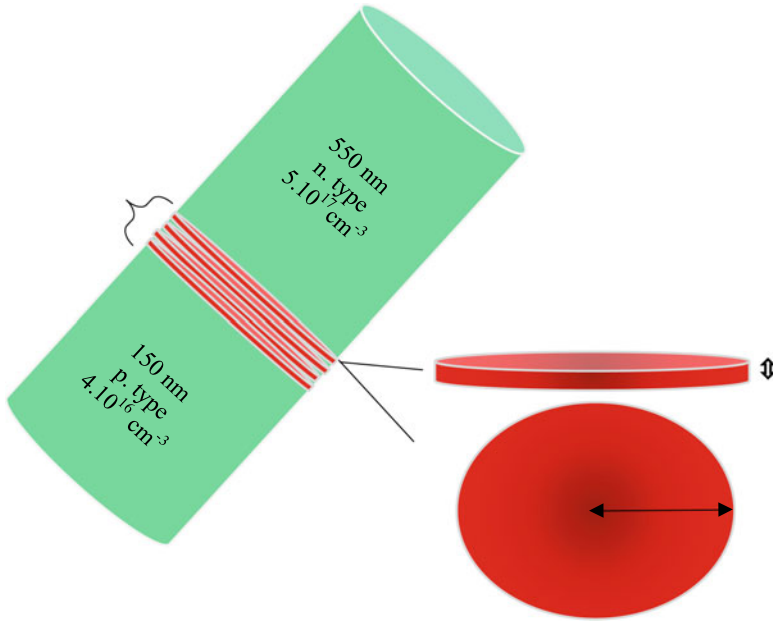
Fig. 1 Structure of Si-based nanowires by insertion of 4 CIGS quantum wells

In these equations, n and p are the electron and hole densities, respectively, and J_n and J_p are the electron and hole current densities, respectively. U and G are the Shockley–Read–Hall (SRH) recombination rate and the photogeneration rate, respectively. For the drift–diffusion model, the electron and hole currents are given by [7]:

$$\vec{j}_n = q\mu_p p \vec{\nabla} \vec{E} + q D_p \vec{J} n \tag{3}$$

$$\vec{j}_p = q\mu_p p \vec{\nabla} \vec{E} + q D_p \vec{J} p \tag{4}$$

D_n and D_p are the thermal diffusion of electrons and holes, respectively. E is the electric field, and μ_n and μ_p are the electron and hole mobilities, respectively. Assuming there is no dependence on θ because of the circular symmetry of the electrostatic potential, the Poisson equation is written [6]:

$$\frac{\partial^2 V}{\partial^2 z} + \frac{1}{r} \cdot \frac{\partial V}{\partial r} + \frac{\partial^2 V}{\partial^2 r} = \frac{-\rho(z, r)}{\varepsilon} \tag{5}$$

2.1 For Quantum Well Regions

The current continuity equations for quantum well systems are given as follows:

$$\frac{dn_{qw}}{dt} = G_{qw} - U_{qw} - \frac{n_b}{\tau_{cn}} - \frac{P_{qw}}{\tau_{en}} = 0 \tag{6}$$

$$\frac{dp_{qw}}{dt} = G_{qw} - U_{qw} - \frac{P_b}{\tau_{cp}} + \frac{P_{qw}}{\tau_{en}} = 0 \tag{7}$$

where U_{qw} and G_{qw} are the Shockley–Read–Hall (SRH) recombination rate and the quantum well photogeneration rate, respectively. Self-consistent coupled Schrödinger Poisson model is used to model the effects of quantum confinement, and it is given on 1D for electrons and holes, respectively, by:

$$\frac{-h^2}{2} \frac{\partial}{\partial x} \left(\frac{1}{m_e^*} \frac{\partial \Psi_i}{\partial x} \right) + E_c(x)\Psi_i - E_i \Psi_i = 0 \tag{8}$$

$$\frac{-h^2}{2} \frac{\partial}{\partial x} \left(\frac{1}{m_h^*} \frac{\partial \Psi_i}{\partial x} \right) - E_v(x)\Psi_i - E_i \Psi_i = 0 \tag{9}$$

where E_c and E_v are the conduction band and valence band edge, respectively, and m_e^* and m_h^* are the effective mass of electrons and holes, respectively. Ψ_i and E_i are the wave function and energy level of the i subband.

3 Results and Discussion

During this part of the work, we used the solar cell qualification conditions: A solar spectrum AM1.5 is a temperature of 300 K. We started the simulation of the *p–i–n* solar cell without QWs. Then, we moved to the simulation of a CIGS/Si QW solar cell for a fixed number of layers: 4 and 8 in order to see the effect of the QW insertion on the characteristic parameters of the cells. Then, we moved to the variation of the number of QWs in order to determine the number of optimal QWs that give us the best performance of the cell. For each simulation, the *J–V* characteristic, the EQE curve and the important characteristic parameters of a solar cell are presented and discussed. Figure 2 shows the evolution of the bandgap energy and the strain as a function of the mole fraction of Indium. It is clear that the bandgap energy increases with the increase of *x*.

On the other hand, we see that the strain decreases with the increase of *x*. Therefore, a good compromise between the bandgap energy and the strain must be reached at a certain optimal value of *x* for a good performance of the QW structure. For this purpose, we fixed the In concentration at $x = 0.1$ which corresponds to a strain of 4.85% and a bandgap energy of 1.06 eV.
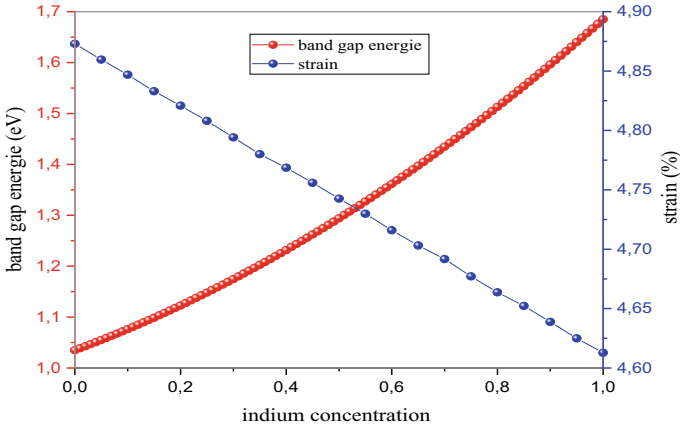
Fig. 2 Variation of the bandgap energy and strain as a function of the indium concentration

Figure 3 represents the external quantum efficiency (EQE) taken as a function of the wavelength of solar cell for both two structures of Si nanowires: pin and with insertion of 8 QW CIGS in the intrinsic region. It is clear that the EQE of a quantum well solar cell takes higher values from the wavelength of 500 nm. The EQE in the range 500–1200 nm of the CIGS/Si QW solar cell with 8 layers increases by a value of 7.12% for a wavelength of 800 nm, for example, which means that the less energetic photons in this wavelength range are effectively absorbed and not lost; thus, they contributed to the creation of electron–hole pairs. Figure 4 shows the current–voltage characteristic $J–V$ of a Si pin nanowire solar cell and a CIGS/Si solar cell by inserting of 4 and 8 quantum wells. As we can see, the short circuit current and open circuit voltage increase with a value of 2.49 mA/cm^2 and 0.08 V, respectively, with the insertion of 8 QWs. This can be explained by the fact that the insertion of QWs is indeed helpful for solar cell to absorb photon with low energy (photon with energy less than 1.12 eV), which means more electron–hole pairs are generated and contribute for creation of additional photocurrent. This improvement leads to an increase in the power delivered by this structure (Fig. 5). After choosing the appropriate concentration, as well as seeing the effect of QW insertion on the cell characteristics, we moved to the simulation of CIGS/Si nanowire-based solar cells by inserting 2, 4, 6, 8, 10, 12 and 14 quantum wells, respectively, to indicate the influence of quantum well introduction on the different structure characteristics. Figures 6 and 7 represent the variation of open circuit voltages and Form Factor as well as the variation of short circuit current and efficiency of a Si pin nanowire-based solar cell and a CIGS/Si solar cell for different numbers of QWs 2, 4, 6, 8, 10, 12 and 14, respectively. As we can see, the V_{oc}, FF and J_{sc} increase by a value of 0.12 V, 25.90% and 9.22 mA/cm^2, respectively, when the number of inserted QWs increases from 0 to 14 wells. Therefore, the efficiency increases by a value of 2.17%, and beyond this number, the efficiency keeps the same value. This is the number

Fig. 3 EQE of the solar cell without and with 8 CIGS/Si QWs
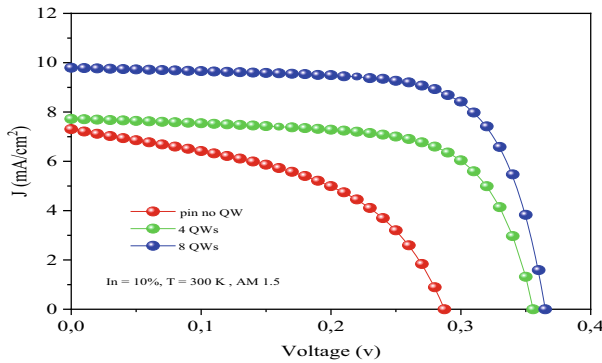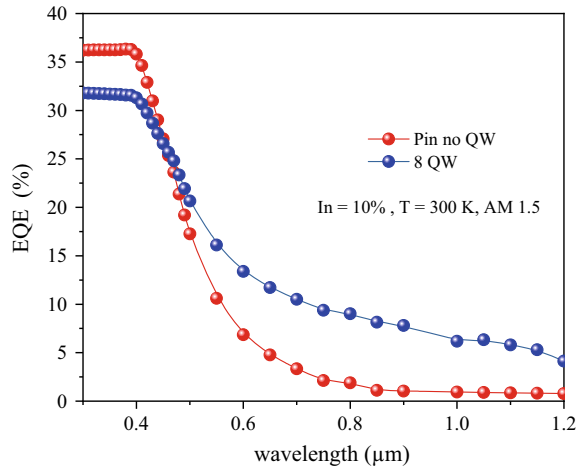
Fig. 4 *J–V* characteristic of the solar cell without and with 4 and 8 CIGS/Si QWs

of quantum wells for which we have saturation. All the characteristic parameters achieved are listed in Table 1.

4 Conclusion

We reported a modeling and characterization investigation of a single Si nanowire-based solar cell with CIGS quantum wells, mainly devoted to discern the effect of the number of wells on the performance of the cell. We approved in the first part that a concentration of 10% of In gives us a good compromise between the gap energy and the strain to reach the best performances of the cell. In the second part, we notice that the insertion of 14 wells in a nanowire increases the short circuit current and
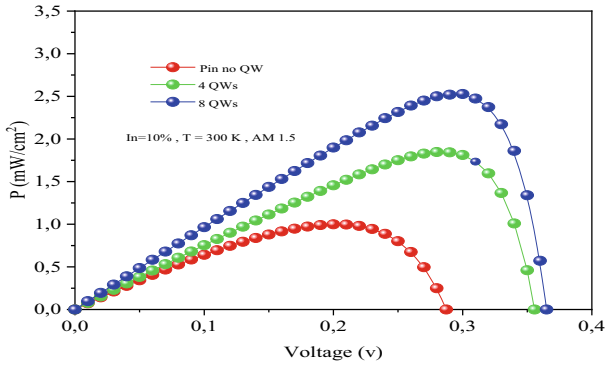
Fig. 5 *P–V* characteristic of the solar cell without and with 4 and 8 CIGS/Si QWs
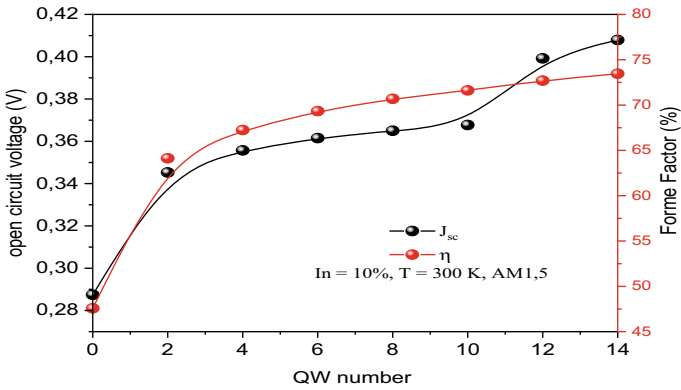
Fig. 6 Variation of the open circuit voltages and the form factor of the solar cell for a variable number of quantum wells
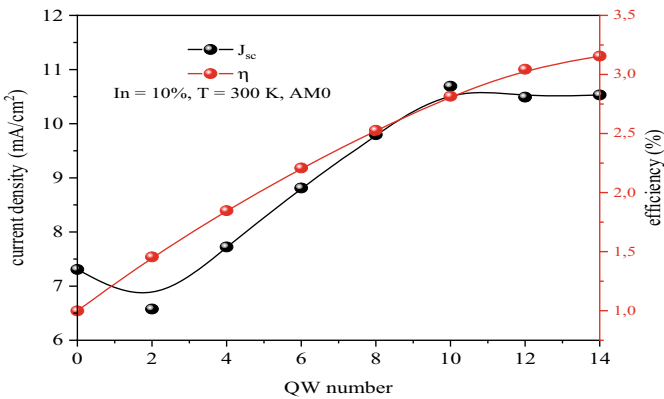
Fig. 7 Variation of open circuit voltages and the form factor of the pin Si and CIGS/Si solar cell for a varying number of quantum wells

Table 1 Characteristic parameters of the pin Si and CIGS/Si solar cell range from 0 to 14 quantum wells

QW number	J_{sc} (mA/cm^2)	V_{oc} (V)	FF (%)	η (%)
0	7.307	0.287	47.57	0.99
2	6.576	0.345	64.12	1.45
4	7.723	0.356	67.25	1.85
6	8.812	0.361	69.34	2.21
8	9.796	0.365	70.68	2.53
10	10.691	0.367	71.62	2.82
12	10.489	0.399	72.69	3.04
14	10.530	0.407	73.46	3.16

the efficiency by a factor of 1.44 and 3.19, respectively. The enhancement of the efficiency and the main parameter of the cell were approved through the study of the QW integration in a single nanowire.

References

1. Lehoucq G (2010) Top-down silicon nanowire transistors. Application to biological detection. Doctoral thesis. Ecole polytechnique Paris Tech, France
2. Rosaz G (2012) 3D integration of Si–SiGe nanowires for the production of vertical transistors 3D nanowire channel. University of Grenoble, France
3. Michallon J (2015) Study and optimization of optical absorption and electronic transport in nanowire based photovoltaic cells. Micro and nanotechnology, microelectronics. University of Grenoble Alpes
4. Aissat A, Bellil W, Bestam R, Vilcot JP (2014) Modeling and simulation of the structure based on the semiconductor III–V for solar cell application. J New Technol Mater (JNTM) 04(01):99–102
5. Zhang Y, Wu J, Aagesen M, Liu H (2015) III–V nanowires and nanowire optoelectronic devices. J Appl Phys 48:463001 (29p)
6. Munteanu D, Autran JL (2012) A 2-D/3-D Schrödinger-Poisson drift-diffusion numerical simulation of radially-symmetric nanowire MOSFETs. In: Nanowires—recent advances, pp 342–370 (Chapter 15)
7. Neu G, Bosacchi A, Franchi S (1991) Temperature dependence of the radiative and nonradiative recombination time in GaAs/AlGaAs quantum-well structures. Phys Rev B Condens Matter 44:3115

Intersubband Optical Properties of Strained InAsSb/AlGaAs Quantum Well Structure

L. Chenini, A. Aissat, and Jean Pierre Vilcot

Abstract In this paper, the intersubband optical properties of strained InAsSb/AlGaAs single quantum well structure have been investigated in a detailed way as depending on the structure parameters. For this aim, first, we have computed the energy eigenvalues and their corresponding wave functions of this structure by means of the solution of the Schrodinger equation in envelope function effective mass approximation. Second, we have calculated some important parameters as strain, band gap energy, effective mass and the conduction band (CB) offset ratio. Finally, the intersubband absorption coefficient properties have been determined as a function of the composition, well thickness and the emission wavelength.

Keywords Semiconductor III–V · Nanostructures · Absorption coefficient · Detection

1 Introduction

Theoretical understanding of the optical properties of quantum wells is of great importance from both basic and technological research. The optical properties related to intersubband transitions (ISBT) within the conduction band of quantum wells have attracted considerable attention during the last years due to the large values of dipole transition matrix elements, the strong quantum-confinement effect and possibility of achieving resonance states [1–6]. This technology has become a leading optoelectronics platform since the first observations of intersubbands transitions in GaAs/Al_xGa_{1-x}As multi-quantum wells by Esaki and Sasaki [7]. The Al_xGa_{1-x}As

L. Chenini · A. Aissat (✉)
Faculty of Technology, University Saad Dahlab Blida 1, Blida, Algeria
e-mail: sakre23@yahoo.fr; aa_aissat@univ-blida.dz

A. Aissat · J. P. Vilcot
Institut d'Electronique, de Microélectronique et de Nanotechnologie (IEMN), UMR 8520, Université des Sciences et Technologies de Lille 1, Avenue Poincaré, CS60069, 59652 Villeneuve-d'Ascq, France
e-mail: jean-pierre.vilcot@univ-lille.fr

and $InAs_{1-x}Sb_x$ systems have been the subject of considerable research and opto-electronic device development activities over the past years [7–14]. By using InAsSb for the well and AlGaAs as barrier materials, quantum well can have large conduction band offsets ratio Q_c, and hence, large intersubband transition energies will be obtained.

In addition to the lack of previous works and research on the effects of structural parameters on the InAsSb/AlGaAs quantum well structure, there has, to our knowledge to date, been no theoretical or experimental analysis, by intersubband transitions despite the great importance of this problem. So, in the present work, we will attempt to investigate theoretically the influence of the antimony composition x and aluminum fraction y in the barrier, well thickness and the emission wavelength on the intersubband absorption coefficient of the single $InAs_{1-x}Sb_x/Al_yGa_{1-y}As$ quantum well structure. A theoretical model used in this study is briefly described; this model is based principally on an interpolation scheme, and the effects of compositional variations in well and barrier regions are properly taken into account in the model and the effects of the well thickness and strain are also included. In the last section, we summarize briefly the conclusions obtained in the present study.

2 Theoretical Model

To find the eigenvalues and corresponding wave functions, we have to solve the Schrödinger equation:

$$(H - E_l)\Psi_l(z) = 0 \tag{1}$$

The Hamiltonian H is expressed as:

$$H = -\frac{\hbar^2}{2m_e^*}\nabla^2 + V_{CB}(z) \tag{2}$$

E_l is the energy of the bound state l, Ψ_l is the envelope wave function and m_e^* is the electron effective mass, $V_{CB}(z)$ is the CB confinement potential taken as follows:

$$V_{CB}(z) = \begin{cases} Q_C.\Delta E_g, & -L_w/2 < z < L_w/2 \\ 0, & \text{elsewhere} \end{cases} \tag{3}$$

where ΔE_g is the difference between the bandgap energies of InAsSb well region and AlGaAs barrier region, which will be given by the following expression:

$$\Delta E_g = 0.6x^2 + 0.37y^2 - 0.789x + 1.22y + 1.779 \tag{4}$$

Q_c is the conduction band offset ratio parameter, L_w is the well width.

Solving the Schrödinger equation, we will obtain:

$$E_l = -\frac{\hbar^2}{2m_e^*}\left(\frac{l\pi}{L_w}\right)^2, \quad l = 1, 2\ldots, \quad \Psi_l = \Psi_0 \sin\left(\frac{l\pi z}{L_w}\right) \tag{5}$$

The strain (ε) is an important element to verify. The difference lattice constant between the epitaxial InAsSb layer and the AlGaAs substrate results in a strain which forced the two lattice constants to be almost equal in the interface. However, there is no stress in the z direction; the strain can be given as a function of *x* and *y* as:

$$\varepsilon(\text{InAsSb/AlGaAs}) = \frac{a_2 - a_1}{a_1} = \frac{-0.421x + 0.0081y - 0.4051}{0.421x + 6.0584} \tag{6}$$

The subscript $a_{1,2}$ denotes the lattice constant for the InAsSb and AlGaAs layers, respectively. The physical parameters of the different binary materials used in this work are taken from Refs. [15, 16] and are summarized in Table1. For both, well (InAs$_{1-x}$Sb$_x$) and barrier (Al$_y$Ga$_{1-y}$As) regions, the corresponding bandgap energies are given, respectively, as the following quadratic relationships:

$$E_g(\text{InAsSb}) = (1 - x).E_g(\text{InAs}) + x.E_g(\text{InSb}) - 0.6.x(1 - x) \tag{7}$$

$$E_g(\text{AlGaAs}) = (1 - y).E_g(\text{GaAs}) + y.E_g(\text{AlAs}) - 0.37.y(1 - y) \tag{8}$$

x and *y* are, respectively, the antimony (Sb) and the aluminum (Al) compositions, and $E_g(\text{InAs})$, $E_g(\text{InSb})$, $E_g(\text{GaAs})$ and $E_g(\text{AlAs})$ are the energy gaps of the binary materials taken, respectively, as 0.359 eV, 0.17 eV, 1.42 eV and 3.01 eV at 300 K. All other parameters as lattice constant (*a*), elastic stiffness coefficients (C_{11} and C_{12}), conduction band deformation potential (a_c) ..., of the InAs$_{1-x}$Sb$_x$ and Al$_y$Ga$_{1-y}$As materials, are noted *P* and are obtained by a linear interpolation of the parameters of binary compounds given in Table 1, as the following expressions:

Table 1 Parameters of the different binary materials used in this work [15, 16]

Parameters	InAs	InSb	GaAs	AlAs
a (Å)	6.0583	6.4794	5.6533	5.6614
C_{11} (10^{11} dyn/cm^2)	8.329	6.608	11.88	11.93
C_{12} (10^{11} dyn/cm^2)	4.526	3.531	5.38	5.72
(ε) = $m*/m0$	0.024	0.013	0.067	0.124
a_c (eV)	−10.2	−15	−11	−5.64
a_v (eV)	1	0.36	−0.85	−2.60
b (eV)	−1.80	−2	−1.85	−2.30
Δ (eV)	0.39	0.81	0.341	0.28
$E_{v,av}$ (eV)	−6.67	−6.09	−6.92	−7.49

$$P(\text{InAs}_{1-x}\text{Sb}_x) = (1 - x).P(\text{InAs}) + x.P(\text{InSb}) \tag{9}$$

$$P(\text{Al}_y\text{Ga}_{1-y}\text{As}) = (1 - y).P(\text{GaAs}) + y.P(\text{AlAs}) \tag{10}$$

Except for the InAsSb, electron effective mass (m_e^*) and the spin orbit splitting energy (Δ) will be given, respectively, as:

$$m_e^*(\text{InAsSb}) = (1 - x).m_e^*(\text{InAs}) + x.m_e^*(\text{InSb}) - 0.027.x.(1 - x) \tag{11}$$

$$\Delta(\text{InAsSb}) = (1 - x).\Delta(\text{InAs}) + x.\Delta(\text{InSb}) - 1.2.x.(1 - x) \tag{12}$$

It is worth to note that the bowing parameter for the AlGaAs of these both parameters is equal to zero.

It is very important to evaluate the degree of carrier confinement in order to design and fabricate good and new electronic and optoelectronic devices. For this purpose, we have to calculate the conduction band offset ratio, defined as follows:

$$Q_C = \frac{E_c(\text{InAsSb}) - E_c(\text{AlGaAs})}{E_g(\text{InAsSb}) - E_g(\text{AlGaAs})} \tag{13}$$

$E_C(\text{InAsSb})$ and $E_C(\text{AlGaAs})$ are the conduction band extrema of the well and barrier regions, respectively.

The conduction band position for each material can be obtained as follow:

$$E_c = E_v + E_g + 2(a_c + a_v)\left(1 - \frac{C_{12}}{C_{11}}\right)\varepsilon + b\left(1 + \frac{2C_{12}}{C_{11}}\right)\varepsilon \tag{14}$$

where the valence band position E_v for each material can be given by:

$$E_v = E_{v,av} + \frac{\Delta}{3} + 2a_v\left(1 - \frac{C_{12}}{C_{11}}\right)\varepsilon + b\left(1 + \frac{2C_{12}}{C_{11}}\right)\varepsilon \tag{15}$$

where $E_{v,av}$ is the average valence band energy, (ε) is the spin orbit splitting energy, C_{11} and C_{12} are elastic stiffness constants, b the valence band shear deformation potential, a_c and a_v are the conduction and valence band hydrostatic deformation potentials. All these parameters are obtained as mentioned above by the linear interpolation method where the corresponding binary values for each material are listed in Table 1. The absorption coefficient (ε) of the conduction band ISBT in $\text{InAs}_{1-x}\text{Sb}_x/\text{Al}_y\text{Ga}_{1-y}\text{As}$ quantum well can be obtained by the following expression:

$$\alpha(\omega) = \frac{\omega}{L_w}\sqrt{\frac{\mu_0}{\varepsilon_0\varepsilon_r}}\sum|M_{nm}|^2(N_n - N_m)\frac{(\hbar/\tau)}{(E_{nm} - \hbar\omega)^2 + (\hbar/\tau)^2} \tag{16}$$

where μ_0 is the vacuum permeability, M_{nm} is the dipole moment matrix, ω is the photon frequency, ε_r is the relative dielectric constant and ε_0 is the vacuum dielectric constant, $N_n - N_m$ is the difference of the electron densities in the conduction subbands n and m, respectively, $E_{nm} = E_n - E_m$, E_n and E_m are the electron energy in the conduction subbands.

3 Results and Discussion

Figure 1 shows the variation of strain, according to the antimony and aluminum mole fractions. Growing InAsSb layers on AlGaAs substrate leads to appear a biaxial compressive strain due to the difference between the lattice parameters of the grown over layers and substrate one. Figure 1 shows that the strain developed in the InAsSb/AlGaAs single quantum well increases with increasing Sb concentration while it reduces with the increase of Al concentration. The InAs$_{1-x}$Sb$_x$ bandgap expression is plotted in Fig. 2. The figure indicates that the InAs$_{1-x}$Sb$_x$ bandgap energy has a nonlinear composition dependency, which is due to the presence of bowing coefficient ($C = 0.6$ eV) multiplied by x^2 term, as we have seen in Eq. (7).

The smallest bandgap energy of InAs$_{1-x}$Sb$_x$ alloy, corresponding to $E_g = 99.65$ meV, occurs for $x \approx 0.66$, a value close to $x \approx 0.60$–0.64 given in Ref. [17]. Several experimental data are collected from recently published papers [15, 18, 19]
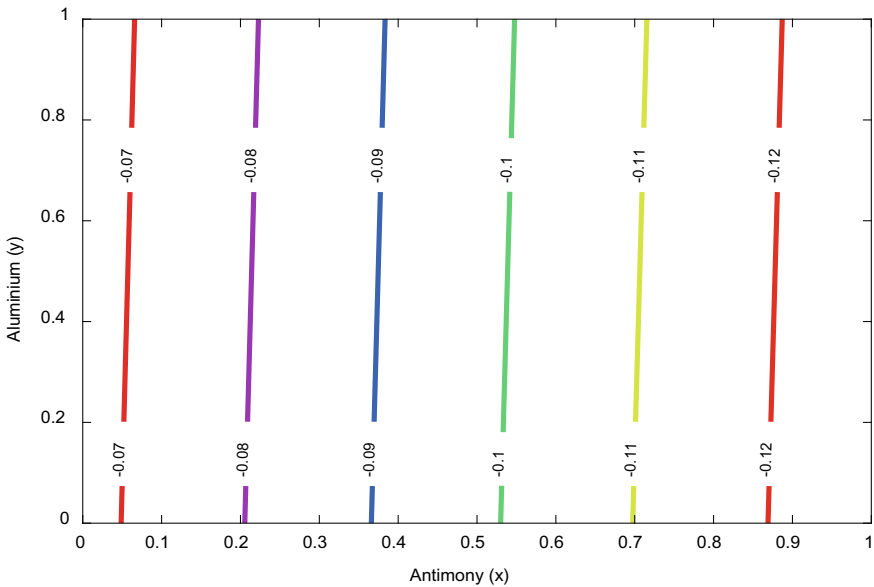
Fig. 1 Variation of strain as a function of the Sb and Al concentrations of the InAsSb/AlGaAs single quantum well
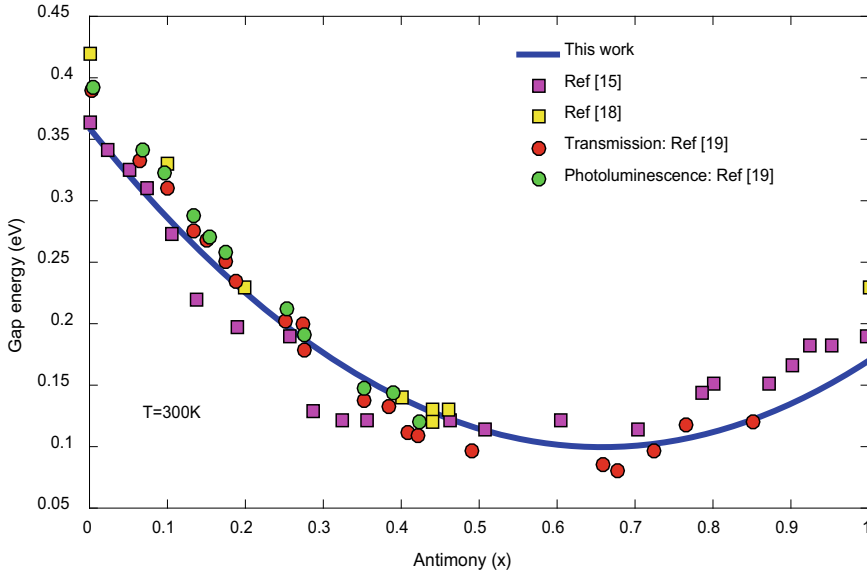
Fig. 2 Variation of the bandgap energy depending on the antimony concentration for ternary InAsSb structure at $T = 300$ K

and are included in Fig. 2. The experimental data are in agreement with the theoretical bandgap simulation of the InAsSb material at $T = 300$ K. Figure 3 illustrates the variation of electron effective mass in the well region as a function of x composition; the experimental effective masses gathered from Refs. [15, 20] were determined at $T = 300$ K. The electron effective mass for the InAs$_{1-x}$Sb$_x$ alloy system has a quadratic dependence with Sb composition. The smallest electron effective mass occurs for $x \approx 0.72$. A good agreement between experimental values of effective mass and our calculations is shown in Fig. 3.

The variation of Q_C with x and y is shown in Fig. 4. From this figure, we can see that increasing both antimony and aluminum composition decreases the band offset ratio. To have a Q_C value of 0.7, we have interest to choose, a little mole fraction of both compositions. So, the antimony mole fraction must be ranged between $0 \leq x \leq 0.25$ and the aluminum fraction in barrier will be taken also ranged between $0 \leq y \leq 0.25$.

In Fig. 5, the variation of the intersubband absorption coefficient with wavelength emission is given for different values of antimony, we set $L_W = 60$ Å and $y = 0.25$. It is clear that increasing the antimony quantity increases the absorption coefficient and a blue-shift toward higher energies will be have.

Effect of the aluminum composition in barrier region on the intersubband absorption coefficient is also investigated. The obtained results are given by Fig. 6, x is taken 0.25 and $L_W = 60$ Å. The peak of absorption coefficient decreases with increasing y and also blue-shift toward higher energies.
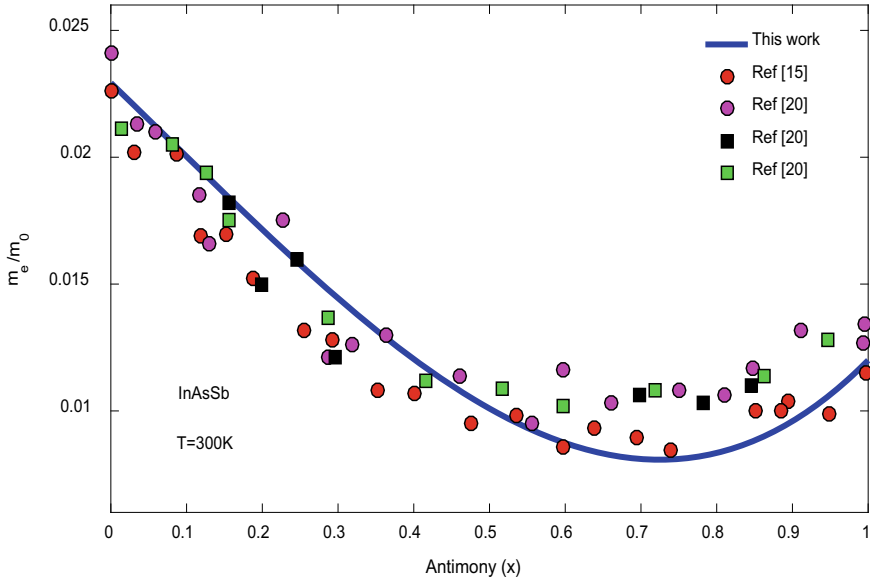
Fig. 3 Variation of the effective mass in the well region as a function of the x composition
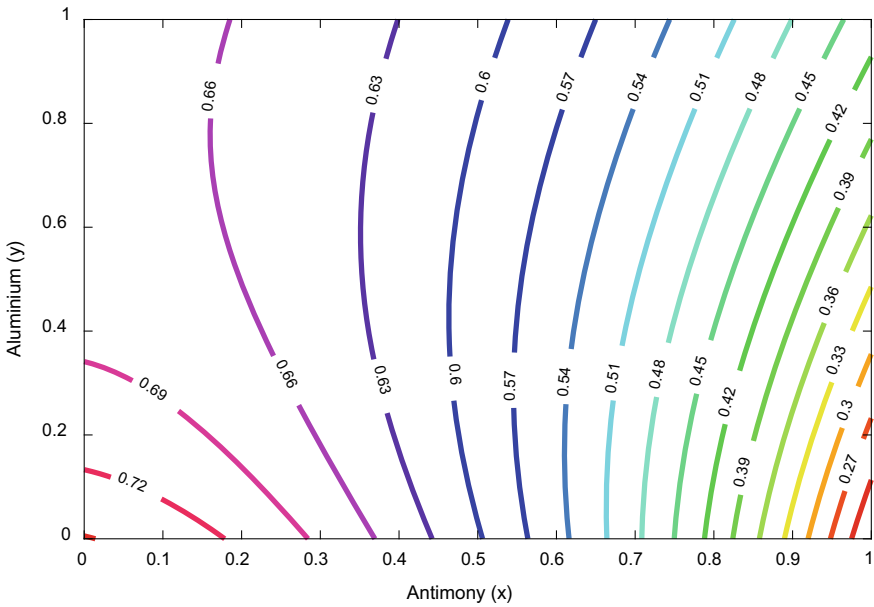
Fig. 4 Conduction band offset ratio Q_c of the InAs$_{1-x}$Sb$_x$/Al$_y$Ga$_{1-y}$As alloy
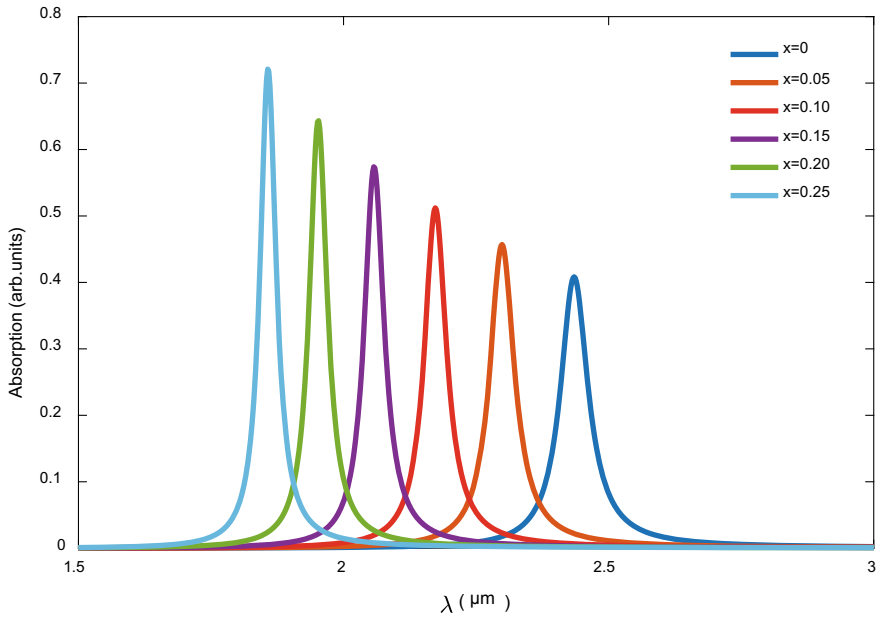
Fig. 5 Simulated intersubband absorption spectra of $InAs_{1-x}Sb_x/Al_{0.25}Ga_{0.75}As$ alloy versus emission wavelength for several antimony values ranged between $0 \leq x \leq 0.25$
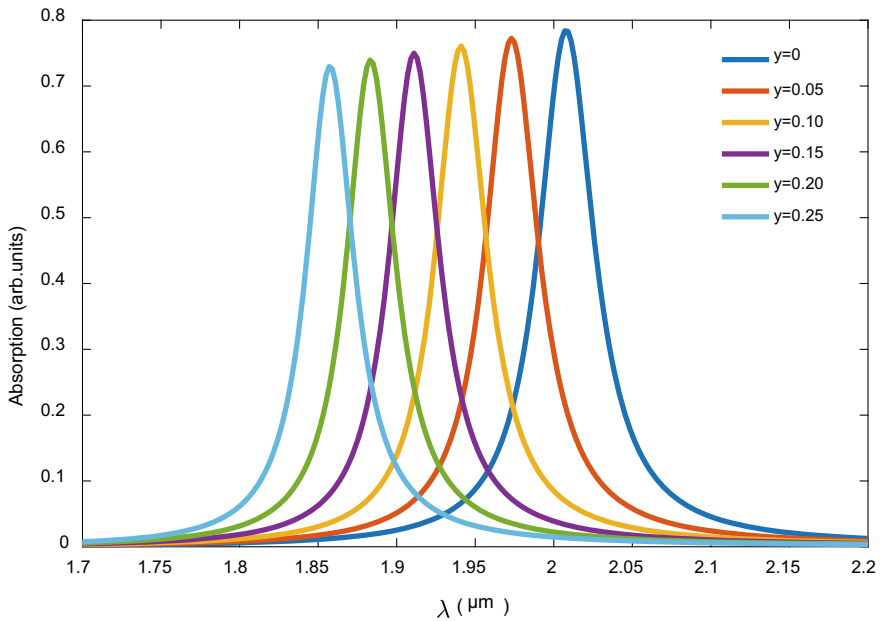
Fig. 6 Simulated intersubband absorption spectra of $InAs_{0.75}Sb_{0.25}/Al_yGa_{1-y}As$ alloy versus emission wavelength for several aluminum fractions in barrier ranged between $0 \leq y \leq 0.25$
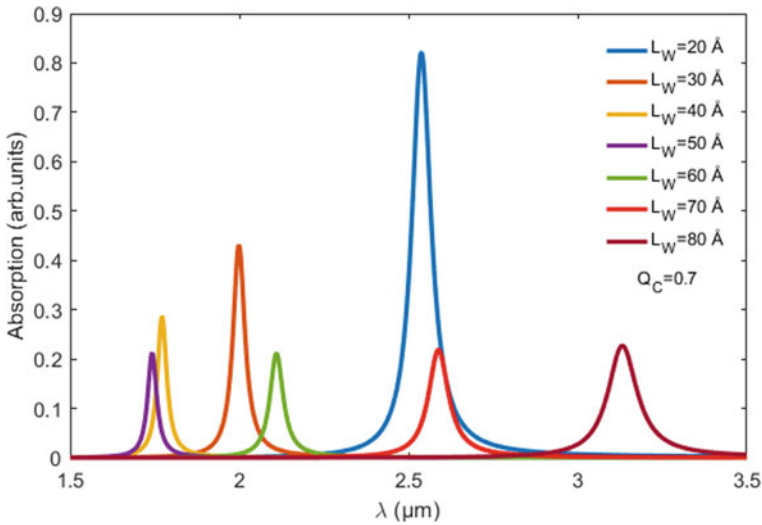
Fig. 7 Simulated intersubband absorption spectra of $InAs_{0.85}Sb_{0.15}/Al_{0.15}Ga_{0.85}As$ alloy versus emission wavelength for several thickness well values ranged between 20 and 80 Å

To show the influence of well width L_W on α, in Fig. 7, we set $x = 0.15$, $y = 0.15$, $T = 300$ K to have $Q_c = 0.70$ and plot the intersubband absorption coefficient as a function of the wavelength emission for different values of L_W. We can see that increasing the well width decreases the peak of the absorption coefficient for values of thickness less than $L_W = 50$ Å but above this value, α will increases. Also, we can see that for a well width values less than $L_W = 50$ Å, the peak of the absorption coefficient moves to the higher energies, but above this value, it shifts toward the lower energies.

4 Conclusion

In this paper, the intersubband optical absorption coefficients in strained $InAs_{1-x}Sb_x/Al_yGa_{1-y}As$ single quantum well are studied by solving the Schrödinger equation. Our results reveal that a red or a blue-shift can be obtained in the intersubband optical transitions as dependent on the shape of the InAsSb/AlGaAs quantum well and also by tuning the antimony concentrations of the InAsSb well and the aluminum concentrations of the AlGaAs barriers. Moreover, the theoretical results reveal that these parameters have significant effects on the peak intersubband absorption coefficients. This theoretical study could potentially give new possibilities in designing new devices based on the intersubband transitions of electrons and operating in the near infrared region.

References

1. Chenini L, Aissat A, Vilcot JP (2020) Investigation of temperature, well width and composition on the intersubband absorption of InGaAs/GaAs quantum wells. In: Proceedings of the 2nd international conference on electronic engineering and renewable energy systems. ICEERE 2020. Saidia, Morocco, 13–15 Apr 2020, pp 325–332

2. Nguyen T, Dzuba B, Cao Y, Senichev A, Diaz RE, Manfra JM, Malis O (2021) Mid infrared intersubband absorption in strain balanced non-polar (In)AlGaN/GaN multi quantum wells. Opt Mater Express 11(9):3284–3297

3. En-nadir R, El Ghazi H, Jorio A, Zorkani I, Kiliç HS (2022) Intersubband optical absorption in (In, Ga)N/GaN double quantum wells considering applied electric field effects. J Comput Electron 21(1):111–118

4. Chenini L, Aissat A, Vilcot JP (2019) Optimization of inter-subband absorption of InGaAsSb/GaAs quantum wells. Superlattices Microstruct 129(3):115–123

5. Kotani T, Arita M, Arakawa Y (2015) Doping dependent blue shift and linewidth broadening of intersubband absorption in non-polar m-plane AlGaN/GaN multiple quantum wells. Appl Phys Lett 107(11):112107

6. Chenini L, Aissat A, Vilcot JP (2018) Theoretical study of intersubband absorption coefficient in GaNAsBi/GaAs quantum wells structures. In: Proceedings of the 1st international conference on electronic engineering and renewable energy systems. ICEERE 2018. Saidia, Morocco, 15–17 Apr 2018, pp 216–224

7. Kurtz SR, Biefeld RM (1995) Magnetophotoluminescence of biaxially compressed InAsSb quantum wells. Appl Phys Lett 66(3):364–366

8. Casallas-Moreno YL, Ramírez-López M, Villa-Martínez G, Martínez-López AL, Macias M, Cruz-Orea A, Mendoza-Álvarez JG (2020) Effect of the Sb content and the n- and p-GaSb(100) substrates on the physical and chemical properties of $InSb_xAs_{1-x}$ alloys for mid-infrared applications: analysis of surface, bulk and interface. J Alloys Compd 861:157936 (1/32)

9. Semakova AA, Bazhenov NL, Mynbaev KD (2018) Luminescence of HgCdTe– and InAsSb-based quantum-well heterostructures, IOP Conf Ser J Phys Conf Ser 1135:012072 (1/4)

10. You JF, Zhao Q, Zhang ZH, Yuan JH, Guo KX, Feddi E (2019) The effect of temperature, hydrostatic pressure and magnetic field on the nonlinear optical properties of AlGaAs/GaAs semi-parabolic quantum well. Int J Mod Phys B 33(27):1950325 (1/12)

11. Abolghasemi A, Kohandani R (2018) Numerical investigation of strain effects on properties on AlGaAs/InGaAs multiple quantum well solar cells. Appl Opt 57(24):7045–7054

12. Geum DM, Kim S, Kang S, Kim H, Park H, Rho I, Ahn SY, Song J, Choi WJ, Yoon E (2018) Room temperature operation of mid-infrared $InAs_{0.81}Sb_{0.19}$ based photovoltaic detectors with an $In_{0.2}Al_{0.8}Sb$ barrier layer grown on GaAs substrates, Opt Express 26(5):6249–6259

13. Ladugin MA, Yarotskaya IV, Bagaev TA, Telegin KY, Andreev AY, Zasavitskii II, Padalitsa AA, Marmalyuk AA (2019) Advanced AlGaAs/GaAs heterostructures grown by MOVPE. Crystals 305(9):1–11

14. Kawazu T (2019) Valence band mixing in GaAs/AlGaAs quantum wells adjacent to self-assembled InAlAs antidots. J Nanomater 2019:1–7

15. Kasap S, Capper P (eds) Springer handbook of electronic and photonic materials. Springer International Publishing

16. Vurgaftman M, Meyer JR, Ram-Mohan LR (2011) Band parameters for III–V compound semiconductors and their alloys. Appl Phys Rev 89(11):5815–5875

17. Rogalski A, Martyniuk P, Kopytko M, Madejczyk. P, Krishna S (2020) InAsSb-based infrared photodetectors: thirty years later on. Sensores 20(24):7047 (1/72)

18. Svensson SP, Sarney WL, Hier H, Lin Y, Wang D, Donetsky D, Shterengas L, Kipshidze G, Belenky G (2012) Band gap of InAs$_{1-x}$Sb$_x$ with native lattice constant. Phys Rev B 86(24):245205 (1/6)
19. Murawski K, Gomółka E, Kopytko M, Grodecki K, Michalczewski K, Kubiszyn Ł, Gawron W, Martyniuk P, Rogalski A, Piotrowski J (2019) Bandgap energy determination of InAsSb epilayers grown by molecular beam epitaxy on GaAs substrates, Prog Nat Sci Mater Int 29(4):472–476
20. Rogalski A, Jozwikowski K (1989) Intrinsic carrier concentration and effective masses in InAs$_{1-x}$Sb$_x$. Infrared Phys 29(1):35–42

Numerical Simulation: Toward High-Efficiency CIGS Solar Cell Through Buffer Layer Replacement

Abdallah Bendoumou, Abderrahim Raidou, Atika Fahmi, Mohamed Lharch, and Mounir Fahoume

Abstract The present contribution reports on the performance of CIGS-based solar cells using Solar Cell Capacitance Simulator (SCAPS) to study the replacement of the CdS buffer layer by others. The principle idea behind this investigation is the improvement of CIGS solar cells conversion efficiency by using buffer layers of non-toxic and abundant materials. The four buffer layers of ZnSe, Zn(O, S), In_2S_3, and SnS_2 have substituted the CdS in the Glass/Mo/CIGS/CdS/ZnO/ZnO:Al/MgF$_2$ structure. SCAPS software was used to analyze the effects of buffer layer thickness, buffer layer donor density, buffer layer defect density, absorber layer thickness, and temperature on photovoltaic parameters: open-circuit voltage V_{oc}, short-circuit current density J_{sc}, Fill Factor FF, and efficiency η. The obtained results indicated that ZnSe, Zn(O, S), In_2S_3, and SnS_2 are good alternative buffer layers with an efficiency of around 28.3%.

Keywords Buffer layer · CIGS · Efficiency · Solar cell · SCAPS 1D

1 Introduction

Converting abundant solar energy directly into electricity represents a variable and very attractive option for cost-effective sustainable production of electrical energy [1]. Photovoltaic (PV) technology, which directly generates electricity from sunlight, is quickly becoming a major player in the future clean energy economy [2]. Most research efforts for the realization of low cost Cu(In, Ga)Se$_2$ (CIGS)-based solar cells have been dedicated to the development of non-vacuum processing methods for the preparation of this type of solar cells [3–6]. The Copper–Indium–Gallium–diSelenide (CIGS)-based solar cell is considered as one of the most promising thin film solar cells due to its many features such as higher efficiency both on module and cell levels, simple processes of manufacturing, excellent durability and stability,

A. Bendoumou · A. Raidou (✉) · A. Fahmi · M. Lharch · M. Fahoume
Materials Physics and Subatomic Laboratory, Department of Physics, Faculty of Science, Ibn Tofail University, B.P 133, 14000 Kénitra, Morocco
e-mail: abderrahim.raidou@uit.ac.ma

and radiation hardness [7]. One of the objectives in the field of CIGS technology in the next years is to improve or at least to maintain the efficiency of these cells by replacing the classical CdS buffer layer by wider band gap and Cd-free materials [8], because it is toxic and detrimental to the environment [9]. CdS is the most common buffer layer material used in the fabrication of CIGS-based solar cells [10]. The buffer layer is an intermediate layer film between the absorber and window layers with two main objectives, to provide structural stability to the device and to fix the electrostatic conditions inside the absorber layer [11, 12]. Usually, the semiconductor compounds with *n*-type conductivity and band gap between 0.2–3.6 eV are used as a buffer layer in the CIGS thin film solar cells [13]. The chalcopyrite solar cells employing alternative buffers can reach the same efficiencies as those with CdS buffers [14]. Moreover, solar cell simulation can be a valuable tool for gaining a deeper knowledge into the physical background of phenomena visible in electrical measurements [15].

In this present work, we studied numerically the typical structure Glass/Mo/CIGS/CdS/i-ZnO/*n*-ZnO:Al/MgF$_2$ using a Solar Cell Capacitance Simulator (SCAPS-1D) [16], with ZnSe [17, 18], Zn(O, S) [19–21], In$_2$S$_3$ [22–24], and SnS$_2$ [25, 26] as alternative buffer layers, and effects of its parameters: thickness, doping, and defect density on photovoltaic cell parameters such as open-circuit voltage V_{oc}, short-circuit current density J_{sc}, Fill Factor FF, and efficiency η. Also, we investigated CIGS thickness and temperature effects on the same output cell parameters.

2 Experimental Details

SCAPS is a one-dimensional solar cell simulation program developed at the Department of Electronics and Information Systems (ELIS) of the University of Ghent, Belgium [27]. Our numerical analysis is achieved by using the numerical simulation program SCAPS, that has been designed to simulate CIGS-based thin film solar cell devices [28].

The simulation is done under standard illumination AM1.5G, 1000 W/m^2, and 300 K. The material parameters of the Glass/Mo/CIGS/buffer_layer/i-ZnO/ZnO:Al/MgF$_2$ structure are obtained from literature, theoretical, experimental data, either by fitting the simulation to experimental data. The molybdenum (Mo) is used as a back metal contact and MgF$_2$ as an anti-reflection layer. The series resistance of 0.21 Ω and shunt resistance of 770 Ω are used. The SCAPS software is used to study and investigate the thickness, doping, and defect density effects of CdS, ZnSe, Zn(O, S), In$_2$S$_3$, and SnS$_2$ buffer layers on open-circuit voltage V_{oc}, short-circuit current density J_{sc}, Fill Factor FF, and efficiency η. And CIGS absorber layer thickness and temperature effects on the same photovoltaic parameters. Schematic structure of the simulated solar cell is illustrated in Fig. 1, which consists of the main layers: the CIGS absorber layer, CdS; ZnSe; Zn(O, S); In$_2$S$_3$; and SnS$_2$ buffer layers,

i-ZnO window layer, and *n*-ZnO:Al transparent conducting oxide (TCO) layer. Some input parameters used in this numerical calculation are displayed in Table 1.

To study the effect of ZnSe, Zn(S, O), In_2S_3, and SnS_2 buffer layers on the photo-voltaic parameters, we keep the same input parameters of the CdS material, except the band gap, electron affinity and relative dielectric permittivity, which are represented in Table 2.

Fig. 1 Schematic structure of CIGS solar cells
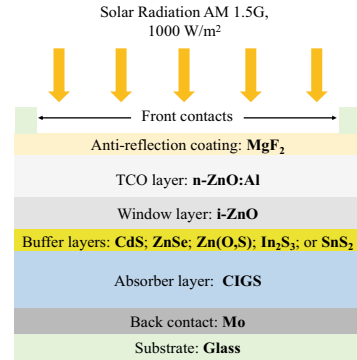
Table 1 Input electronic material parameters employed in the numerical study of the CIGS-based solar cell

Parameters	Layers			
	CIGS	CdS	i-ZnO	ZnO:Al
Thickness (nm)	Variable	Variable	100	150
E_g (eV)	1.15 [29]	2.4 [30]	3.3 [10]	3.3 [10]
χ (eV)	4.5 [29]	4.0 [30]	4.4 [10]	4.4 [10]
ε_r	13.6 [29]	10.0 [30]	9.0 [10]	9.0 [10]
N_c (cm^{-3})	2.2E+18	2.2E+18 [30]	2.2E+18 [31]	2.2E+18 [31]
N_v (cm^{-3})	1.8E+18	1.8E+19 [30]	1.8E+19 [31]	1.8E+19 [31]
v_{t-e} (cm/s)	1.0E+5	1.0E+5	1.0E+7	1.0E+7
v_{t-h} (cm/s)	1.0E+5	5.0E+4	1.0E+7	1.0E+7
μ_n (cm^2/Vs)	50	50	100	100
μ_p (cm^2/Vs)	25	25	25	25
N_D (cm^{-3})	0	Variable	1.0E+18 [31]	1.0E+20 [31]
N_A (cm^{-3})	5.0E+17	0.0	0.0	0.0
N_t (cm^{-3})	2.0E+15	Variable	1.0E+16 [31]	1.0E+16 [31]

Table 2 Input parameters of ZnSe, Zn(S, O), In_2S_3, and SnS_2 buffer layers

Parameters	Layers			
	ZnSe	Zn(S, O)	In_2S_3	SnS_2
E_g (eV)	2.9 [32]	2.7 [33, 34]	2.8 [35, 36]	2.24 [37]
χ (eV)	4.09 [32]	4.3 [33, 34]	4.25 [35]	4.24 [37]
ε_r	10 [32]	10 [33, 34]	13.5 [36]	10 [37]

3 Results and Discussion

3.1 Effect of Buffer Layer Thickness

Figure 2 illustrates the effect of CdS, ZnSe, Zn(O, S), In_2S_3, and SnS_2 buffer layers thickness, which have been varied in the range from 10 to 100 nm, on photovoltaic parameters: V_{oc}, J_{sc}, FF, and η while the CIGS absorber layer thickness was fixed of 2 μm. Figure 2a shows a very important increase of V_{oc} with thickness increase from 10 to 30 nm of CdS buffer layer and remains almost constant up to 40 nm. Afterward puts down a slight decrease. For ZnSe, Zn(O, S), In_2S_3, and SnS_2 buffer layers, the V_{oc} decreases slightly with thickness increase. The larger V_{oc} is obtained for structures having ZnSe, Zn(O, S), and SnS_2 buffer layers that possess a large band gap. For larger band gap buffer layers, the amount of photons absorbed outside the hole diffusion length region increases with increasing the buffer layer thickness which lowers the recombination rate and thus increasing the V_{oc} [38, 39]. In Fig. 2b, the J_{sc} is found almost constant with increasing of thickness from 10 to 30 nm for all buffer layers CdS, ZnSe, Zn(O, S), In_2S_3, and SnS_2. After that, it marks a drop with thickness increase. From Fig. 2c, d, it can be observed the FF and efficiency have almost the same variation, and optimal thickness is situated in the range 30–40 nm for CdS buffer layer with an efficiency of 27.2%. This result is in good agreement with the work of Moradi et al. [13, 40]. For other buffer layers, the optimum thickness is situated in the range of 10–20 nm. The best efficiency of around 28.3% is obtained for structures containing buffer layers ZnSe, Zn(O, S), In_2S_3, and SnS_2 at low thickness of 10–20 nm. When the thickness of buffer layer increases, a large number of photons are absorbed into the n-type buffer layer before arriving in the p-type absorber layer, which causes the reduction of the current and thus the efficiency, because of the higher recombination rate due to low diffusion length of the minority carriers of the buffer layer [37].

3.2 Effect of Buffer Layer Donor Density

The effect of buffer layer doping concentration N_D (cm^{-3}) on output parameters of CIGS solar cells is scrutinized. It has been varied from 10^{11} to 10^{22} cm^{-3}. Figure 3a
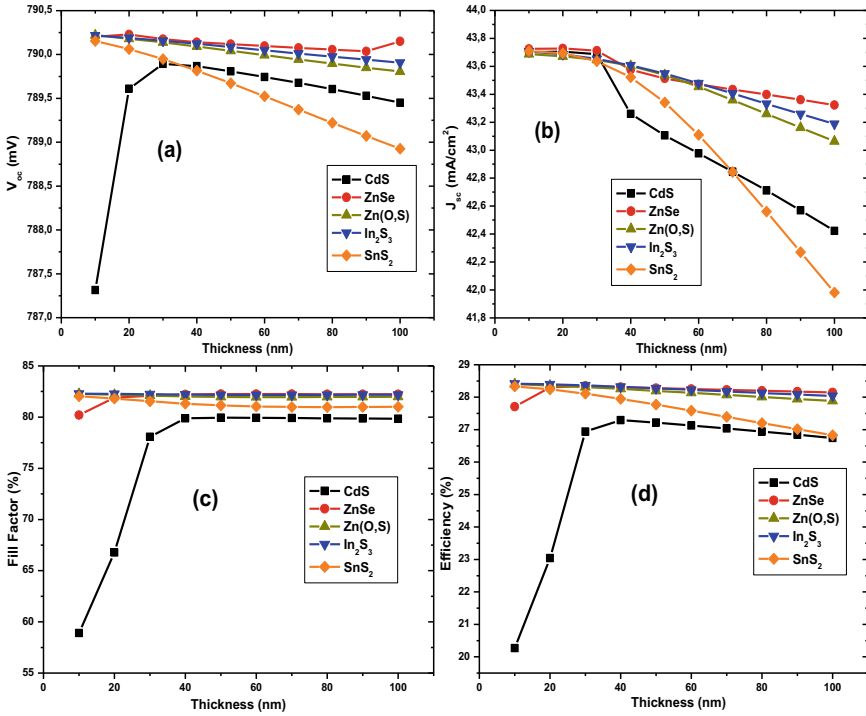
Fig. 2 Variation of: **a** V_{oc}, **b** J_{sc}, **c** FF and **d** η as a function of CdS, ZnSe, Zn(O, S), In$_2$S$_3$, and SnS$_2$ buffer layers thickness

presents the variation of V_{oc} as a function of doping concentration. It is clear, the V_{oc} of CdS/CIGS and ZnSe/CIGS structures is not influenced by the doping density up to 10^{15} cm^{-3} and afterward decreases to reach a minimum of 790 mV at 10^{18} cm^{-3} and remains constant for CdS/CIGS structure, and increases to reach a maximum value of 790 mV at 10^{18} cm^{-3} and remains constant for ZnSe/CIGS structure. Figure 3b–d shows the variation of J_{sc}, FF, and efficiency, respectively. It can be seen, the curves of J_{sc}, FF, and efficiency of CdS/CIGS and ZnSe/CIGS structures are similar in the three previous figures and reach their maximum values 43.5 mA/cm^2, 82.3%, and 28.3%, respectively, at 10^{18} cm^{-3}. For Zn(O,S)/CIGS, In$_2$S$_3$/CIGS, and SnS$_2$/CIGS structures, noticeable stability of V_{oc}, J_{sc}, FF, and efficiency is shown in Fig. 3a–d. Furthermore, in Fig. 3, all curves intersect at the carrier concentration 10^{18} (cm^{-3}) and have the same value above this point, indicating the optimum buffer layers donor density is 10^{18} cm^{-3}.
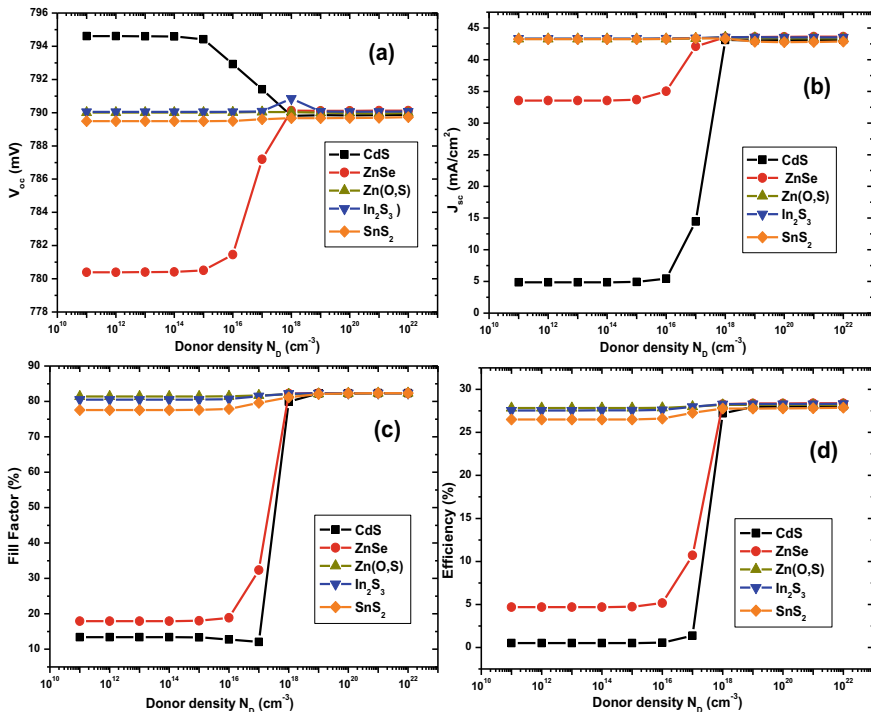
Fig. 3 Variation of: **a** V_{oc}, **b** J_{sc}, **c** FF, and **d** η as a function of CdS, ZnSe, Zn(O, S), In$_2$S$_3$, and SnS$_2$ buffer layer donor density N_D (cm^{-3})

3.3 Effect of Buffer Layer Defect Density

This subsection is dedicated to examining the impact of defect density N_t (cm^{-3}) in CdS, ZnSe, Zn(O, S), In$_2$S$_3$, and SnS$_2$ buffer layers on CIGS solar cell performance. The defect density has been varied from 10^{10} to 10^{20} cm^{-3}. The solar cell parameters V_{oc}, J_{sc}, FF, and η are obtained and reported in Fig. 4a–d, respectively. It can be observed that when N_t is less than 10^{14} cm^{-3} for the CdS buffer layer, and less than 10^{15} cm^{-3} for ZnSe, Zn(O, S), In$_2$S$_3$, and SnS$_2$ buffer layers, it has no influence on the solar cell parameters. The decrease of J_{sc} (Fig. 4b) is due to the reduction of the number of photogenerated carriers that must participate in the photocurrent; this observation can also be attributed to losses by deep penetration of low-energy photons, and the incomplete absorption of certain photons [41]. With the increase of N_t (Fig. 4d), the recombination rate also increases which in turn decreases the efficiency [42]. In this way, for better device performance it is necessary to keep N_t under 10^{14} cm^{-3} for CdS and 10^{15} cm^{-3} for ZnSe, Zn(O, S), In$_2$S$_3$, and SnS$_2$ buffer layers. The best efficiency of simulated CIGS solar cell for CdS/CIGS structure is 27.8%, and for structures such as ZnSe/CIGS, Zn(O, S)/CIGS, In$_2$S$_3$/CIGS, and SnS$_2$/CIGS have almost the same best efficiency equals to 28.5%.
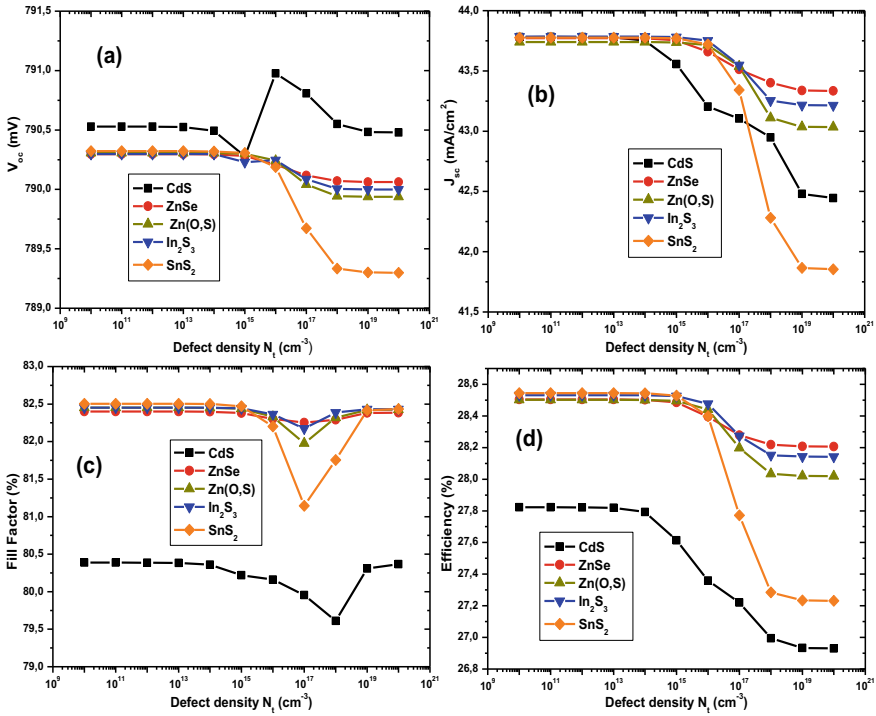
Fig. 4 Variation of: **a** V_{oc}, **b** J_{sc}, **c** FF and **d** η as a function of CdS, ZnSe, Zn(O, S), In$_2$S$_3$, and SnS$_2$ buffer layers defect density N_t (cm^{-3})

3.4 Effect of Absorber Layer Thickness

The device performance as a function of CIGS absorber layer thickness is performed for thickness in the range 400–3400 nm as illustrated in Fig. 5. The simulation results (Fig. 5a) show that open-circuit voltage (V_{oc}) increases strongly with the CIGS absorber layer thickness. It can be seen that all V_{oc} curves of CdS/CIGS, ZnSe/CIGS, Zn(O, S)/CIGS, In$_2$S$_3$/CIGS, and SnS$_2$/CIGS structures are congruent. Figure 5b reveals an increasing trend in current density (J_{sc}) with varying the absorber layer thickness. It is clear that as the thickness increases, current density also increases accordingly. Figure 5c exhibits an increasing of FF with the increase of CIGS absorber thickness up to 2800 nm for all buffer layers, then drops with the further thickness increasing, which can be attributed to the increasing series resistance [42]. Figure 5d illustrates an increasing trend in efficiency with CIGS layer thickness increase, and the best efficiency 29.82% is obtained for the ZnSe/CIGS, Zn(O, S)/CIGS, and In$_2$S$_3$/CIGS structures at a thickness of 3400 nm, 29.3 and 28.74% for SnS$_2$/CIGS and CdS/CIGS structures, respectively, at the same CIGS absorber layer thickness. Consequently, all photovoltaic parameters such as V_{oc}, J_{sc}, FF, and η are increased with increasing the thickness of CIGS, while at a low thickness of the absorber layer
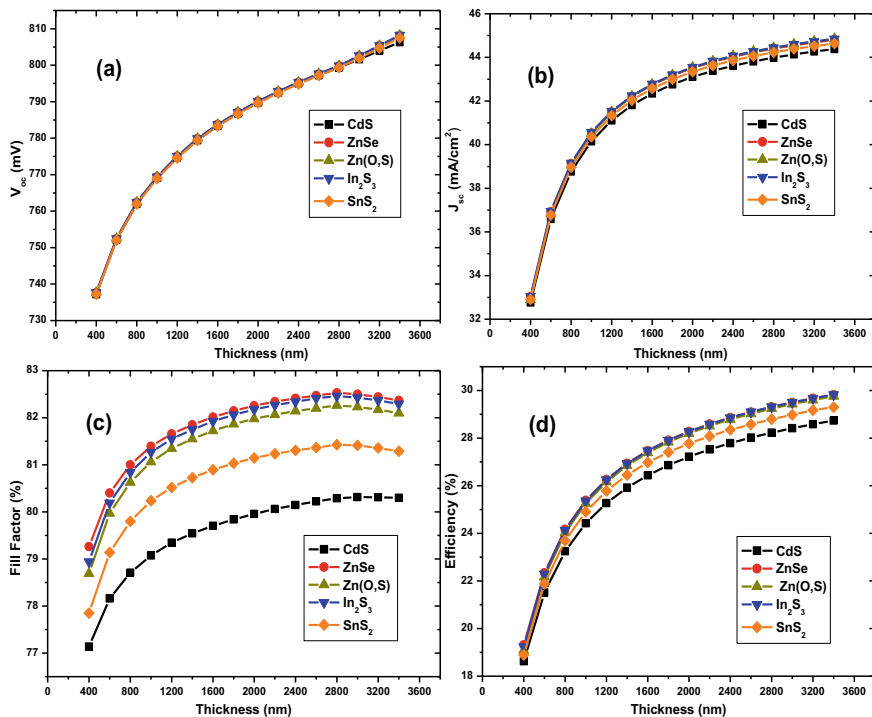
Fig. 5 Variation of: **a** V_{oc}, **b** J_{sc}, **c** FF, and **d** η as a function of CIGS absorber layer thickness

all parameters are low due to the high recombination of photogenerated carriers that recombine before reaching the buffer layer/CIGS layer interface [43, 44]. As the absorber layer thickness increases, it absorbs a large number of photons, which generates more and more electron–hole pairs, due to the built-in potential and electrical field, charge carriers get separated, which further rise in J_{sc} and V_{oc} will be observed, it increases the efficiency of the device [36, 45, 46].

3.5 Effect of Temperature

The study of the behavior of solar cells with temperature (T) is important since in terrestrial applications, they are generally exposed to temperatures ranging from 288 to 323 K and to even higher temperatures in space and concentrator systems [47]. Figure 6. presents the effect of temperature on photovoltaic parameters of CIGS solar cell. The temperature has been varied from 300 to 400 K, and the CIGS absorber layer thickness is kept constant at 2 μm. From Fig. 6a, it can be observed that all V_{oc} curves of CdS/CIGS, ZnSe/CIGS, Zn(O, S)/CIGS, In$_2$S$_3$/CIGS, and SnS$_2$/CIGS structures are congruent and drop linearly with temperature increase.

The decrease in open-circuit voltage causes electrons to gain sufficient energy at high temperatures and recombine with a hole before they could reach the depletion region and collected [48–50]. Figure 6b illustrates a slight increase of current density of CdS/CIGS and ZnSe/CIGS structures and a slight drop of Zn(O, S)/CIGS, In$_2$S$_3$/CIGS, and SnS$_2$/CIGS structures with operating temperature increase. In Fig. 6c, d, the Fill Factor and efficiency drop with increase in temperature. When we increase the temperature, the energy of electron increases; thus, the band gap of material decreases, and the electron gaining a large amount of energy recombines with holes; therefore, rate of recombination increases resulting in the decrement in V_{oc}, J_{sc}, FF, and efficiency [43, 51].

The best efficiency of simulated CIGS solar cell is obtained at room temperature 300 K for all structures such as 28.28% for ZnSe/CIGS, Zn(O, S)/CIGS, and In$_2$S$_3$/CIGS, 27.77% for SnS$_2$/CIGS, and 27.22% for CdS/CIGS.
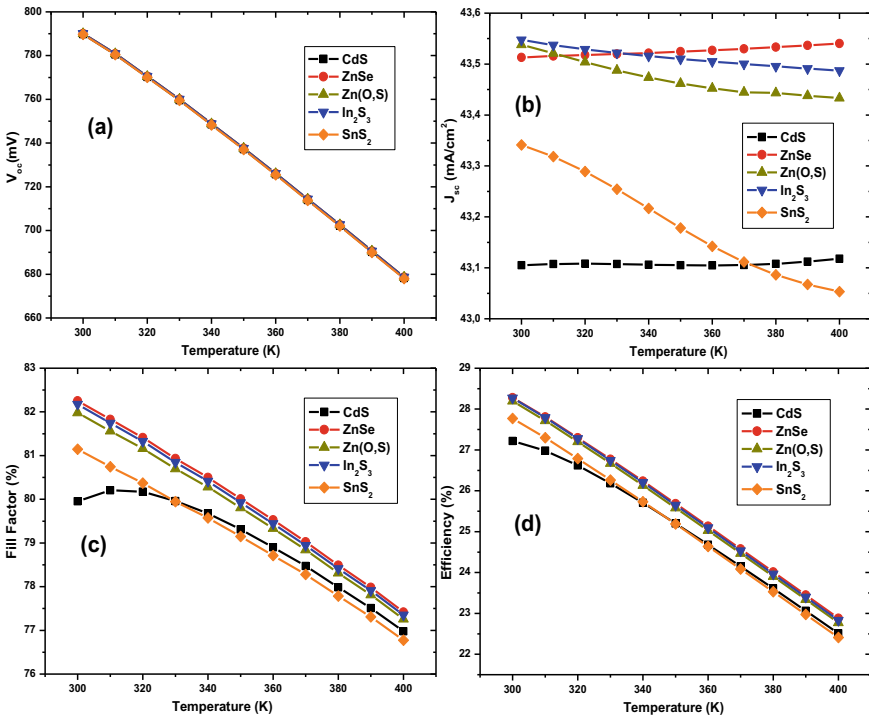
Fig. 6 Variation of: **a** V_{oc}, **b** J_{sc}, **c** FF, and **d** η as a function of temperature

4 Conclusion

In this study, the performance of CIGS-based solar cells has been investigated using CdS, ZnSe, Zn(O, S), In_2S_3, and SnS_2 buffer layers. Photovoltaic parameters were observed by varying buffer layer thickness, buffer layer donor density, buffer layer defect density, absorber layer thickness, and temperature. To get a larger efficiency, the optimum CdS buffer layer thickness is located in the range of 30–40 nm. For other buffer layers, the optimum thickness was found in the range of 10–20 nm. Optimum donor density and defect density are 10^{18} cm^{-3} and 10^{14}–10^{15} cm^{-3}, respectively, for all buffer layers. It has been found that the efficiency increased with the CIGS absorber layer thickness, and it was highest at room temperature.

Furthermore, for all previously studied parameters, our simulation always revealed the best efficiency of about 28.3% for CIGS-based solar cells using ZnSe, Zn(O, S), In_2S_3, and SnS_2 as good alternative buffer layers, which are non-toxic and accessible materials.

References

1. Reinhard P, Chirilă A, Blösch P, Pianezzi F, Nishiwaki S, Buecheler S, Tiwari AN (2013) Review of progress toward 20% efficiency flexible CIGS solar cells and manufacturing issues of solar modules. IEEE J Photovoltaics 3:572–580
2. Powalla M, Paetel S, Hariskos D, Wuerz R, Kessler F, Lechner P, Wischmann W, Friedlmeier TM (2017) Advances in cost-efficient thin-film photovoltaics based on Cu(In, Ga)Se$_2$. Engineering 3:445–451
3. Jang J, Lee JS, Hong K-H, Lee D-K, Song S, Kim K, Eo Y-J, Yun JH, Chung C-H (2017) Cu(In, Ga)Se$_2$ thin film solar cells with solution processed silver nanowire composite window layers: buffer/window junctions and their effects. Sol Energy Mater Sol Cells 170:60–67
4. Hibberd CJ, Chassaing E, Liu W, Mitzi DB, Lincot D, Tiwari AN (2010) Non-vacuum methods for formation of Cu(In, Ga)(Se, S)$_2$ thin film photovoltaic absorbers. Prog Photovolt Res Appl 18:434–452
5. Chung C-H, Hong K-H, Lee D-K, Yun JH, Yang Y (2015) Ordered vacancy compound formation by controlling element redistribution in molecular-level precursor solution processed CuInSe$_2$ thin films. Chem Mater 27:7244–7247
6. Todorov T, Mitzi DB (2010) Direct liquid coating of chalcopyrite light-absorbing layers for photovoltaic devices. Eur J Inorg Chem 2010:17–28
7. Bouabdelli MW, Rogti F, Maache M, Rabehi A (2020) Performance enhancement of CIGS thin-film solar cell. Optik 216:164948
8. Serhan J, Djebbour Z, Darga A, Mencaraglia D, Naghavi N, Renou G, Lincot D, Guillemeoles J-F (2010) Electrical characterization of CIGSe solar cells metastability with Zn(S, O, OH)–ZnMgO interface buffer layers. Sol Energy Mater Sol Cells 94:1884–1888
9. Chen J, Shen H, Zhai Z, Li Y, Li S (2020) Cd-free Cu(InGa)Se$_2$ solar cells with eco-friendly a-Si buffer layers. Appl Surf Sci 512:145729
10. Saadat M, Amiri O, Rahdar A (2019) Optimization of (Zn, Sn)O buffer layer in Cu(In, Ga)Se$_2$ based solar cells. Sol Energy 189:464–470

11. Mostefaoui M, Mazari H, Khelifi S, Bouraiou A, Dabou R (2015) Simulation of high efficiency CIGS solar cells with SCAPS-1D software. Energy Proc 74:736–744

12. Jamil N, Abdulla S, Muhammed A, Pogrebnjak A, Ivashchenko M (2012) Design and fabrication heterojunction solarcell of Si–CdS–ZnO thin film. Presented at the proceedings of the international conference nanomaterials: applications and properties

13. Moradi M, Teimouri R, Saadat M, Zahedifar M (2017) Buffer layer replacement: a method for increasing the conversion efficiency of CIGS thin film solar cells. Optik 136:222–227

14. Heriche H, Rouabah Z, Bouarissa N (2016) High-efficiency CIGS solar cells with optimization of layers thickness and doping. Optik 127:11751–11757

15. Pettersson J, Platzer-Björkman C, Zimmermann U, Edoff M (2011) Baseline model of graded-absorber Cu(In, Ga)Se$_2$ solar cells applied to cells with Zn$_{1-x}$Mg$_x$O buffer layers. Thin Solid Films 519:7476–7480

16. Burgelman M, Nollet P, Degrave S (2000) Modelling polycrystalline semiconductor solar cells. Thin Solid Films 361–362:527–532

17. Qu J, Zhang L, Wang H, Song X, Zhang Y, Yan H (2019) Simulation of double buffer layer on CIGS solar cell with SCAPS software. Opt Quant Electron 51:383

18. Ghavami F, Salehi A (2020) High-efficiency CIGS solar cell by optimization of doping concentration, thickness and energy band gap. Mod Phys Lett B 34:2050053

19. Ramanathan K, Mann J, Glynn S, Christensen S, Pankow J, Li J, Scharf J, Mansfield L, Contreras M, Noufi R (2012) A comparative study of Zn(O, S) buffer layers and CIGS solar cells fabricated by CBD, ALD, and sputtering. In: Conference record of the IEEE photovoltaic specialists conference

20. Garris RL, Mansfield LM, Egaas B, Ramanathan K (2017) Low-Cd CIGS solar cells made with a hybrid CdS/Zn(O, S) buffer layer. IEEE J Photovoltaics 7:281–285

21. Gour KS, Parmar R, Kumar R, Singh VN (2020) Cd-Free Zn(O, S) as alternative buffer layer for chalcogenide and Kesterite based thin films solar cells: a review. J Nanosci Nanotechnol 20:3622–3635

22. Witte W, Spiering S, Hariskos D (2014) Substitution of the CdS buffer layer in CIGS thin-film solar cells. Vak Forsch Prax 26:23–27

23. Khoshsirat N, Md Yunus NA (2016) Numerical analysis of In$_2$S$_3$ layer thickness, band gap and doping density for effective performance of a CIGS solar cell using SCAPS. J Electr Mater 45:5721–5727

24. Osman Y, Fedawy M, Abaza M, Aly M (2020) Solar cell performance enhancement with optimized CIGS absorber bandgap and buffer layer. J Phys Conf Ser 1447:012057

25. Jariwala A, Chaudhuri TK, Toshniwal A, Patel S, Kheraj V, Ray A, Mukhopadhyay I, Ray A, Pati R (2018) SnS$_2$ films deposited from molecular ink as Cd-free alternative buffer layer for solar cells. AIP Conf Proc 1961:030026

26. Ullah S, Bouich A, Ullah H, Mari B, Mollar M (2020) Comparative study of binary cadmium sulfide (CdS) and tin disulfide (SnS$_2$) thin buffer layers. Sol Energy 208:637–642

27. Burgelman M, Decock K, Khelifi S, Abass A (2013) Advanced electrical simulation of thin film solar cells. Thin Solid Films 535:296–301

28. Decock K, Zabierowski P, Burgelman M (2012) Modeling metastabilities in chalcopyrite-based thin film solar cells. J Appl Phys 111:043703

29. Mishra S, Bhargava K, Deb D (2019) Numerical simulation of potential induced degradation (PID) in different thin-film solar cells using SCAPS-1D. Sol Energy 188:353–360

30. Mathur AS, Singh BP (2020) Study of effect of defects on CdS/CdTe heterojunction solar cell. Optik 212:164717

31. Bauer A, Sharbati S, Powalla M (2017) Systematic survey of suitable buffer and high resistive window layer materials in CuIn$_{1-x}$Ga$_x$Se$_2$ solar cells by numerical simulations. Sol Energy Mater Sol Cells 165:119–127

32. Tousif MdN, Sakib M, Ferdous AA, Hoque MdA (2018) Investigation of different materials as buffer layer in CZTS solar cells using SCAPS. JOCET 6:293–296

33. Zhang H, Cheng S, Yu J, Lai Y, Zhou H, Zheng Q, Jia H (2017) Effect of Zn(O, S) buffer layer on Cu$_2$ZnSnS$_4$ solar cell performance from numerical simulation. J Appl Sci Eng 20:39–46

34. Buffière M, Harel S, Guillot-Deudon C, Arzel L, Barreau N, Kessler J (2015) Effect of the chemical composition of co-sputtered Zn(O, S) buffer layers on Cu(In, Ga)Se$_2$ solar cell performance. Phys Status Solidi A 212:282–290

35. Baig F, Khattak YH, Beg S, Soucase BM (2019) Numerical analysis of a novel CNT/Cu$_2$O/Sb$_2$Se$_3$/In$_2$S$_3$/ITO antimony selenide solar cell. Optik 197:163107

36. Tripathi S, Sadanand, Lohia P, Dwivedi DK (2020) Contribution to sustainable and environmental friendly non-toxic CZTS solar cell with an innovative hybrid buffer layer. Sol Energy 204:748–760

37. Belarbi F, Rahal W, Rached D, Benghabrit S, Adnane M (2020) A comparative study of different buffer layers for CZTS solar cell using Scaps-1D simulation program. Optik 216:164743

38. Yasin S, Waar ZA, Zoubi TA (2020) Development of high efficiency CZTS solar cell through buffer layer parameters optimization using SCAPS-1D. Mater Today Proc 33:1825–1829

39. Green MA (1982) Solar cells: operating principles, technology, and system applications. Prentice-Hall, Englewood Cliffs

40. Moradi M, Teimouri R, Zahedifar M, Saadat M (2016) Optimization of Cd$_{1-y}$Zn$_y$S buffer layer in Cu(In, Ga)Se$_2$ based thin film solar cells. Optik 127:4072–4075

41. Guirdjebaye N, Ngoupo AT, Ouédraogo S, Tcheum GLM, Ndjaka JMB (2020) Numerical analysis of CdS-CIGS interface configuration on the performances of Cu(In, Ga)Se$_2$ solar cells. Chin J Phys 67:230–237

42. Lin L, Jiang L, Qiu Y, Fan B (2018) Analysis of Sb$_2$Se$_3$/CdS based photovoltaic cell: a numerical simulation approach. J Phys Chem Solids 122:19–24

43. Sadanand, Dwivedi DK (2019) Theoretical investigation on enhancement of output performance of CZTSSe based solar cell. Sol Energy 193:442–451

44. Wanda MD, Ouédraogo S, Ndjaka JMB (2019) Theoretical analysis of minority carrier lifetime and Cd-free buffer layers on the CZTS based solar cell performances. Optik 183:284–293

45. Simya OK, Mahaboobbatcha A, Balachander K (2016) Compositional grading of CZTSSe alloy using exponential and uniform grading laws in SCAPS-ID simulation. Superlattices Microstruct 92:285–293

46. Bag A, Radhakrishnan R, Nekovei R, Jeyakumar R (2020) Effect of absorber layer, hole transport layer thicknesses, and its doping density on the performance of perovskite solar cells by device simulation. Sol Energy 196:177–182

47. Heriche H, Rouabah Z, Bouarissa N (2017) New ultra thin CIGS structure solar cells using SCAPS simulation program. Int J Hydrogen Energy 42:9524–9532

48. Sadanand, Dwivedi DK (2020) Modeling of CZTSSe solar photovoltaic cell for window layer optimization. Optik 222:165407

49. Lin P, Lin L, Yu J, Cheng S, Lu P, Zheng Q (2014) Numerical simulation of Cu$_2$ZnSnS$_4$ based solar cells with In$_2$S$_3$ buffer layers by SCAPS-1D. J Appl Sci Eng 17:383–390

50. Li J, Wang H, Luo M, Tang J, Chen C, Liu W, Liu F, Sun Y, Han J, Zhang Y (2016) 10% efficiency Cu$_2$ZnSn(S, Se)$_4$ thin film solar cells fabricated by magnetron sputtering with enlarged depletion region width. Sol Energy Mater Sol Cells 149:242–249

51. Green MA (2003) General temperature dependence of solar cell performance and implications for device modelling. Prog Photovolt Res Appl 11:333–340

Influence of an Alternating Phase on the Electron Heating in Capacitively Coupled Radio-Frequency Discharges

Abdelhak Missaoui, Morad El kaouini, and Hassan Chatei

Abstract In this work, we use a fluid approach to investigate the effect of including an alternating phase on the electron heating mechanism in capacitively coupled radio-frequency discharges. The fluid model coupled to the Poisson equation is numerically solved by using the finite difference method. The obtained results show that by including an alternating phase in the excitation waveform, the electron density increases in the bulk plasma region, while the electron heating increases in the sheath regions where electrons are heated and cooled alternatively during a radio-frequency cycle. Moreover, it is found that an increase in the pressure or the driving frequency can contribute to a more significant increase in electron heating when an alternating phase is added.

Keywords Electron heating · RF plasma · Low pressure · Fluid model · Driving frequency

1 Introduction

In recent years, the micro and nano-electronic industries have seen a great development in the miniaturization of electronic circuit chips [1]. Capacitively coupled radio-frequency (RF) plasma discharges operated at low pressure are interesting in microelectronic applications, where these discharges are essential for etching, deposition of thin films and modification of the surface properties [2]. The characteristics of the radio-frequency discharges can be influenced by several parameters such as the pressure of the background gas, the driving frequency, the secondary electron emission and the excitation mode of the discharge. These radio-frequency discharges are

A. Missaoui (✉) · H. Chatei
Faculty of Sciences, BV Mohammed VI, Mohammed I University, B.P. 524, Oujda 60000, Morocco
e-mail: abdelhakmissaoui92@gmail.com

M. E. kaouini
Polydisciplinary Faculty of Nador, Mohammed I University, B.P 300, Selouane Nador62700, Morocco

generally driven by a single frequency sinusoidal waveform or by a multi-frequency waveform [3, 4]. Recently, it has been shown that dual-frequency excitations can influence the plasma parameters in capacitive discharges [5]. On the other hand, a multi-frequency excitation waveform provides advantages for many plasma applications. It has been found that these waveforms could optimize the deposit uniformity and the deposition rate in the plasma-enhanced chemical vapor deposition (PECVD) process [6, 7]. Meanwhile, the electron heating which plays a crucial role in sustaining the discharge was also examined in a radio-frequency discharge driven by a dual-frequency excitation and by a multi-frequency excitation waveform where the phase was considered zero or constant [5, 8].

However, the effect of an alternating phase on the electron heating has not yet been carried out. In this work, we focus on the effect of including an alternating phase on the electron heating mechanism in a radio-frequency argon plasma discharge by using a fluid model.

The present paper is organized as follows. In Sect. 2, we describe the mathematical model. In Sect. 3, the numerical results are presented and discussed. Finally, the conclusion is drawn in Sect. 4.

2   Model Description

In this simulation, we use a fluid model consisting of the continuity equations for the charged particles (electrons, ions), the momentum transfer equations written in the the drift-diffusion approximation and the electron energy balance equation which are, respectively, given as

$$\frac{\partial n_k}{\partial t} + \frac{\partial J_k}{\partial x} = S \tag{1}$$

$$J_k = \mp n_k \mu_k E - D_k \frac{\partial n_k}{\partial x} \tag{2}$$

$$\frac{\partial}{\partial t}\left(\frac{3}{2} n_e k_B T_e\right) = -\frac{\partial q_e}{\partial x} - e J_e . E - H_i S \tag{3}$$

where $k = (e, +)$ refer to electrons and positive ions, respectively, n_k is the number density of plasma species k and S is the source term. The mobility coefficients μ_k and the diffusion coefficients D_k for the charged particles in argon gas are taken to be the same to those used in the Ref. [9]. In order to determine the electric field in the discharge, the fluid equations are coupled in a self-consistent way to the Poisson equation given by

$$\frac{\partial^2 V}{\partial x^2} = \frac{e}{\varepsilon_0}(n_e - n_+), \quad E = -\frac{\partial V}{\partial x}. \tag{4}$$

where ε_0 is the permittivity of free space. V and E are, respectively, the electric potential and the electric field. The first and the last terms on the right-hand side of Eq. (3) are, respectively, the gradient of the conductive heat flux of electron energy and the electron energy loss [10].

The term $-eJ_eE$ represents the total electron heating rate which includes two parts and can be expressed as

$$S_h = eD_e \frac{\partial n_e}{\partial x} E + e\mu_e n_e E^2 \tag{5}$$

The first term in Eq. (5) represents the electron pressure heating while the second part is the electron Ohmic heating.

The problem requires the resolution of the coupled system of Eqs. (1)–(4) with the boundary and the initial conditions as those given in Refs. [5, 10].

At the upper electrode $(x = D)$

$$J_+ = n_+\mu_+E, \quad J_e = k_s n_e - \gamma J_+, \quad T_e = 0.5(eV) \quad \text{and} \quad V = 0. \tag{6}$$

At the lower electrode $(x = 0)$

$$J_+ = n_+\mu_+E, \quad J_e = -k_s n_e - \gamma J_+, \quad T_e = 0.5(eV) \quad \text{and} \quad V = V_a. \tag{7}$$

where V_a is the applied voltage, γ is the secondary electron emission coefficient and k_s is the electron recombination coefficient. The initial conditions used in this simulation are as follows

$$n_e = n_+ = n_\varepsilon \left[\varepsilon + 16\left(1 - \frac{x}{D}\right)^2 \left(\frac{x}{D}\right)^2\right], \quad T_e = 1(eV) \quad \text{and} \quad V = 0. \tag{8}$$

Here, ε is a small positive number and n_ε is the initial density of the charged particles [5].

3 Results and Discussion

In order to study the electron heating in the discharge under the effect of an excitation voltage with and without including the alternating phase, we present in this section the simulation results obtained after 5000 radio-frequency cycles. The argon was used as the background gas and the discharge geometry adopted in this study consists of two symmetrical large electrodes separated by the distance $D = 2.54$ cm. The schematic of the simulated region and the applied voltages used in this study is given in Fig. 1. The first excitation waveform applied without an alternating phase is given as

$$V_a = V_{rf}\sin(2\pi ft) \tag{9}$$
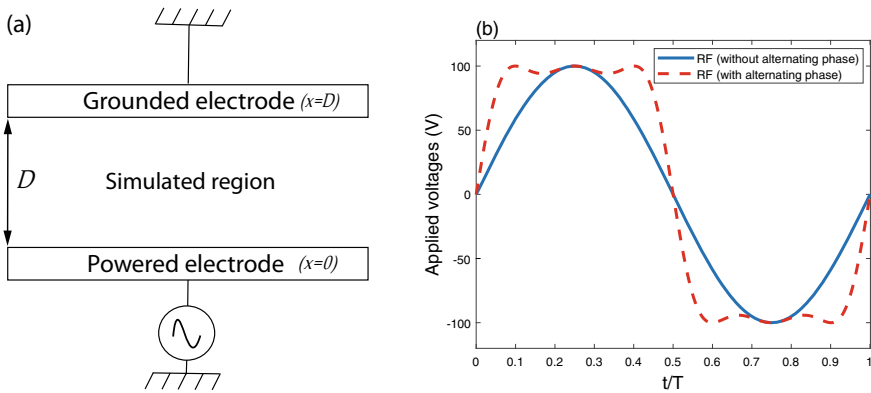
Fig. 1 a Schematic of the simulated region and **b** the applied voltages used in this study
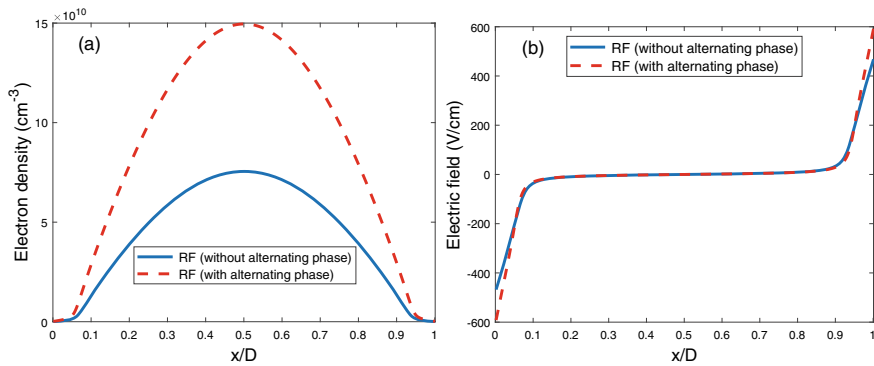
Fig. 2 Profiles of the cycle-averaged **a** electron density, **b** electric field without and with the alternating phase. At $P = 1$ Torr and $f = 13.56$ MHz

while the second excitation waveform with an alternating phase is given by [11]

$$V_a = V_{rf}\sin(2\pi ft + \alpha(t)) \tag{10}$$

where $\alpha(t) = \sin(4\pi ft)$ is the alternating phase, $f = 13.56$ (MHz) is the driving frequency and $V_{rf} = 100$ (V) is the amplitude of the applied excitation waveform.

Fig. 2a shows the cycle-averaged electron density in the space between electrodes in the two cases with and without introducing the alternating phase in the applied voltage. The maximum of the electron density in the middle of the discharge for an excitation without alternating phase is in satisfactory agreement with the experimental measurements given in Ref. [12]. In addition, the electron density in the case of including the alternating phase is obviously higher than that in the case without the alternating phase. This is because the electric field in the sheath regions near the electrodes increases when the alternating phase is considered as can be seen from

Fig. 2b. Consequently, electrons gain more energy in this case and hence, the ionization rate increases leading to high electron density. The results also show that the alternating phase has no effect on the electric field in the bulk plasma region where the electric field is almost zero due to the electrical neutrality.

The spatio-temporal profiles of the electron heating in the two cases with and without an alternating phase at pressure of 0.3 Torr and driving frequency of 3.39 MHz are illustrated in Fig. 3. It can be seen that the electron heating is mainly related to the sheath expansion and the maximums of the electron heating are located near the sheath edges. The positive values in these regions indicate that the electrons are heated by the electric field, while the negative values indicate that the electrons are cooled due to the diffusion term in Eq. (5). Therefore, electrons are heated and cooled alternatively in these regions. Moreover, the electron heating in the case with the alternating phase is higher compared to that in the case without alternating phase. This is because of the increase in the electric field and the electron density gradient when the alternating phase is added as can be seen in Fig. 2.

In order to examine the effect of pressure on the electron heating in the discharge, we present in Fig. 4, the spatio-temporal distributions of the electron heating by changing the pressure from 0.3 to 1 Torr and keeping the driving frequency fixed at 3.39 MHz for the two cases with and without alternating phase. As can be seen in the figure, the increase in the pressure leads to a decrease in the sheath length while it causes an increase in the electron heating in the two cases of excitation. Thus, when the pressure increases from 0.3 to 1 Torr, the maximum of the electron heating in the sheath regions increases from 0.0034 to 0.027 Watt.cm^{-3} in the case without the alternating phase, while in the case with the alternating phase, this maximum increases from 0.017 to 0.061 Watt.cm^{-3}.

In Fig. 5, the pressure is taken to be 1 Torr while the driving frequency is increased from 3.39 to 13.56 MHz. It is clear that when the driving frequency increases, the electron heating also increases in both cases with and without alternating phase with a decrease in the sheath length. The results in the case without alternating phase are in good agreement with the earlier simulation results shown in [13]. Furthermore, when
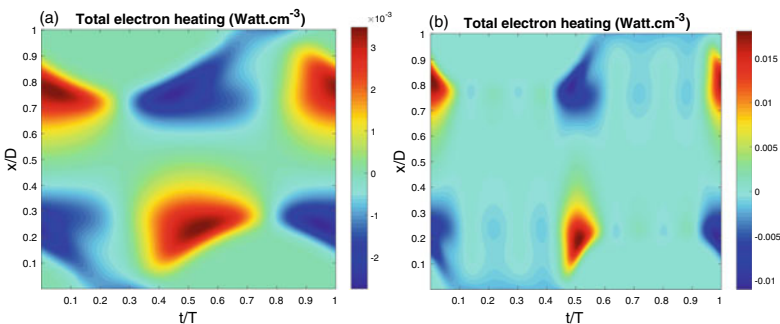
Fig. 3 Spatio-temporal profiles of the electron heating at $P = 0.3$ Torr and $f = 3.39$ MHz, **a** without alternating phase and **b** with alternating phase
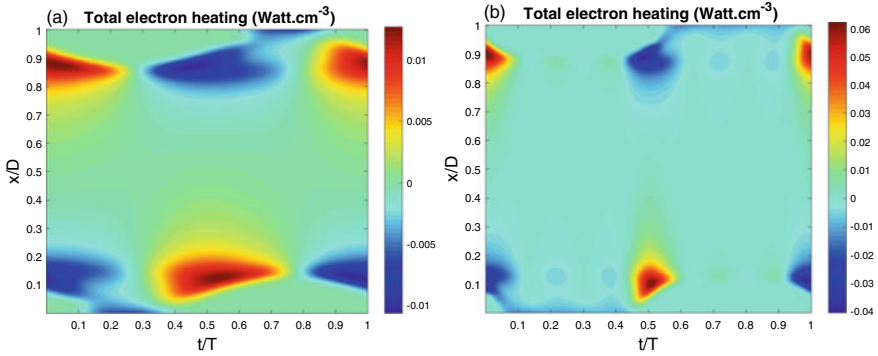
Fig. 4 Spatio-temporal profiles of the electron heating at $P = 1$ Torr and $f = 3.39$ MHz, **a** without alternating phase and **b** with alternating phase
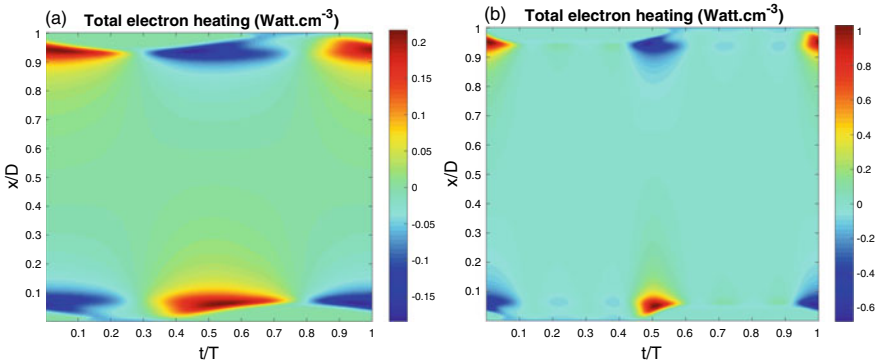
Fig. 5 Spatio-temporal profiles of the electron heating at $P = 1$ Torr and $f = 13.56$ MHz, **a** without alternating phase and **b** with alternating phase

the alternating phase is included, the small and the large oscillations of the electric field lead to an increase in the electron heating, which means that the electrons can be more heated due to rapid expansion of the sheaths which can be useful for sustaining the discharge.

4   Conclusion

In this study, a one-dimensional (1D) fluid model for an argon radio-frequency capacitive discharge at low pressure is developed to investigate the effects of an alternating phase on the electron heating mechanism. It is found that by applying an excitation waveform with an alternating phase, the electron density increases in the bulk plasma region, while the electron heating increases in the sheath regions due to the

rapid expansion of the sheaths. Moreover, as the pressure and the driving frequency increase, the electron heating also increases because of the increase in the electric field and the density gradient in the sheath regions. The results also show that electrons are alternatively heated and cooled during a radio-frequency cycle and the alternating phase can strengthen electrons to be more heated to sustain the discharge.

References

1. Brunet M, Kotb HM, Bouscayrol L, Scheid E, Andrieux M, Legros C, Schamm-Chardon S (2011) Nanocrystallized tetragonal metastable ZrO2 thin films deposited by metal-organic chemical vapor deposition for 3D capacitors. J Thin Solid Films 519:5638. https://doi.org/10.1016/j.tsf.2011.03.006
2. Makabe T, Petrovic ZL (2014) Physics of radio-frequency plasmas. 2nd (ed), CRC Press
3. Missaoui A, El kaouini M, Chatei H (2021) Numerical study of the effect of applied voltage on simultaneous modes of electron heating in RF capacitive discharges. J Lect Notes Electr Eng 681:285–291. https://doi.org/10.1007/978-981-15-6259-4_29
4. Kruger F, Wilczek S, Mussenbrock T, Schulze J (2019) Voltage waveform tailoring in radio frequency plasmas for surface charge neutralization inside etch trenches. J Plasma Sources Sci Technol 28:075017. https://doi.org/10.1088/1361-6595/ab2c72
5. Zhao L, Liu Y, Samir T (2017) Effects of gas pressure on plasma characteristics in dual frequency argon capacitive glow discharges at low pressure by a self-consistent fluid model. J Phys Plasmas 26:125201. https://doi.org/10.1088/1674-1056/26/12/125201
6. Johnson EV, Verbeke T, Vanel J, Paul J (2010) Nanocrystalline silicon film growth morphology control through RF waveform tailoring. J Phys D Appl Phys 43:412001. https://doi.org/10.1088/0022727/43/41/412001
7. Johnson EV, Delattre PA, Booth JP (2012) Microcrystalline silicon solar cells deposited using a plasma process excited by tailored voltage waveforms. J Appl Phys Lett 100:133504. https://doi.org/10.1063/1.3699222
8. Derzsi A, Lafleur T, Booth J-P, Korolov I, Donkó Z (2015) Experimental and simulation study of a capacitively coupled oxygen discharge driven by tailored voltage waveforms. J Plasma Sources Sci Technol 25:015004. https://doi.org/10.1088/0963-0252/25/1/015004
9. Missaoui A, El kaouini M, Chatei H (2021) Effect of driving frequency on dust particle dynamics in radiofrequency capacitive argon discharge. J Plasma Phys 87:905870212. https://doi.org/10.1017/S0022377821000258
10. Missaoui A, El kaouini M, Chatei H (2021) The influence of electron recombination rate on the plasma characteristics in RF argon capacitive discharge. J Mater Today Proc 45:7404–7407. https://doi.org/10.1016/j.matpr.2021.01.542
11. Bastykova NK, Donko Z, Kodanova SK, Ramazanov TS, Moldabekov ZA (2016) Manipulation of dusty plasma properties via driving voltage waveform tailoring in a capacitive radiofrequency discharge. J IEEE Trans Plasma Sci 44:545–548. https://doi.org/10.1109/TPS.2015.2503731
12. Greenberg KE, Hebner GA (1993) Electron and metastable densities in parallel-plate radio-frequency discharges. J Appl Phys 73:8126–8133. https://doi.org/10.1063/1.353451
13. Samir T, Liu Y, Zhao L-L, Zhou Y-W (2017) Effect of driving frequency on electron heating in capacitively coupled RF argon glow discharges at low pressure. J Chin Phys B 26:115201. https://doi.org/10.1088/1674-1056/26/11/115201

Computational Simulation of High-Efficiency HTL-Free Sb$_2$Se$_3$-Based Solar Cell

Abdellah Benami [iD]**, Abdelmajid El Khalfi, Youssef Achenani, lhoussayne Et-taya, and Lahoucine Elmaimouni**

Abstract Antimony selenide (Sb$_2$Se$_3$) is a new photovoltaic material that has recently gained popularity due to its direct band gap, high absorption coefficient, and low cost. So far, its efficiency has not been able to match that of a three most commercialized thin-film solar cell, such as CdTe, silicon, and CIGS. In this work, a hole transport layer-free (HTL) Sb$_2$Se$_3$-based solar cell is simulated in one dimension using the solar cell capacitance simulator (SCAPS-1D). The device key parameters with respect to the thickness, doping concentration, and defect density of absorber as well as the operating temperature were investigated. The efficiency for Sb$_2$Se$_3$ HTL free with thickness of 1600 nm and doping concentration of 10^{15} cm^{-3} at room temperature can yield a 28%. The findings of this study could pave the way for the development of a new generation of low-cost, high-efficiency thin-film solar cells.

Keywords Sb$_2$Se$_3$ solar cell · SCAPS-1D · Simulation

1 Introduction

Due to the abundance of materials on Earth, their low cost, and the simplicity of the manufacturing process, the emerging generation of thin-film solar cells (TFSCs) is undergoing extensive research. Sb$_2$Se$_3$ is one of the most promising absorber materials for efficient TFSCs due to its excellent optoelectronic properties such as a tunable band gap of 1.1–1.3 eV, low toxicity, high optical absorption coefficient ($>10^5$ cm^{-1} in the visible region) and excellent electron and hole mobilities of 16.9 cm^2 V^{-1} S^{-1} and

A. Benami (✉) · A. El Khalfi · Y. Achenani · Et-taya
LM3ER-OTEA, Department of Physics, Faculty of Sciences and Techniques, Moulay Ismail University of Meknes, Boutalamine, B.P. 509, 52000 Errachidia, Morocco
e-mail: a.benami@fste.umi.ac.ma

Et-taya
e-mail: l.ettaya@edu.umi.ac.ma

L. Elmaimouni
ERMAM, Faculté Polydisciplinaire d'Ouarzazate, Université Ibn Zohr, 45000 Ouarzazate, Morocco

42 cm^2 V^{-1} S^{-1}, respectively [1–3]. Moreover, Sb$_2$Se$_3$ is physically, thermodynamically, and chemically stable with a unique one-dimensional (1D) crystal structure and highly anisotropic properties [1].

The current best power conversion efficiency (PCE) of Sb$_2$Se$_3$ TFSC is around 10.7%, which is comparable to that of conventional TFSCs [4, 5]. Despite the rapid improvements in Sb$_2$Se$_3$ efficiency over the last five years, it remains far behind perovskite (PCE = 25.5%), CIGS (PCE = 23.4%), CZTSSe (PCE = 13%) [6], and Shockley–Queisser limit efficiency of above 30% [7]. The limit PCE of Sb$_2$Se$_3$ is caused by large V_{oc} deficit which can be attributed to interface recombination and defect state in the bulk [8]. So, the next objectives include increasing the PCE of the Sb$_2$Se$_3$ TFSC and digging deeper into its features.

Sb$_2$Se$_3$ TFSC can be produced in three ways: high vacuum, low vacuum, and non-vacuum. Messina et al. prepared Sb$_2$Se$_3$ film by chemical bath deposition and obtained a PCE of 0.66% [9]. In 2014, Liu et al. used high vacuum thermal evaporation to create Sb$_2$Se$_3$ device with a PCE of 2.1% [10]. Close-space sublimation technique was used by Li et al. in 2019 to deposit Sb$_2$Se$_3$ solar cells with a PCE of 9.2% [3]. In 2022, Zhao et al. developed a solution post-treatment technique for fabricating TFSCs using alkali metal fluorides and obtained a PCE of 10.7% [5].

In this work, we used SCAPS program to simulate the performance of FTO/ETM/absorber layer/back contact device configuration. The thickness, doping concentration, and total defect density of the absorber, as well as the operating temperature, were evaluated as device critical parameters.

2 Device Structure and Simulation Methodology

2.1 Device Structure

As shown in Fig. 1a, the proposed *p–n* heterojunction solar cell structure consists of glass/FTO/ETM/absorber layer/back contact. In this structure, FTO (SnO$_2$:F) serves as a front contact, with a layer of *n*-type doped CdS serving as an electron transport material (ETM). *P*-type Sb$_2$Se$_3$ serves as the absorber layer, with a flat band serving as the back contact. The energy band diagram of the proposed structure is shown in Fig. 1b.

2.2 Simulation Methodology

The one-dimensional solar capacitance simulator (SCAPS-1D) version 3.3.10 was used for numerical simulation. It was created by Burgelman and co-workers at the Ghent University, Belgium. It is open to the scientific community and is free of charge. Since its initial release, it has been improved. Different types of solar

Fig. 1 **a** Structure and **b** band diagram of the proposed CdS/Sb$_2$Se$_3$ heterojunction solar cell
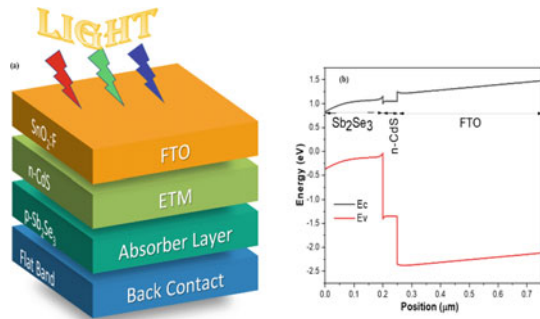
cells can be simulated using SCAPS-1D such as perovskite [11], CIGS [12], CZTS [13, 14], CZTSSe [7], and Sb$_2$Se$_3$ [3, 15]. Furthermore, the results obtained using the SCAPS simulator are consistent with previous experimental results reported by other researchers. The software is based on Poisson's equation, hole continuity, and electron continuity [7].

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{q}{\varepsilon}\left[p(x) - n(x) + N_D - N_A + \rho_p - \rho_n \right] = 0 \tag{1}$$

$$\frac{1}{q}\frac{dJ_p}{dx} = G_{op}(x) - R(x) \tag{2}$$

$$\frac{1}{q}\frac{dJ_n}{dx} = -G_{op}(x) + R(x) \tag{3}$$

where Ψ is the electrostatic potential, ε is the dielectric constant, q is the electron charge, N_A and N_D are densities of acceptor-like and donor-like, p, n, ρ_p, ρ_n, J_p, and J_n are hole concentration, electron concentration, hole distribution, electron distribution, current densities of holes and electrons, respectively. Gop is the optical generation rate, and R is the net recombination from direct and indirect recombination.

The parameters used in SCAPS-1D are listed in Table 1. Flat band model is used for the FTO/semiconductor and metal interfaces/semiconductor. The operating temperature is set at 300 K, and the illumination is an AM 1.5 spectrum with a light power of 1000 W/m^2. The optical absorption of all layers is set from a model and is given by Eq. (4):

$$\alpha(hv) = \left(\alpha_0 + \beta_0 \frac{E_g}{hv} \right)\sqrt{\frac{hv}{E_g} - 1} \quad \text{and} \quad \alpha = 0 \quad \text{if } hv < E_g \tag{4}$$

where $\alpha_0 = 10^5$ cm^{-1} and $\beta_0 = 10^{-12}$ cm^{-1}.

Table 1 Parameters used in the simulation [3, 7, 11]

Material properties	FTO	n-CdS	p-Sb$_2$Se$_3$
Thickness (nm)	500	50	200[a]
Bandgap (eV)	3.6	2.4	1.2
Electron affinity (eV)	4	4.2	4.04
Dielectric permittivity (eV)	9	10	18
Density of states in CB (cm^{-3})	2.2×10^{18}	2.2×10^{18}	2.2×10^{18}
Density of states in VB (cm^{-3})	1.8×10^{19}	1.8×10^{19}	1.8×10^{19}
Thermal velocity of electron (cm/s)	10^7	10^7	10^7
Thermal velocity of hole (cm/s)	10^7	10^7	10^7
Electron mobility (cm^2/Vs)	10^2	10^2	15
Hole mobility (cm^2/Vs)	25	25	5.1
Donor density (cm^{-3})	10^{17}	10^{18}	0
Acceptor density (cm^{-3})	0	0	10^{15a}

[a] Variable parameter
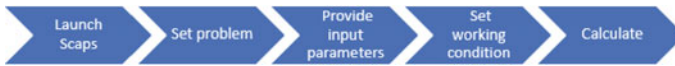
Fig. 2 SCAPS simulation process

Figure 2 explains the SCAPS working procedure (Run SCAPS first, then select Set Problem; after that, we provide the input parameters, working conditions, and then begin the calculation).

3   Results and Discussion

3.1   *Effect of Absorber Layer Thickness and Doping Concentration*

The absorber thickness and doping concentration are two of the most important parameters that influence the TFSC's performance. As a result, the thickness and doping of Sb$_2$Se$_3$ were changed from 200 to 1600 nm, and from 10^{13} cm^{-3} to 10^{19} cm^{-3}, respectively, to see how they affected the output parameters. It is clear from Fig. 3a–d, for doping of 10^{15} cm^{-3}, that V_{oc}, J_{sc}, and PCE rise as absorber thickness increases, whereas FF initially decreases and then increases. When the absorber thickness is 200 nm, the PCE is around 13.44%, and when the thickness increased to 1600 nm, the PCE rises to 21.5%. This can be explained by increasing the absorber thickness, which causes more photons to be absorbed [16].
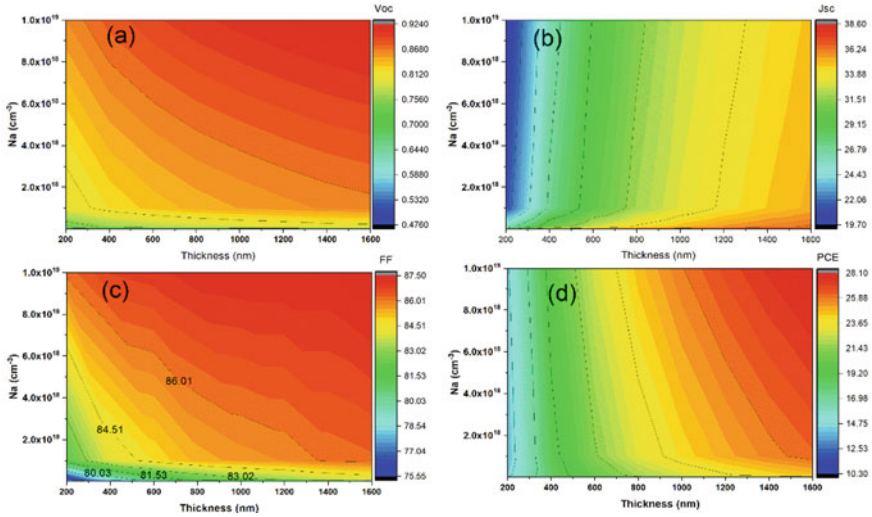
Fig. 3 Contour plot of output performance **a** V_{oc}, **b** J_{sc}, **c** FF, and **d** PCE as a function of thickness and doping concentration

Doping concentration has the same effect on PCE, V_{oc}, and FF, which all rise as doping concentration rises. However, J_{sc} remained constant than drop when the doping is greater than 10^{15} cm^{-3}. The PCE increases from 10.34 to 28.08% when the thickness and doping concentration of the absorber layer rise from 200 to 1600 nm and 10^{13} cm^{-3} to 10^{19} cm^{-3}, respectively.

3.2 Influence of Defect Density of the Absorber Layer

The optoelectronic properties of thin-film devices are highly dependent on the material quality. In the absorber layer, defects are unavoidable. They can be found on the surface as well as in bulk. We introduce neutral defect state density in Sb$_2$Se$_3$ to quantitatively investigate the influence of absorber layer defect states on photovoltaic device performance. The effect of defect density in the range of 10^{12} to 10^{17} cm^{-3} on cell performance was investigated, in this study, while all other layers were kept constant at the values specified in Table 1, with the active layer thickness and concentration N_a at 1600 nm and 10^{15} cm^{-3}, respectively. Figure 4 depicts the influence of defect density variation on the device's photovoltaic properties. The device's key parameters are almost constant up to a density of 10^{15} cm^{-3}, but they deteriorate beyond that point. The PCE decreases from 21.5% (for defect density of 10^{12} cm^{-3}) to 9.71% (for defect density of 10^{17} cm^{-3}). As the defect density rises, the carrier recombination rate rises as well. In other words, higher defect density leads to increased carrier trapping and a deterioration in carrier transmission. The defect
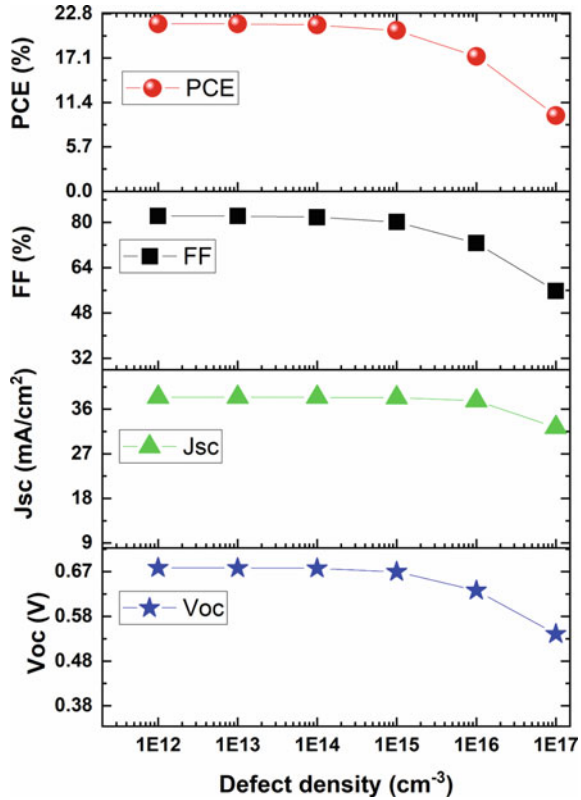
density should not exceed 10^{15} cm^{-3}, according to this study, in order to improve
the output of the proposed device.

3.3 Effect of Working Temperature

Another important factor influencing device performance is the working temperature.
It has been reported that the photovoltaic panels are rated to operate in a tempera-
ture range ranging from −45 to 85 °C. In this section, we look at how changing
the temperature from 240 to 360 K affects the results. We keep the thickness and
doping concentration at 800 nm and 10^{15} cm^{-3}, respectively, to minimize matter
consumption. As shown in Fig. 5, as the temperature rises, V_{oc} decreases. As the V_{oc}
decreases, the PCE decreases as well, as shown in the inset of Fig. 5. The decrease in
PCE as a function of working temperature is due to an increase in series resistance
and a reduction in the energy band gap [7].

Fig. 5 *J–V* characteristics and PCE (inset shows the decreasing) of Sb$_2$Se$_3$ as a function of temperature
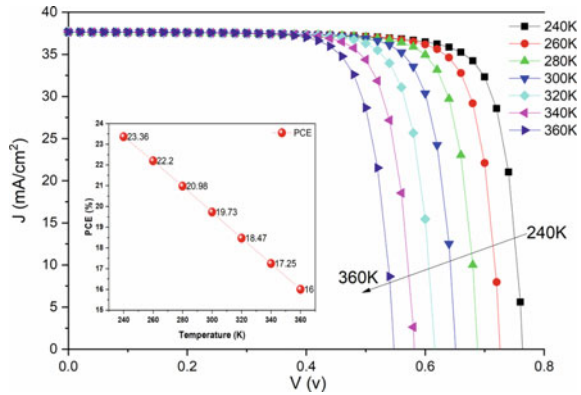
4 Conclusion

The third-generation Sb$_2$Se$_3$ solar cell has received a lot of attention as a potential replacement for thin-film technology based on CIGS and CdTe materials. Despite this, solar cells made from these materials operate poorly, with low reported conversion efficiencies due to a number of limiting constraints. As a result, adequate solutions for understanding and resolving these difficulties are required. In this context, we proposed using SCAPS-1D software to simulate the heterojunction FTO/*n*-CdS/*p*-Sb$_2$Se$_3$ device in order to improve the efficiency. Many parameters that have a substantial impact on the performance of the TFSC are explored. The thickness, doping concentration, defect density of the absorber layer, and temperature all have an impact on output parameters such as PCE, J_{sc}, FF, and V_{oc}. The maximum PCE of HTL-free Sb$_2$Se$_3$-based TFSC was around 28%. Our findings can be applied to the production of high Sb$_2$Se$_3$-based TFSC.

References

1. Cheng Y, Xin C, Zhao J, Wang J, Gong M, Miao H, Hu X (2022) Constructing lateral sulfur-gradient Sb$_2$(S$_x$Se$_{1-x}$)$_3$ heterostructures for Sb$_2$Se$_3$ nanorod photocathodes with enhanced photoelectrochemical properties. Electrochim Acta 403:139610
2. Lai Y, Chen Z, Han C, Jiang L, Liu F, Li J, Liu Y (2012) Preparation and characterization of Sb$_2$Se$_3$ thin films by electrodeposition and annealing treatment. Appl Surf Sci 261:510–514
3. Li Z-Q, Ni M, Feng X-D (2020) Simulation of the Sb$_2$Se$_3$ solar cell with a hole transport layer. Mater Res Exp 7:016416
4. Tang R, Wang X, Lian W, Huang J, Wei Q, Huang M, Yin Y, Jiang C, Yang S, Xing G, Chen S, Zhu C, Hao X, Green MA, Chen T (2020) Hydrothermal deposition of antimony selenosulfide thin films enables solar cells with 10% efficiency. Nat Energy 5:587–595
5. Zhao Y, Wang S, Jiang C, Li C, Xiao P, Tang R, Gong J, Chen G, Chen T, Li J, Xiao X (2022) Regulating energy band alignment via alkaline metal fluoride assisted solution post-treatment enabling Sb$_2$(S,Se)$_3$ solar cells with 10.7% efficiency. Adv Energy Mater 12:2103015

6. NREL (2020) Best research-cell efficiencies. https://www.nrel.gov/pv/cell-efficiency.html. Accessed: Oct 2020
7. Et-taya L, Ouslimane T, Benami A (2020) Numerical analysis of earth-abundant $Cu_2ZnSn(S_xSe_{1-x})_4$ solar cells based on spectroscopic ellipsometry results by using SCAPS-1D. Sol Energy 201:827–835
8. Liang G, Chen X, Ren D, Jiang X, Tang R, Zheng Z, Su Z, Fan P, Zhang X, Zhang Y, Chen S (2021) Ion doping simultaneously increased the carrier density and modified the conduction type of Sb_2Se_3 thin films towards quasi-homojunction solar cell. J Mater 7:1324–1334
9. Messina S, Nair MTS, Nair PK (2009) Antimony selenide absorber thin films in all-chemically deposited solar cells. J Electrochem Soc 156:H327
10. Liu X, Chen J, Luo M, Leng M, Xia Z, Zhou Y, Qin S, Xue D-J, Lv L, Huang H, Niu D, Tang J (2014) Thermal evaporation and characterization of Sb_2Se_3 thin film for substrate Sb_2Se_3/CdS solar cells. ACS Appl Mater Interfaces 6:10687–10695
11. Ouslimane T, Et-taya L, Elmaimouni L, Benami A (2021) Impact of absorber layer thickness, defect density, and operating temperature on the performance of $MAPbI_3$ solar cells based on ZnO electron transporting material. Heliyon 7:e06379
12. Daoudia A, El Hassouani Y, Benami A (2016) Investigation of the effect of thickness, band gap and temperature on the efficiency of CIGS solar cells through SCAPS-1D. Int J Eng Tech Res (IJETR) 6:71–75
13. Benami A (2019) Effect of CZTS parameters on photovoltaic solar cell from numerical simulation. J Energy Power Eng 13:32–36
14. Haddout A, Fahoume M, Qachaou A, Raidou A, Lharch M, Elharfaoui N (2019) Influence of composition ratio on the performances of kesterite solar cell with double CZTS layers—a numerical approach. Sol Energy 189:491–502
15. Basak A, Singh UP (2021) Numerical modelling and analysis of earth abundant Sb_2S_3 and Sb_2Se_3 based solar cells using SCAPS-1D. Sol Energy Mater Sol Cells 230:111184
16. Mamta, Maurya KK, Singh VN (2021) Efficient Sb_2Se_3 solar cell with a higher fill factor: a theoretical approach based on thickness and temperature. Sol Energy 230, 803–809

Optimization of Cd-Free CZTSSe Kesterite Device with Different BSF Layers by SCAPS-1D

Lhoussayne Et-taya and **Abdellah Benami**

Abstract In recent years, the low-cost and non-toxic element-based solar cells have attracted the researcher community since the last decade due to their clean and green status and their performance in several domains on everything in photovoltaic technology. Copper zinc tin sulfur-selenium (CZTSSe) is a good candidate for these elements. In this study, we have used the SCAPS-1D simulator (Solar Cell capacitance Simulator) to investigate the performance of Cd-free CZTSSe kesterite solar cells with the TiO_2 as an alternative buffer layer. The effect of the window layer's thickness and shunt resistance of device has been examined. Also, the current voltage ($J–V$) and quantum efficiency (QE) characteristics of different back surface field materials have been analyzed and improved. The optimized CZTSSe structure achieves a maximum power conversion efficiency of 30.52%, 31.21%, and 31.29% with CZGSe, $CuSbS_2$, and Sb_2S_3 as BSF layers, respectively.

Keywords Solar cells · Cd-free buffer layer · BSF · CZTSSe · SCAPS-1D

1 Introduction

The third generation of solar cells comprises low-cost, non-toxic and abundant elements, in particular the CZTSSe kesterite material captivated many scientists to use it as absorber thin-film solar cells (TFSCs). This is due to the excellent optoelectrical properties which are manifested in the tuning of direct energy bandgap, intrinsic p-type conductivity and high optical absorption coefficient ($>10^4$ cm^{-1}), which can absorb ~ 90% of the incident photon in the range of solar spectrum [1]. Bandgap and electron affinity of $CZTS_xSe_{1−x}$ material are adjusted with the change of composition ratio of sulfur where $x = S/(S + Se)$ and the composition of x is between 0

L. Et-taya (✉) · A. Benami
LM3ER-OTEA, Department of Physics, Faculty of Sciences and Techniques, Moulay Ismail University of Meknes, BP 509 Boutalamine, 52000 Errachidia, Morocco
e-mail: l.ettaya@edu.umi.ac.ma

A. Benami
e-mail: a.benami@fste.umi.ac.ma

and 1, resulting in a change of bandgap (0.95–1.5 eV) and electron affinity (4.35–4.5 eV) [2], which is close to the ideal for a single-junction solar cell which allows absorption of the solar spectrum by reaching the Shockley-Queisser limit at 32% of efficiency. Another reason for showing high interest in CZTSSe is its long-life stability [1, 3]. All the above properties are sufficient to choose the CZTSSe among the main candidates for thin-film solar cell devices and a good replacement for the chalcopyrite conventional such as CdTe, CIGS and GaAs which are composed by toxic elements like Cd, As and rare earth element hindering their properties for the commercial photovoltaic technologies [4, 5].

Due to its low efficiency of 12.62% [6] and the toxicity of cadmium (Cd) in the heterojunction CdS/CZTSSe, it allows us to modify and improve the structure of CZTSSe solar cell.

Many studies used the back surface field between the Mo and the absorber layer to improve the overall performance of the CZTSSe kesterite. When compared to a conventional device, the device with the back surface field layer produces a significantly higher output voltage [7]. The addition of BSF lowers the barrier height of the rear contact and reduces the loss of recombination of minority carriers at the rear contact [8]. As a result, the number of photons absorbed in the CZTSSe layer increases, as does the cell's efficiency [9, 10].

SCAPS-1D simulator was used to perform numerical simulation of the proposed device structure in order to optimize output performance in this contribution. The impact of the window layer and shunt resistance on the performance of Mo/p-CZTSSe/n-TiO$_2$/i-ZnO/ZnO:Al device structure with and without BSF layer is demonstrated in this study.

2   Methodology and Cell Structure

SCAPS-1D is one of the best and popular software used for the simulation of solar cell devices and fitting the simulation results with the experimental data. SCAPS program has been developed by the Department of Electronics and Information Systems of the University of Gent, Belgium [11].

In the first step, we have simulated the structure Mo/p-CZTSSe/n-TiO$_2$/i-ZnO/ZnO:Al as shown in Fig. 1a, with CZTSSe as an active layer, ZnO: Al (aluminum doped ZnO: AZnO) as an optical window which must combine two essential properties, namely electrical conductivity and optical transparency. The ZnO layer serves as a window layer, and the n-type TiO$_2$ layer acts as a buffer layer, creating a p-n heterojunction with the p-type absorbent layer CZTSSe. Molybdenum (Mo), its main role is to collect the charges generated in the solar cell.

Secondly, we optimized our structure by inserting the BSF layer between the CZTSSe absorber layer and Mo back contact using different layers: CuSbS$_2$, Sb$_2$S$_3$ and CZGSe materials as BSF as shown in Fig. 1b.

We employed the AM 1.5G spectrum with a 1000 W/m^2 incident power at 300 K for the simulation. It should be mentioned that the absorption coefficient for each
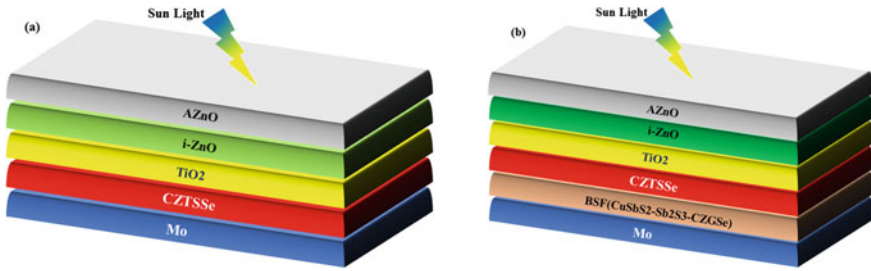
Fig. 1 **a** Conventional and **b** optimized solar cell structure

layer is shown in Fig. 2 and was determined using the following formula:

$$\alpha = \frac{4\pi k}{\lambda} \qquad (1)$$

where k and λ are extinction coefficient and wavelength, respectively.

The physical parameters used for the numerical simulation are summarized in Table 1.

Fig. 2 Absorption coefficient of different layers
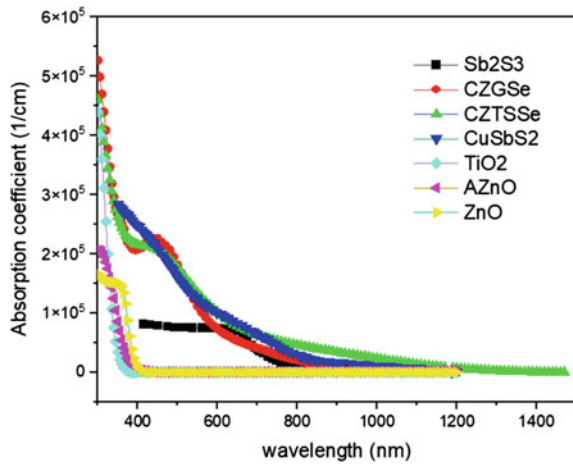
Table 1 Physical parameters of different layers used in the simulation [2, 4, 12–15]

Material properties	ZnO:Al	i-Zno	n-TiO$_2$	p-CZTSSe	p-Sb$_2$S$_3$	p-CuSbS$_2$	p-CZGSe
d (nm)	200[a]	50	50	2000	100	100	100
E_g (eV)	3.3	3.3	3.2	1.146	1.62	1.45	1.4
χ (eV)	4.4	4.4	4.26	4.41	3.7	4.05	3.54
ε (eV)	9	9	9	13.6	7.08	12	13.6
N_c (cm^{-3})	2.2×10^{18}	2.2×10^{18}	2×10^{18}	2.2×10^{18}	2×10^{19}	1.23×10^{20}	2.2×10^{18}
N_v (cm^{-3})	1.8×10^{19}	1.8×10^{19}	1.8×10^{19}	1.8×10^{19}	1×10^{19}	1.78×10^{20}	1.8×10^{19}
V_{the-} (cm/s)	10^7	10^7	10^7	107	10^7	6.8×10^8	10^7
V_{thp} (cm/s)	10^7	10^7	10^7	10^7	10^7	6.07×10^8	10^7
μn (cm^2/Vs)	10^2	10^2	20	10^2	9.8	4	10^2
μh (cm^2/Vs)	25	25	10	25	10	4	25
N_D (cm^{-3})	10^{20}	10^{19}	10^{18}	0	0	0	0
N_A (cm^{-3})	0	10^{19}	0	10^{18}	5×10^{18}	5×10^{18}	5×10^{18}
α (cm^{-1})	File	File	File	File	File	File	File

d Thickness, E_g bandgap, χ electron affinity, ε dielectric permittivity, N_c/N_v density of states in CB/VB, V_{the-}/V_{thp} thermal velocity of electron/hole, μ_n/μ_h electron/hole mobility, N_D/N_A donor/acceptor density and α absorption coefficient
[a] is a variable parameter

3 Results and Discussions

3.1 *J–V Characteristics and Quantum Efficiency of TFSC with and Without BSF Material*

Figure 3 depicts the *J–V* characteristics and external quantum efficiency of the CZTSSe structure solar cell without and with CuSbS$_2$, Sb$_2$S$_3$ and CZGSe BSF layers. The introduction of the BSF layer improved the overall performance of the photovoltaic solar cell when compared to the conventional structure, as shown in Fig. 3a. As a result, the short current density (J_{sc}) circuit and open circuit voltage (V_{oc}) increase from 40 to 46 mA/cm^2 and from 0.71 to 0.87 V, respectively, with and without BSF layer. As a result, BSF allows the designed structure to be optimized by reducing carrier recombination at the device's back surface. In Fig. 3b, due to the increased photon collection at longer wavelengths, the quantum efficiency is greater than 90% in the 500–1000 nm range after adding the BSF layer [16].
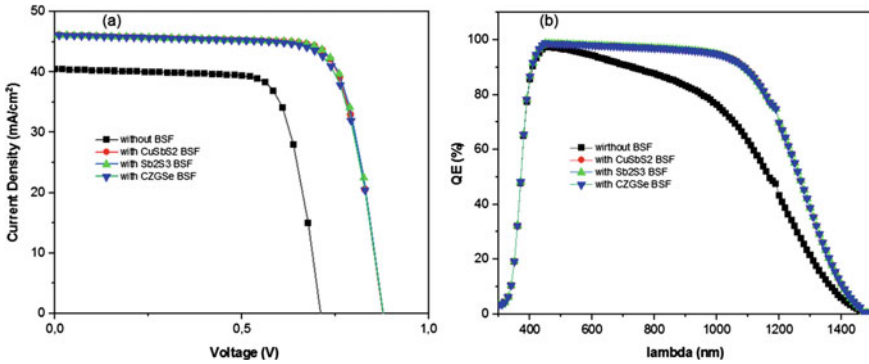
Fig. 3 **a** *J*–*V* characteristics curve and **b** quantum efficiency of solar cell wit and without BSF

3.2 Effect of Window Layer on Solar Cell Device

The thickness of the AZnO window layer is varied from 50 to 500 nm, while the other parameters remain constant as shown in Table 1, and the effect on photovoltaic parameters is studied as shown in Fig. 4. J_{sc} and PCE decrease as AZnO thickness increases, but FF increases and Voc remains nearly constant. It is due to that AZnO has high transparency and a low resistivity as an electrode, which allows it to lower series resistance and improve fill factor (FF) therefore for enhancement of overall performance because the window layer serves two functions: optical transmission and electrical conduction, allowing light to pass through to the absorber and collect electrons.

3.3 Effect of Shunt Resistance on Solar Cell Device

Figure 5 depicts a study of the effect of the shunt resistance on the photovoltaic performances of CZTSSe solar cell. The shunt resistance (R_{sh}) is varied from 100 to 1000 Ω cm^2, while the series resistance is held constant at 1.5 Ω cm^2 and the other parameters at the values specified in Table 1. As shown in Fig. 5, all key parameters of the proposed solar cell increase as the value of Rsh is increases. At R_{sh} = 1000 Ω cm^2, the higher efficiency in the conventional CZTSSe-based solar cell and optimized with CuSbS$_2$, Sb$_2$S$_3$ and CZGSe BSF layers is 21.85% and (31.55%, 31.67% and 30.86%), respectively. The results of the output parameters of solar cells explored in the literature as a function of Rsh are in good agreement with the simulation results [9].
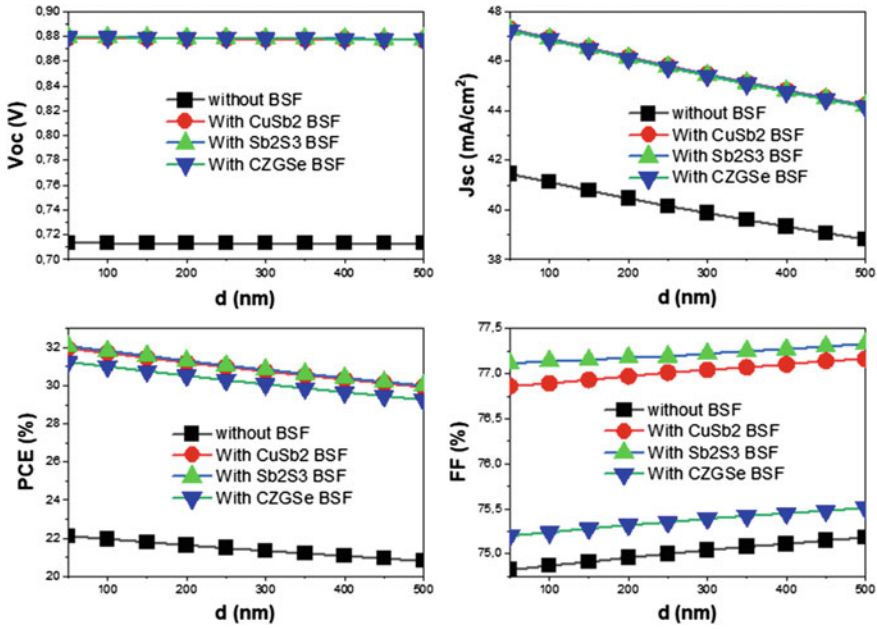
Fig. 4 Photovoltaic parameters of solar cell by varying the thickness of window layer
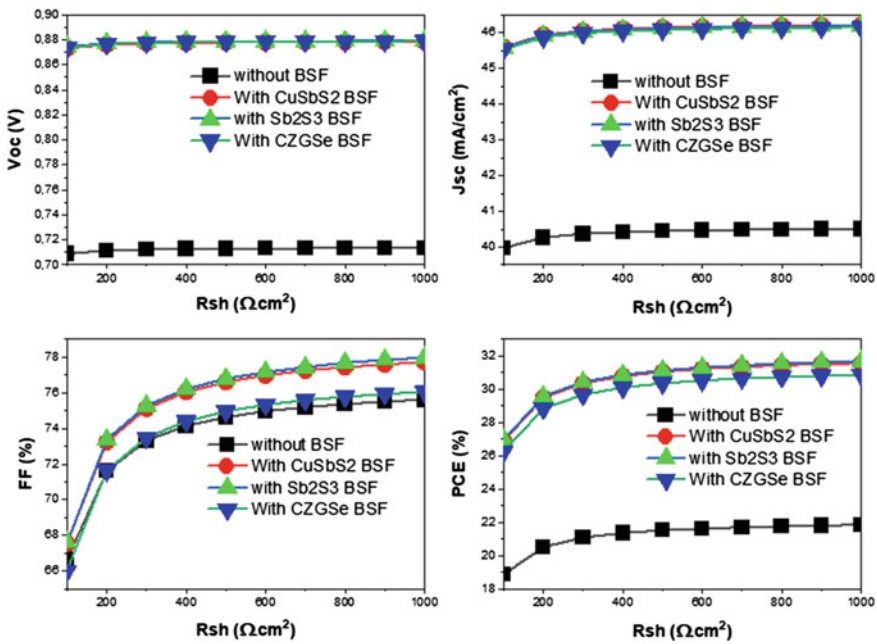
Fig. 5 Photovoltaic parameters of solar cell and with BSF layer at different resistance shunt

4 Conclusion

In the simulation contribution provided, we increased the performance of the CZTSSe kesterite by employing the CuSbS$_2$, Sb$_2$S$_3$ and CZGSe materials as a BSF layer. In this study, we utilized the SCAPS-1D tool to evaluate the performance of standard CZTSSe solar cells and compare them to novel ones changed by the BSF. The effects of modifications in window layer thickness and shunt resistance on the output parameter photovoltaics of the device solar cells are investigated. CZTSSe attained a maximum efficiency of 22.12% without BSF and then increased to 31.95%, 32.07% and 31.24% with CuSbS$_2$, Sb$_2$S$_3$ and CZGSe as BSF layers, respectively. The aforesaid results will be utilized to guide the production of kesterite CZTSSe high-efficiency solar cell-based thin films.

References

1. Rahman MS et al (2021) Bilayer CZTS/Si absorber for obtaining highly efficient CZTS solar cell. Sol Energy 230:1189–1198
2. Et-taya L, Ouslimane T, Benami A (2020) Numerical analysis of earth-abundant Cu$_2$ZnSn(S$_x$Se$_{1-x}$)$_4$ solar cells based on spectroscopic ellipsometry results by using SCAPS-1D. Sol Energy 201:827–835
3. AlZoubi T, Moustafa M (2019) Numerical optimization of absorber and CdS buffer layers in CIGS solar cells using SCAPS. Int J Smart Grid Clean Energy 8:291–298
4. Sadanand et al (2021) Comparative study of the CZTS, CuSbS$_2$ and CuSbSe$_2$ solar photovoltaic cell with an earth-abundant non-toxic buffer layer. Solar Energy 222:175–185
5. Benami A (2019) Effect of CZTS parameters on photovoltaic solar cell from numerical simulation. J Energy Power Eng 13
6. Son D-H et al (2019) Effect of solid-H2S gas reactions on CZTSSe thin film growth and photovoltaic properties of a 12.62% efficiency device. J Mater Chem A 7(44):25279–25289
7. Khattak YH et al (2018) Effect of CZTSe BSF and minority carrier life time on the efficiency enhancement of CZTS kesterite solar cell. Curr Appl Phys 18(6):633–641
8. Omrani MK et al (2018) Improve the performance of CZTSSe solar cells by applying a SnS BSF layer. Solid-State Electron 141:50–57
9. Biplab SRI et al (2020) Performance enhancement of CIGS-based solar cells by incorporating an ultrathin BaSi2 BSF layer. J Comput Electron 19(1):342–352
10. Enayati Maklavani S, Mohammadnejad S (2020) Reduction of interface recombination current for higher performance of p+-CZTS$_x$Se$_{(1-x)}$/p-CZTS/n-CdS thin-film solar cells using Kesterite intermediate layers. Solar Energy 204:489–500
11. Burgelman M et al (2013) Advanced electrical simulation of thin film solar cells. Thin Solid Films 535:296–301
12. Et-taya L, Benami A, Ouslimane T (2022) Study of CZTSSe-based solar cells with different ETMs by SCAPS. Sustainability 14(3):1916
13. Salem MS et al (2022) Numerical analysis and design of high performance HTL-free antimony sulfide solar cells by SCAPS-1D. Opt Mater 123:111880
14. Xiao Y, Wang H, Kuang H (2020) Numerical simulation and performance optimization of Sb2S3 solar cell with a hole transport layer. Opt Mater 108:110414
15. Yang S et al (2020) Numerical modelling of the performance-limiting factors in CZGSe solar cells. J Phys D Appl Phys 53
16. Barman B, Kalita PK (2021) Influence of back surface field layer on enhancing the efficiency of CIGS solar cell. Sol Energy 216:329–337

Analysis of 2D Simulation of Hydrogenated Silicon Nitride Plasma Discharge in CCP Reactor for Thin Film Solar Cell Deposition

Meryem Grari, Yassmina Guetbach, Sara Said, CifAllah Zoheir, and Abdenacer Essalhi

Abstract In this work, we are interested in the one-dimensional (1D) and two-dimensional (2D) numerical simulation of the deposition of thin films of hydrogenated silicon nitride in a capacitive coupled plasma reactor (CCP) at low pressure of 0.3 Torr and low temperature of about 573 K. The radiofrequency plasma discharge, driven by a sinusoidal voltage of frequency 13.56 MHz, is modeled by a macroscopic–microscopic approach such that the densities and energy are calculated by the fluid approach where the source terms are calculated by the microscopic approach. The fluid equations are solved by the numerical finite element method. The results obtained show the evolution of fundamental characteristics of the plasma discharge. By comparing the 1D and 2D evolution of these characteristics, we have concluded that the contribution of the 2D simulation is very important to be neglected. Thus, the consideration of 2D simulation provides a better description and understanding of the characteristics of a radiofrequency plasma discharge in order to properly develop a deposit of thin films of hydrogenated silicon nitride more efficient and less expensive.

Keywords Plasma discharge · Hydrogenated silicon nitride · Macroscopic-microscopic approach · Thin films deposition

1 Introduction

Plasma-enhanced chemical vapor deposition (PECVD) of silicon-based thin films is the main commercial use of films for microelectronic and photovoltaic applications [1, 2]. Hydrogenated silicon nitride films, which are formed from the decomposition of radiofrequency (RF) plasma discharge containing silicon, ammonia and hydrogen gases, represent the largest use of plasma deposition surface treatment in the semiconductor industry [3, 4].

M. Grari (✉) · Y. Guetbach · S. Said · C. Zoheir · A. Essalhi
LETSER Laboratory, Department of Physics, Mohamed First University, Oujda, Morocco
e-mail: grarimery@gmail.com; m1.grari@ump.ac.ma

The modeling of a radiofrequency plasma discharge combines three major phenomena: fluid flow, chemical reactions and plasma dynamics. The most difficult part is the analysis of plasma interactions where charged particles are both affected by external electric and magnetic fields [5]. Indeed, inter-particle collisions occur on time and space scales which are generally much shorter than those of the applied fields or the fields obtained from the average motion of the particles [6]. The resulting coherent system is nonlinear and very difficult to analyze, which has prompted researchers to consider various approximations in numerical simulation in order to reduce the complexity of the problem.

Several methods have been used for the simulation of plasma discharges. Among them are the kinetic [7, 8], fluid [9, 10] and hybrid [4, 6] models. By using the hybrid (macroscopic–microscopic) model, we aim to preserve the accuracy of kinetic simulations and at the same time reduce the computational load. In this context of optimization, several researchers have considered one-dimensional (1D) simulation [10, 11] and less expensive calculation methods as a key to simplification. However, although the calculations resulting from the 1D simulation are simple and less expensive, it is possible that very important information is not revealed because of this dimension restriction on a single axis.

In this work, we are interested in the study of the 1D and 2D spatio-temporal variation of the fundamental characteristics of RF plasma of hydrogenated silicon nitride where the emphasis is on the importance of the contribution of 2D simulation, where we use the finite element method for the spatial discretization. Thus, we seek to validate the numerical model used by comparing the obtained results with those of the literature. Then, we investigate the use of 2D versus 1D simulation in order to determine the most adequate simulation for a good description and understanding of the evolution of the characteristics of a radiofrequency plasma discharge.

2 Simulation Model

The use of plasma deposition for microelectronics seemed doubtful because the uniformity of film thickness is very low [12]. 2D simulations are particularly useful since they can address the issue of plasma uniformity and the spatiotemporal dynamics of plasma along the radial and axial direction.

2.1 Plasma Model

The fluid model is presented by the set of Eqs. (1)–(6) combined by the Poisson equation (7), assuming that the density and energy vary in the radial r and axial z directions, taking into account the hypotheses presented in the works [13–15].

- Electron and ion transport

$$\frac{\partial n_e}{\partial t} + \nabla \boldsymbol{\Gamma}_e = S_e \tag{1}$$

$$\frac{\partial n_i}{\partial t} + \nabla \boldsymbol{\Gamma}_i = S_i \tag{2}$$

- Electron and ion flux

$$\boldsymbol{\Gamma}_e = -n_e \mu_e \boldsymbol{E} - \nabla(n_e D_e) \tag{3}$$

$$\boldsymbol{\Gamma}_i = n_i \mu_i \boldsymbol{E} - \nabla(n_i D_i) \tag{4}$$

- Electron energy

$$\frac{\partial n_\varepsilon}{\partial t} + \nabla \boldsymbol{\Gamma}_\varepsilon + E \boldsymbol{\Gamma}_e = S_\varepsilon \tag{5}$$

$$\boldsymbol{\Gamma}_\varepsilon = -n_\varepsilon \mu_\varepsilon \boldsymbol{E} - \nabla(n_\varepsilon D_\varepsilon) \tag{6}$$

- Electric field

$$\varepsilon_0 \nabla \boldsymbol{E} = e(n_i - n_e) \tag{7}$$

where $\mu_e = e/m_e v_{eN}$, $\mu_i = Z_i e/m_i v_{iN}$ are the electron and ion mobility $D_e = k_B T_e/m_e v_{eN}$ and $D_i = k_B T_i/m_i v_{iN}$ are the electron and ion diffusivity, $n_\varepsilon = n_e \varepsilon$ and $D_\varepsilon = n_e D_e$ are the electron energy density and electron energy diffusivity.

- Source terms

$$S_e = \sum x_r k_r N_n n_e \tag{8}$$

$$S_i = \sum x_r k_r N_n n_e \tag{9}$$

$$S_\varepsilon = \sum x_r k_r N_n n_e \varepsilon_r \tag{10}$$

$$k_r = \sqrt{2e/m_e} \int \varepsilon \sigma_r(\varepsilon) f_0(\varepsilon) \mathrm{d}\varepsilon \tag{11}$$

Table 1 Energy of reactions NH_3 SiH_4

Reactions	Energy (eV)
1: $e + NH_3 \Rightarrow e + NH_3^*$	0.42
2: $e + NH_3 \Rightarrow 2e + NH_3^+$	10.2
3: $e + SiH_4 \Rightarrow e + SiH_4^*$	0.27
4: $e + SiH_4 \Rightarrow e + SiH_4$	8.01
5: $e + SiH_4 \Rightarrow e + SiH_2 + H_2$	2.2
6: $e + SiH_4 \Rightarrow e + SiH_3 + H$	0.4

Table 2 Kinetic coefficient of reactions NH_3 SiH_4

Reactions	Kinetic coefficient (m^3 s^{-1} mol^{-1})
7: $SiH_4 + SiH_2 \Rightarrow Si_2H_6$	2.8×10^7
8: $SiH_4 + H \Rightarrow SiH_3 + H_2$	1.9×10^6
9: $SiH_4 + NH \Rightarrow HSiNH_2 + H_2$	3.6×10^6
10: $SiH_4 + NH_2 \Rightarrow SiH_3 + NH_3$	2.4×10^6
11: $SiH_4 + NH \Rightarrow SiH_3 + NH_2$	2.4×10^7

2.2 Chemical Reactions

The calculation of the source terms (8)–(10), in the fluid model, is based on the Maxwellian distribution function f_0 plus the cross sections σ, which takes into account the most reactive elastic and inelastic collisions presented in Table 1.

The cross-sectional data used in simulation, SiH_4 NH_3 and H_2 collisions are given by Morgan [16].

The chemical species considered in this work are presented in Table 2. These species are very important for plasma chemistry and have an influence on the deposition process.

2.3 Boundary Conditions

The boundary conditions used in this section are similar to those presented by Samir et al. [11], Bavafa et al. [14].

The boundary condition for Poisson's equation is the potential value on the electrodes:

$V = 0$ electrical potential at the cathode.

$V_{rf} = V_0 \sin(\omega t)$ electrical potential at the anode.

The boundary condition for electrons has a flow proportional to their thermal velocity, while for ions it has a zero gradient near the walls:

$$\Gamma_e = \frac{v_{\text{th}} n_e}{2} - \gamma \Gamma_i \tag{12}$$

$$\Gamma_i = \mu_i n_i \nabla V \tag{13}$$

where v_{th} is the thermal speed of electrons and γ is the secondary emission coefficient of electrons.

$$v_{\text{th}} = \sqrt{\frac{8 k_B T_e}{\pi m_e}} \tag{14}$$

The energy flow of electrons toward the internal walls is given by:

$$\boldsymbol{\Gamma}_{\varepsilon} = \frac{5 v_{\text{th}} n_{\varepsilon}}{6} - \gamma \varepsilon_i \boldsymbol{\Gamma}_i \tag{15}$$

3   Results and Discussion

In this section, we present the fundamental characteristics of the radiofrequency plasma discharge for the deposition of a thin film of hydrogenated silicon nitride in a CCP reactor driven by a sinusoidal voltage of frequency 13.56 MHz such that the initial gas temperature is 573 K. Pressure is assumed to be 0.3 Torr, and RF voltage is assumed to be 130 V, applied to the anode. The results are presented for a discharge time of 7.31 µs.

The density and temperature of electrons are among the fundamental characteristics of RF plasma discharge; they are important parameters to consider in the fabrication of the thin silicon film. Indeed, the electron density is proportional to the rate of deposition of the thin film, so the analysis of this characteristic allows us to assess the uniformity of the deposited thin film. The electronic temperature is proportional to the ion bombardment energy [17], which allows assessing the temperature at the surface to avoid its deterioration.

Firstly, we will analyze the evolution of the characteristics of the plasma discharge in terms of electronic density and temperature in order to validate the numerical model whether in 1D or 2D simulation. Then, we compare the results obtained by the 1D simulation with the 2D one to determine the importance of the contribution of the 2D simulation.

Figure 1a, b show the evolution of the electron density, respectively in 1D and 2D. The density reaches a maximum value in the plasma mass; this means that the birth of electrons takes place in this volume due to the strong collisions between particles. The density gradually decreases from the center toward the electrodes because the neutralization reaction occurs faster for electrons and ions in the gains. The electrons
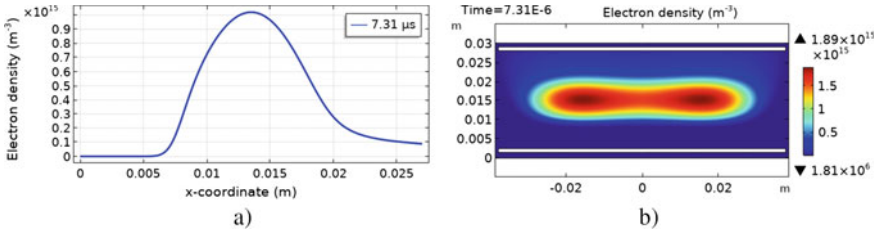
Fig. 1 **a** 1D and **b** 2D electron density (m^{-3})

move toward the plasma mass with high energy which is involved in the ionization of the gas to produce the positive ions which bombard the substrate.

Figure 2a, b show the evolution of the electronic temperature respectively in 1D and 2D. The electron temperature increases sharply in the region of the cathode and remains relatively uniform throughout the plasma mass. The 1D and 2D electronic speed is shown in Fig. 3a, b; in the cathode region, the velocity is strong due to the high electric field in this region (Fig. 4); it means that the electrons are strongly accelerated and therefore their temperature is maximum in the cathode region. In the plasma mass, Fig. 3 shows that there is a significant drop in electron velocity in this region, due to collisions and the decrease in electric field. In the region of the anode, the electric field is strong but the effect of the diffusion allows to decrease the velocity in this region.
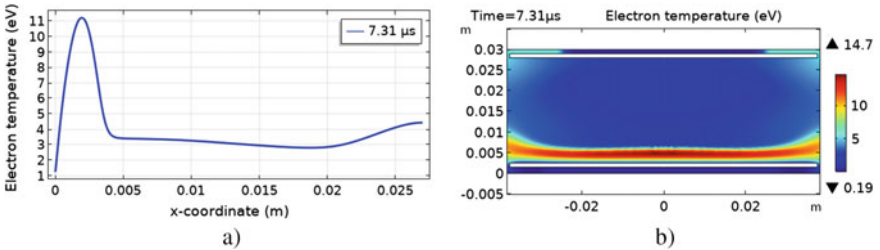
Fig. 2 **a** 1D and **b** 2D electronic temperature (eV)
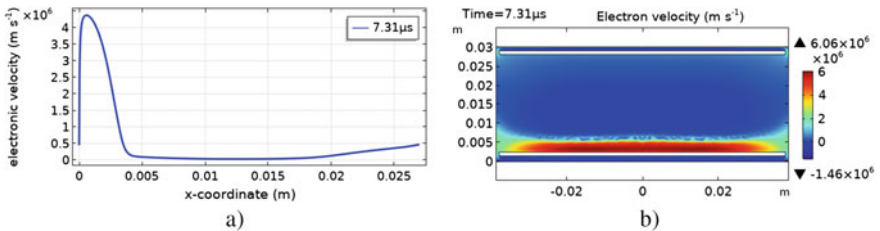
Fig. 3 **a** 1D and **b** 2D electronic velocity (m s^{-1})
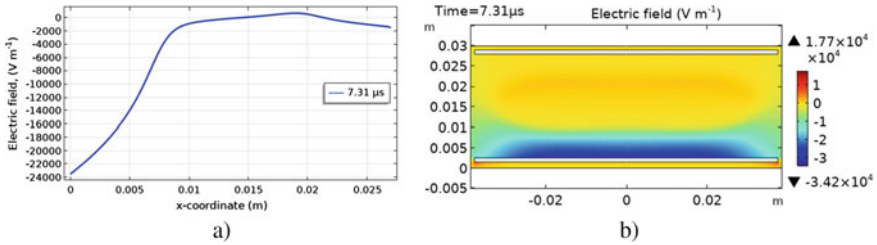
Fig. 4 **a** 1D and **b** 2D electric field (V m^{-1})

By comparing these results with other simulation works Bavafa [14], Liu [17] and Su [18], and also by experimental works and Fauroux [19], we notice that the characteristics of the plasma in both 1D and 2D simulations follow the normal evolution of a radiofrequency discharge in a CCP reactor, hence the validity of the numerical model used.

The maximum density in 2D has a value almost a double (1.98×10^{15} m^{-3}) compared to 1D (1.02×10^{15} m^{-3}). This means that there are a number of electrons that are added: These are the electrons that move in the radial direction. Therefore, the 2D simulation reveals a high electron density compared to the one obtained by the 1D simulation, which cannot be neglected. In addition, due to the increase in the number of electrons, an increase in the maximum temperature is expected. Indeed, Fig. 2 show that the temperature in 2D (14.6 eV) is more than that of 1D (11.2 eV). Thus, we can conclude that the use of the 2D simulation has succeeded to reveal the spatio-temporal evolution of the characteristics of the discharge closer to reality compared to the 1D one, which results in a better understanding of the physico-chemical processes of the reactor, and consequently provide a better optimization in the cost and the quality of the deposition of the thin film of hydrogenated silicon nitride.

4 Conclusions

In this work, we have performed a one-dimensional (1D) and two-dimensional (2D) spatio-temporal simulation of a capacitive coupled plasma reactor using silicon diluted with ammonia and hydrogen, where the fluid equations are solved by the numerical finite element method. Our objective is to validate the numerical model and to reveal the interest of 2D simulation compared to 1D simulation. To do this, we have studied the 1D and 2D spatio-temporal evolution of the fundamental characteristics of the plasma discharge of hydrogenated silicon nitride. Then, we have compared the results obtained from the 1D and 2D simulation.

The results show that the spatio-temporal evolution of the fundamental characteristics of the plasma discharge follows the normal evolution of the low-pressure and low-temperature RF plasma discharge. In addition, the 2D simulation allowed

us to obtain a higher electron density (1.98×10^{15} m^{-3}), almost a double, and a quite higher electron temperature (14.6 eV) compared to those found using the 1D simulation (1.02×10^{15} m^{-3} for density and 11.2 eV for temperature). The 1D simulation seems to be unsatisfactory to explain the characteristics of plasma discharges. Consequently, the use of 2D simulation is primordial and significantly better than the 1D.

Due to the ability of the 2D simulation to reveal the evolution of radicals in the radial direction as well as axial direction, we deduce that the use of 2D simulation reveals much better information and closer to reality than the one obtained through 1D simulation. Therefore, the 2D simulation presents an important tool for understanding and optimizing plasma devices, with the aim of having a uniform deposition of a thin film of hydrogenated silicon nitride without deterioration of the surface.

References

1. Chen TC, Kuo TW, Lin YL, Ku CH, Yang ZP, Yu IS (2018) Enhancement for potential-induced degradation resistance of crystalline silicon solar cells via anti-reflection coating by industrial PECVD methods. Coating 8(12):418
2. Grari M, Zoheir C, Yousfi Y et al (2021) Effect of pressure and space between electrodes on the deposition of SiNxHy films in a capacitively coupled plasma reactor. Chin Phys B 30(5):055205
3. Koutsoureli M, Siannas N, Papaioannou G (2017) Temperature accelerated discharging processes through the bulk of PECVD silicon nitride films for MEMS capacitive switches Microelectron Reliab 76:634
4. Grari M, Zoheir C (2020) Numerical modeling of non-equilibrium plasma discharge of hydrogenated silicon nitride (SiH$_4$/NH$_3$/H$_2$). Int J Eng 33(8):1449
5. Celiberto R, Capitelli M, Colonna G, D'Ammando G, Esposito F, Janev RK et al (2017) Elementary processes and kinetic modeling for hydrogen and helium plasmas. Atoms 5(2):18
6. Wang XF, Jia WZ, Song YH, Zhang YY, Dai ZL, Wang YN (2017) Hybrid simulation of electron energy distributions and plasma characteristics in pulsed RF CCP sustained in Ar and SiH4/Ar discharges. Phys Plasmas 24(11):113503
7. Ravari F, Fazeli SM, Bozorgzadeh HR, SadeghzadehAhari J (2017) Kinetic model study of dry reforming of methane using cold plasma. Phys Chem Res 5(2):408
8. Rousso A, Mao X, Chen Q, Ju Y (2019) Kinetic studies and mechanism development of plasma assisted pentane combustion. Prod Combust Inst 37(4):5603
9. Zhou J, Pérez-Grande D, Fajardo P, Ahedo E (2019) Numerical treatment of a magnetized electron fluid model within an electromagnetic plasma thruster simulation code. Plasma Sources Sci Technol 28:115004
10. Grari M, Zoheir C (2019) Numerical modeling of plasma silicon discharge for photovoltaic application. Mater Today Proc 13:888
11. Samir T, Liu Y, Zhao LL, Zhou YW (2017) Effect of driving frequency on electron heating in capacitively coupled RF argon glow discharges at low pressure. Chin Phys B 26:115201
12. Lichtenberg AJ (2005) Principles of plasma discharges and materials processing. Wiley, Hoboken
13. Grari M, Zoheir C (2020) Numerical characteristics of silicon nitride SiH$_4$/NH$_3$/H$_2$ plasma discharge for thin film solar cell deposition. In: 2021 Proceeding of the 2nd international conference on electronic engineering and renewable energy, ICEERE 2020. Springer, Singapore, p 230

14. Bavafa M, Ilati H, Rashidian B (2008) Comprehensive simulation of the effects of process conditions on plasma enhanced chemical vapor deposition of silicon nitride. Semicond Sci Technol 23(9):095023
15. Hao D, Jia C, Linhong J, Yuchun S (2012) Simulation of cold plasma in a chamber under high-and low-frequency voltage conditions for a capacitively coupled plasma. J Semicond 33(10):104004
16. Morgan database, www.lxcat.net. Retrieved on 27 Oct 2016
17. Liu Y, Booth JP, Chabert P (2018) Plasma non-uniformity in a symmetric radiofrequency capacitively-coupled reactor with dielectric side-wall: a two dimensional particle-in-cell/Monte Carlo collision simulation. Plasma Sources Sci Technol 27(2):025006
18. Su LW, Chen W, Uchino K, Kawai Y (2018) Two-dimensional simulations of multi-hollow VHF SiH_4/H_2 plasma. AIP Adv 8(2):025316
19. Fauroux A, Vandenabeele C, Pflug A, Lucas S (2020) Experimental and theoretical study of a magnetron DC-PECVD acetylene discharge: determination of the main species and reactions taking place in the plasma. J Surfcoat 126195

Effect of Next-Nearest-Neighbors Intersite Coupling on the Band Structure of a One-Dimensional Photonic Crystal

Mohamed El Ghafiani, Yamina Rezzouk, Soufyane Khattou, Madiha Amrani, Mohammed Moutaouekkil, and El Houssaine El Boudouti

Abstract   We investigate the effect of the next-nearest-neighbor (NNN) interaction on the band structures of one-dimensional (1D) photonic crystal in the context of the Green's function approach. Substantial effect has been remarked when the length of these intersite couplings changes; in particular, we show the appearance of a periodic set of flat-bands whose flatness get destroyed if the length of the couplings get tuned, giving rise to absolute bandgaps inside the first Brillouin zone. Furthermore, changing the material's nature of the NNN intersite couplings shows that when the NNN hopping parameter is so small there exists double degenerate nearly flat-bands in the middle of the bandgaps. As the hopping parameter increases, the degeneracy is lifted and the width of the bandgaps decreases until it disappears completely.

Keywords   Next-nearest-neighbors · 1D photonic crystals · Band structures

1   Introduction

The next-nearest-neighbor (NNN) intersite coupling is an important mechanism and plays a non-trivial role in modulating the properties of real materials [1]. The influence of such interaction phenomena has attracted considerable attention to study various physical applications like entanglement of the Heisenberg chain [2], evolution of the low-energy charge and spin dynamics of quasi-1D spin chains [3], the spectrum of plasmon excitations in graphene considering the NNN tight-binding model [4], etc. In the optical context, the effect of the NNN coupling has widely been discussed on graphene like nanotube structures [5], coupled optical microcav-

M. El Ghafiani (✉) · Y. Rezzouk · S. Khattou · M. Amrani · M. Moutaouekkil · E. H. El Boudouti
LPMR, Département de Physique, Faculté des Sciences, Université Mohammed I, Oujda, Morocco
e-mail: mohamed.elghafiani@ump.ac.ma

ities [6], Bose–Einstein condensation in optical lattices [7], quantum signature of breathers in 1D ultracold bosons in deep optical lattices [8], photonic superlattice to implement 1D random mass Dirac equation on a chip [9], etc. In addition, the behavior of band gaps in one-dimensional waveguides taking into account only the first-nearest-neighbour has been intensively studied [10]. However, to our knowledge, the effect of next-nearest-neighbor on the bandgaps of photonic structures based on waveguides has not been treated before.

In this paper, we investigate analytically and numerically the effect of NNN intersite coupling on the band structures of a 1D periodic photonic waveguide. Such a geometry can be realized in the experiment, since the NNN intersite couplings can be fabricated by a photonic arrays of photonic lattice [11]. Our purpose is to clarify the special influence of NNN interactions coupling on the band structures of our system. The calculation results show that the band structures undergo substantial changes. By allowing for NNN intersite coupling, absolute bandgaps are created inside the first Brillouin zone, in addition to the appearance of flat-bands. A flat-band is a completely dispersionless energy band that extends in the whole Brillouin zone. As the kinetic energy is completely quenched, which suppresses wave transport, particle interaction becomes dominant, leading to some exotic or unconventional correlated ground states [12]. In the past years, systems exhibiting flat-bands have attracted considerable interest and are demonstrated in a variety of two-dimensional (2D) lattice systems, including the 2D Lieb lattice [13], kagome lattice [14], dice lattice [15]. The analytical calculations in this paper were carried out using the Green's function method[10].

This paper is organized as follows : In Sec. 2, we give a brief review of the Green's function approach [10] which enables to derive the dispersion relation for the system under study [Fig. 1a]. In Sec. 3, we give an analytical and numerical discussion of the effects of NNN interaction on the band structure diagram of our system. Section 4 is devoted to the conclusion.
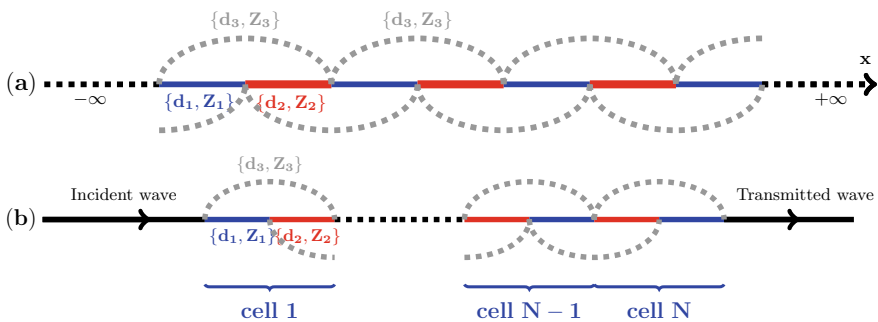
Fig. 1 **a** Infinite periodic photonic waveguide made of two alternating wires of lengths d_1 and d_2 and impedances Z_1 and Z_2 respectively, with NNN interaction included by means of intersite coupling wires of length d_3 and impedance Z_3 (gray dashed lines). **b** Finite structure formed of N cells sandwiched between two semi-infinite wires

2 Theoretical Model

The geometry of our system under study is presented in Fig. 1a which is a periodic 1D photonic crystal made of two alternating segments having lengths d_1 and d_2 and impedances Z_1 and Z_2, respectively, with NNN interactions introduced by means of segments of length d_3 and impedance Z_3. Our first step is to know the Green's function of our system from which one can derive any property of the system being studied.

Let us first recall the Green's function of the elementary segments constituting the infinite system under study. The inverse Green's function in the interface space $M_i = \{0, d_i\}$ of a wire of length d_i and impedance Z_i is given by a (2×2) matrix as follows [10]

$$g_1^{-1}(M_i, M_i) = \begin{pmatrix} -\frac{F_i C_i}{S_i} & \frac{F_i}{S_i} \\ \frac{F_i}{S_i} & -\frac{F_i C_i}{S_i} \end{pmatrix} = \begin{pmatrix} a_i & b_i \\ b_i & a_i \end{pmatrix}, \tag{1}$$

where $F_i = -j\frac{\omega}{Z_i}$, $S_i = -j\sin(\frac{\omega\sqrt{\varepsilon_r}}{c}d_i)$, and $C_i = \cos(\frac{\omega\sqrt{\varepsilon_r}}{c}d_i)$. ω, c and ε_r are the angular frequency, the speed of light and material's relative permittivity. Using Eq. (1), we can construct the inverse Green's function in the interface space M of the infinite waveguide shown in Fig. 1a, namely

$$g^{-1}(M, M) =$$

$$\begin{pmatrix} \ddots & \ddots & \ddots & & \ddots & & & \ddots & & \\ & b_3 & b_1 & a_1 + a_2 + 2a_3 & & b_2 & & b_3 & & \\ & & b_3 & b_2 & a_1 + a_2 + 2a_3 & & b_1 & & b_3 & \\ & & & b_3 & & b_1 & a_1 + a_2 + 2a_3 & b_2 & b_3 \\ & & & & \ddots & & \ddots & & \ddots & \ddots & \ddots \end{pmatrix}.$$

$$\tag{2}$$

Our system is periodic in the x direction; the Fourier transform $[g(k, M, M)]^{-1}$ of the infinite matrix [Eq. (2)] in a unit cell of length $D = d_1 + d_2$, indicated by a blue dashed square, is given as follows :

$$g^{-1}(k; M, M) = \begin{pmatrix} a_1 + a_2 + 2a_3 + b_3 \left(e^{ik_B d_3} + e^{-ik_B d_3}\right) & b_1 + b_2 e^{-ik_B D} \\ b_1 + b_2 e^{ik_B D} & a_1 + a_2 + 2a_3 + b_3 \left(e^{ik_B d_3} + e^{-ik_B d_3}\right) \end{pmatrix}, \tag{3}$$

where $D = d_1 + d_2$ is the period and k_B is the Bloch wave number of the infinite periodic structure. Note that k_B is real inside the allowed bands and complex inside the gaps of the infinite system. Then, one can deduce the dispersion relation of the infinite photonic crystal using the following equation [10]

$$det\left(g^{-1}(k; M, M)\right) = 0. \tag{4}$$

3 Analytical and Numerical Results

3.1 Effect of NNN Intersite Coupling Wire's Length on the Band Structures

First, we study the effect of changing the length of the NNN intersite coupling wires on the band structures. For this matter, we consider a simple 1D periodic photonic crystal made of segments of lengths $d_1 = d_2 = d$ and impedance Z with NNN interactions included by means of intersite coupling wires of length d_3 and the same impedance Z [Fig. 1a]. In this case, the dispersion relation [Eq. (4)] takes the simple form

$$\sin(\epsilon \ \Omega) \ \cos(k_B d) + \sin(\Omega) \ \cos(\epsilon \ k_B d) - \sin(\Omega(1 + \epsilon)) = 0, \qquad (5)$$

where $\epsilon = d_3/d$ and $\Omega = \frac{\omega\sqrt{\varepsilon_r}}{c}d$.

Figure 2 gives the band structures (i.e., the dimensionless frequency Ω versus $k_B d$) for the infinite system [Eq. (5)] presented in Fig. 1a along with the transmission

Fig. 2 Band structures along with the transmission coefficient through a finite system made of $N = 16$ [Fig. 1b] cells for $\epsilon = 1.8, 2$ and 2.3 respectively
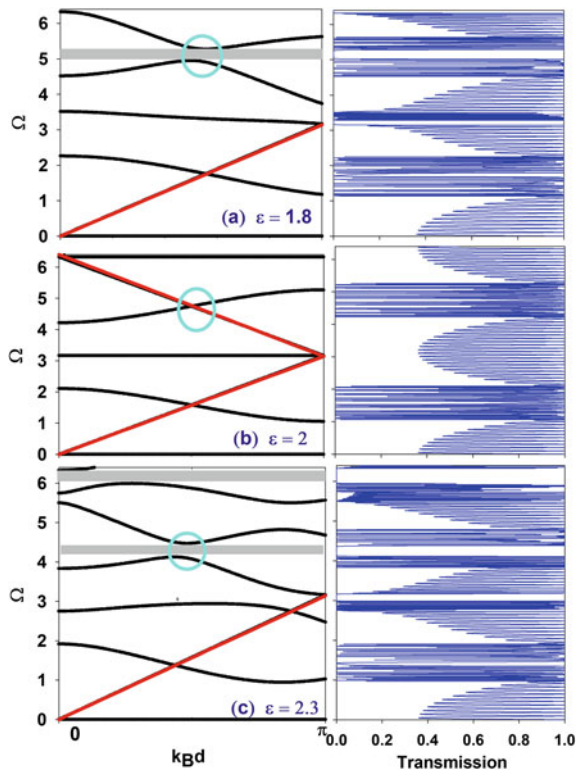
coefficient through a finite structure [Fig. 1b] made of $N = 16$ cells for three values of ϵ. We remark a very good agreement between the two curves, i.e., where there is a gap (the gray area) in the band structures, the transmission coefficient vanishes. The non-dispersive band indicated by a red line in Fig. 2b represents the eigenvalues of our system in the absence of NNN interaction. However, by allowing for NNN interaction two solutions appear, decoupled if ϵ is an integer (e.g., $\epsilon = 2$) as shown in Fig. 2b. These two solutions could be both real as in the region $1.2 < \Omega < 2.3$, or one real and another complex for $0 < \Omega < 1.2$. If ϵ is detuned from its integer case, the two solutions interfere at the crossing point in Fig. 2b indicated by cyan circle and give rise to what we call absolute bandgaps indicated by gray area in Figs. 2a and 2c where the two solutions are both complex. Moreover, the introduction of NNN interaction gives rise also to flat-bands, especially if ϵ is an integer (e.g., $\epsilon = 2$) as shown in Fig. 2b, but as long as ϵ is no longer an integer (e.g., $\epsilon = 1.8$ or $\epsilon = 2.3$) as shown in Figs. 2a and 2c, these flat-bands get affected broadly, thus become dispersive. Indeed, if $\epsilon = n$ where n is an integer, we can expand $\sin(n\Omega)$ and $\sin(\Omega(n + 1)\Omega)$ using Chebychev polynomials so that we can factor out $\sin(\Omega)$ from both terms as follows

$$\sin(n\Omega) = \sin(\Omega)U_{n-1}(\cos(\Omega)) \text{ and } \sin(\Omega(n + 1)) = \sin(\Omega)U_n(\cos(\Omega)) , \quad (6)$$

where U_n is Chebyshev polynomial of second kind. Then, by using Eq. (6) one can write Eq. (5) as follows

$$\sin(\Omega)\{U_{n-1}(\cos(\Omega)) \cos(k_B d) + \cos(\epsilon \ k_B d) - U_n(\cos(\Omega))\} = 0 \quad (7)$$

Equation (7) shows the decoupling of the term $\sin(\Omega)$ that is responsible of flat-bands at the reduced frequencies $\Omega = n\pi$. These flat-bands appear only when $\epsilon = n$ with n being an integer; otherwise, we do not expect them to remain completely flat as illustrated in Fig. 2a, c.

Another peculiar point in Fig. 2 is the opening of absolute bandgaps (indicated by cyan circles in Fig. 2a, c) inside the first Brillouin zone which happens only in the upper half of the band structure diagrams (i.e., $\pi \leq \Omega \leq 2\pi$). However, in the frequency domain $0 \leq \Omega \leq \pi$, there is no lifting of degeneracy at the crossing points between red and black curves. Indeed, one can show easily that $\Omega = k_B d$ satisfies Eq. (5) irrespective of the value of ϵ, giving rise to a non-dispersive band.

3.2 Effect of NNN Intersite Coupling Wire's Impedance on the Band Structures

Now, we focus on discussing the effect of the material's nature of the NNN intersite coupling wires on the band structures diagram. In this section, we consider that

$d_2 = d_1 = d$, $d_3 = 2d$, $F_2 = 2F_1$, and take $\delta = \frac{F_3}{F_1} = \frac{Z_1}{Z_3}$ as a variable quantity. In this case, Eq. (4) becomes

$$4\left[C^2 - \frac{5}{4}S^2 - \cos(k_B D)\right]C^2 + 6\,C^2\left[C' - \cos(k_B D)\right]\delta + \left[C' - \cos(k_B D)\right]^2\delta^2 = 0 \tag{8}$$

where $S = \sin(\Omega)$, $C = \cos(\Omega)$ and $C' = \cos(2\Omega)$, with $\Omega = \frac{\omega\sqrt{\varepsilon_r}}{c}d$ and $D = 2d$.

If for instance we set $\delta = 0$ (i.e., $Z_3 \to \infty$) in Eq. (8) (i.e., we eliminate the NNN coupling), then we recover the dispersion relation of our system without NNN intersite interaction [10], namely

$$\cos(k_B D) = \cos^2(\Omega) - \frac{5}{4}\,\sin^2(\Omega). \tag{9}$$

In the contrary, if δ is so large (i.e., $Z_3 \to 0$) in such a way that we can neglect the other terms, Eq. (8) becomes simply

$$\left(C' - \cos(k_B D)\right)^2 = 0, \text{ i.e., } \cos(k_B D) = \cos(2\Omega), \tag{10}$$

which is the dispersion relation for an infinite 1D photonic crystal of impedance Z. In this case, we have short-circuited our system and the waves tend to traverse the

Fig. 3 **a** Band structures of the system without NNN coupling. **b**–**d** give the band structures of the system with NNN interaction allowed for $\delta = 0.1$, 1 and 10 respectively
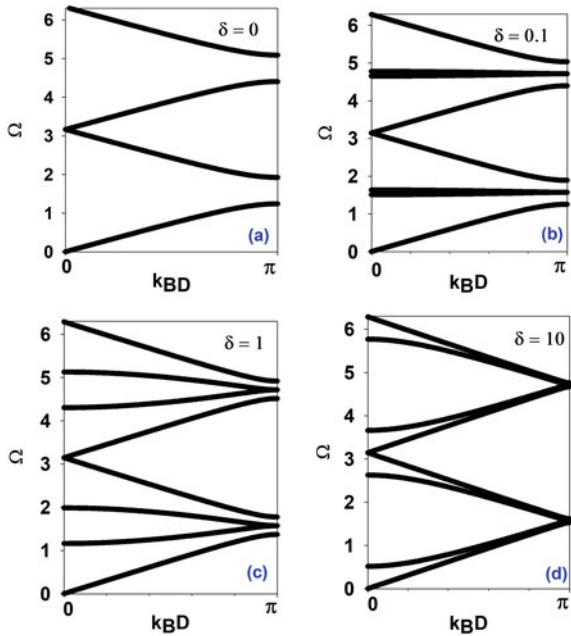
segments of length d_3 and impedance Z_3 without interacting with the other segments. Furthermore, Eq. (10) does not admit any bandgaps, which means that by increasing δ the bandgaps of the system get eliminated.

Figure 3a gives the band structure for the system presented in Fig. 1a without NNN interaction introduced [Eq. (9)]. The band's edges are given by $\cos(k_B D) = \pm 1$ [Eq. (9)]. At the center of the first Brillouin zone [i.e., $\cos(k_B D) = 1$] they are given by $\Omega = n\pi$. For these frequencies, the gaps close [Fig. 3a]. However, the band's edges at the limit of the Brillouin zone [i.e., $\cos(k_B D) = -1$] are given by $\cos(\Omega) = \pm 1/3$ (i.e., $\Omega = 1.23, 1.91, 4.37, 5.05$). These values are in accordance with the numerical results in Fig. 3a.

Figures 3b–d show the dispersion curves [Eq. (8)] for our system with NNN interaction allowed. When δ is small we notice the appearance of two nearly flat-bands at the center of the two bandgaps that get split up as δ increases. Indeed, when $\delta << 1$ Eq. (8) becomes to first order in δ

$$\cos^2(\Omega)\left\{4\left[C^2 - \frac{5}{4}S^2 - \cos(k_B D)\right] + 6\left[C_3 - \cos(k_B D)\right]\delta\right\} = 0, \qquad (11)$$

which shows a decoupling of $\cos^2(\Omega)$ that gives flat-bands with double degeneracy at the reduced frequencies $\Omega = \frac{n\pi}{2}$, where n is an integer. As δ increases, this degeneracy is lifted and the bandgaps get split up as shown in Figs. 3b–d. At a certain value of δ ($\delta \approx 0.5$) the bandgaps of our system get eliminated completely. When δ becomes so large (i.e., $Z_3 \to 0$), the band structure of our system tends to approach that corresponding to an infinite wire of impedance Z whose dispersion relation is given by Eq. (10). In this case, the system gets short-circuited and the waves will tend to traverse the system made of the intersite coupling wires with less resistance.

4 Conclusion

In summary, we have studied the effect of introducing NNN interaction on the band structures diagram of a 1D periodic photonic crystal. This study has been performed in two parts, first we discussed the effect of changing the length of the NNN intersite coupling wires. We have demonstrated that when the length of the NNN intersite couplings is a multiple integer of the periodicity of the infinite system (i.e., $\epsilon = n$ where n is an integer), there exist two decoupled solutions for our system, as well as a periodic set of flat-bands. Deviating ϵ from this special case makes the two decoupled solutions interfere which destroys the flatness of the flat-bands, hence the appearance of absolute bandgaps inside the first Brillouin zone. The second part was devoted to the study of the effect of changing the impedance of the NNN intersite coupling wires, whereby allowing for small NNN interaction, we have observed the appearance of double degenerate nearly flat-bands in the middle of the bandgaps of

our system. Further increasing δ causes the two flat-bands to become dispersive which makes the bands split up and reduces the width of the gaps. Above a certain value of δ, they may get eliminated, especially if δ is sufficiently tuned. The theoretical results presented in this work can be realized experimentally using coaxial cables in the radio-frequency domain for example. This work is in progress.

References

1. Xiong D, Zhang Y, Zhao H (2014) Temperature dependence of heat conduction in the fermi-pasta-ulam-β lattice with next-nearest-neighbor coupling. Phys Rev E 90:022118
2. Gu SJ, Li H, Li YQ, Lin HQ (2004) Entanglement of the Heisenberg chain with the next-nearest-neighbor interaction. Phys Rev A 70:052302
3. Kumar U, Nocera A, Price G, Striwinter K, Johnston S, Datta T (2020) Spectroscopic signature of next-nearest-neighbor hopping in the charge and spin dynamics of doped one-dimensional antiferromagnets. Phys Rev B 102:075134
4. Kadirko V, Ziegler K, Kogan E (2013) Next-nearest-neighbor tight-binding model of plasmons in graphene. Graphene 2:97101
5. Ghanbaripour H, Rezania H (2020) The effects of next-to-nearest neighbor hopping amplitude on electrical properties of graphine like nanotude structure. J Solid State Sci Technol 9:51002
6. Laha A, Biswas A, Ghosh S (2017) Next-nearest-neighbor resonance coupling and exceptional singularities in degenerate optical microcavities. J Opt Soc Am B 34:20502058
7. Zaleski T, Kopeć T (2010) Effect of next-nearest-neighbour hopping on bose-einstein condensation in optical lattices. J Phys B At Molecular Opt Phys 43:85303
8. Djoufack ZI, Kenfack-Jiotsa A, Nguenang J-P (2017) Quantum signature of breathers in 1d ultracold bosons in optical lattices involving next-nearest neighbor interactions. Int J Mod Phys B 31:1750140
9. Keil R, Zeuner JM, Dreisow F, Heinrich M, Tüunnermann A, Nolte S, Szameit A (2013) The random mass dirac model and long-range correlations on an integrated optical platform. Nat Commun 4:1-9
10. Dobrzyński L, Akjouj A, Leveque G, Al-Wahsh H, Pennec Y, Ghouila-Houri C, Talbi A, Djafari-Rouhani B, Jin Y et al (2020) Photonics. Elsevier
11. Caselli N, Riboli F, La China F, Gerardino A, Li L, Linfield EH, Pagliano F, Fiore A, Intonti F, Gurioli M (2015) Tailoring the photon hopping by nearestneighbor and next-nearest-neighbor interaction in photonic arrays. ACS Photonics 2:565–571
12. Tang L, Song D, Xia S, Xia S, Ma J, Yan W, Hu Y, Xu J, Leykam D, Chen Z (2020) Photonic flat-band lattices and unconventional light localization. Nanophotonics 9:11611176
13. Zhang S, Jin L (2019) Flat band in two-dimensional non-hermitian optical lattices. Phys Rev A 100:043808
14. Chern G-W, Chien C-C, Di Ventra M (2014) Dynamically generated flat-band phases in optical kagome lattices. Phys Rev A 90:013609
15. Kolovsky AR, Ramachandran A, Flach S (2018) Topological flat wannier-stark bands. Phys Rev B 97:045120

Y-shaped Demultiplexer Based on Asymmetric Loop Photonic Waveguides

Mimoun El-Aouni, Youssef Ben-Ali, Ilyass El Kadmiri, and Driss Bria

Abstract In this work, we investigate the possibility to realize a Y-shaped demultiplexer based on asymmetric loop photonic waveguides. This demultiplexer contains one input line and two output lines (two transmission channels). These two output channels are grafted at the same position with the input line. This system creates the Fano resonances, which are very sensitive to the length of each waveguide of asymmetric loops in the proposed structure. These Fano resonances are characterized by a specific profile in the transmission spectrum. We base on the theory of interface response in a continuous medium, which allows us to calculate the Green function of any composite material. The Green function allows us to determine the transmission rates T_1 and T_2 through the two channels and the reflection rate R through the input line. This system can be used as an electromagnetic waveguides, demultiplexer, and filter of two frequencies in each channel with high performance.

Keywords Photonic demultiplexer · Asymmetric loops · Fano resonance modes

1 Introduction

Electromagnetic demultiplexer (system with several transmission channels) based on photonic system waveguides has been the subject of many theoretical and experimental works to develop integrated technology [1, 2]. It is a very important device (demultiplexer) in microwave and electromagnetic communication systems because of its higher transmission rate and quality factor Q [3–5]. Recently, many researchers have focused their work on the study of the propagation of electronic, photonic, and phononic waves in demultiplexer waveguide systems, because of their ability to

M. El-Aouni (✉) · Y. Ben-Ali · I. El Kadmiri · D. Bria
Laboratory of Materials, Waves, Energy and Environment, Team of Waves, Acoustics, Photonics and Materials, Mohamed I University, Oujda, Morocco
e-mail: mimounelaouni01@gmail.com

Y. Ben-Ali
Engineering Sciences Laboratory (LSI), Multidisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, B.P. 1223, Taza Gare, Morocco

guide, filter, select, and separate waves with a very good quality. For example, Ben-Ali et al. [6] proposed a coaxial cable defective Y-shaped demultiplexer photonic waveguides with one input line and two periodical output lines. This device presents very high transmission rates and is able to separate two electromagnetic incident mixed signals and guide each one in a channel. They have presented a theoretical and experimental (based on the coaxial cables in the radio frequency domain) study of photonic demultiplexers based on detuned resonators. The demultiplexer consists of Y-shaped structures with one input line and two output lines. The results demonstrate the presence of Fano or an electromagnetic induced transparency (EIT) resonance [7]. The width of the Fano or EIT resonances can become zero for an appropriate choice of the resonators' lengths. In addition to photonic waveguide demultiplexers, we can find phononic or acoustic demultiplexers. For example, El Kadmiri et al. [8] realized a simple phononic Y-shaped demultiplexer based on two output lines, and each output line is containing a single asymmetric resonator. The result shows the existence of a complete transmission beside a zero transmission enables to select a given frequency on one output line, by canceling the second line transmission and the input line reflection. The position and the width at half the maximum of the acoustic induced transparency AIT resonance depend strongly on the varying lengths of the waveguides which should be chosen appropriately. Mouadili and Djafari-Rouhani [9] have given an analytical demonstration of the possibility to realize a simple phononic demultiplexer based on Fano and acoustic induced transparency resonances. The demultiplexer consists of a Y-shaped waveguide with an input line and two output lines. Each output line contains two resonators grafted either at a given position or at two positions far from the input line. This last system can create the AIT resonance, which can be tuned by appropriately choosing the lengths of the different waveguides constituting these systems. In this work, we present a numerical study of the transmission and reflection rates of electromagnetic waves propagating in a one-dimensional photonic demultiplexer consisting of the input line and two outputs lines (two transmission channels). Each output line consists of a finite length segment contacted to two guides forming between them a loop, and the two output lines are grafted to the input line (see Fig. 1). This system considered in question can be used in the field of electromagnetic and microwave telecommunications, namely the guidance of electromagnetic and microwave waves, filtering, and wave separation. All spectra transmission and reflection are calculated on the basis of the Green function method [10–12].

2 Theoretical Formalism

We use the Green function method to calculate the transmission and reflection coefficients. For this, we need only the inverse of the Green function of the whole system in the interfaces to space $M = \{1, 2, 3, 4, 5\}$. This process claims the knowing of the
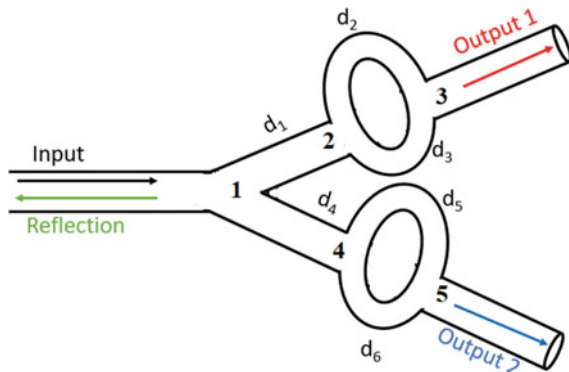
Fig. 1 Y-shaped demultiplexer with one input line and two output lines. In the first output line, an asymmetrical loop of length d_2 and d_3 is inserted between positions 2 and 3 contact with a segment of length d_1. In the second output line, an asymmetrical loop of length d_5 and d_6 is inserted between positions 4 and 5 contact with a segment of length d_4

inverse Green function of elementary constituents, namely the inverse of the interface Green function element of the three semi-infinite waveguides shapes input and two output lines: $g_s^{-1}(1, 1) = g_s^{-1}(3, 3) = g_s^{-1}(5, 5) = A_0 = -F_s$.

We suppose all that the guides and loops have the same characteristics (i.e., $F_1 = F_2 = F_3 = F_4 = F_5 = F_6 = F_s = F$). The inverse of the Green function of the previous constituents gives us the inverse of the Green function of the composite structure in the space of the interface $M = \{1, 2, 3, 4, 5\}$:

$$
g^{-1}(M, M)
$$
$$
= \begin{pmatrix}
A_0 + A_1 + A_4 & B_1 & 0 & B_4 & 0 \\
B_1 & A_1 + A_2 + A_3 & B_2 + B_3 & 0 & 0 \\
0 & B_2 + B_3 & A_0 + A_2 + A_3 & 0 & 0 \\
B_4 & 0 & 0 & A_4 + A_5 + A_6 & B_5 + B_6 \\
0 & 0 & 0 & B_5 + B_6 & A_0 + A_5 + A_6
\end{pmatrix} \quad (1)
$$

where $A_0 = -F$; $A_i = \frac{-FC_i}{S_i}$; $B_i = \frac{F}{S_i}$; $F = \alpha_i = -j\frac{\omega}{c}\sqrt{\varepsilon_1}\left(c = \frac{3\times10^8\,\text{m}}{\text{s}}\right)$; $j = \sqrt{-1}$; $C_i = \cosh(\alpha_i d_i)$ and $S_i = \sinh(\alpha_i d_i)$.

The function $g(M, M)$ is given by the following relation:

$$
g(M, M) = \frac{1}{\det[g^{-1}(MM)]} \begin{pmatrix}
a & b & c & d & e \\
b & f & g & h & i \\
c & j & k & l & m \\
d & n & o & p & q \\
e & r & s & t & u
\end{pmatrix} \quad (2)
$$

The reflection coefficient in the input line of the demultiplexer is given by:

$$r = -1 - 2Fg(1, 1) \text{ with } g(1, 1) = \frac{a}{\det\left[g^{-1}(MM)\right]} \tag{3}$$

The two transmission coefficients in the first and second channels are given, respectively, by:

$$t_1 = -2Fg(1, 3) \text{ with } g(1, 3) = \frac{c}{\det\left[g^{-1}(MM)\right]} \tag{4}$$

$$t_2 = -2Fg(1, 5) \text{ with } g(1, 5) = \frac{e}{\det\left[g^{-1}(MM)\right]} \tag{5}$$

The expression of the reflection rate in the input line is given by:

$$R = |r|^2 \tag{6}$$

and the transmission rates in the two output lines are given, respectively, by:

$$T_1 = |t_1|^2 \text{ and } T_2 = |t_2|^2 \tag{7}$$

The conservation law $R + T_1 + T_2 = 1$ can be easily deduced from Eqs. (6) and (7).

3 Results and Discussions

In this work, we consider the photonic demultiplexer which is presented in Fig. 1 and we present the transmission rate T_1 through the first output line (red line), T_2 through the second output line (blue line), and the reflection rate R in the input line (green line) for different values of the geometrical parameters such as $d_1 = d_4, d_2 \neq d_3$, and $d_5 \neq d_6$, and the values of the relative permittivity of the each waveguides are equal ($\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = \varepsilon_5 = \varepsilon_6 = 2.3$).

3.1 Asymmetrical Loops in the Two Channels of the Electromagnetic Demultiplexer

In this part, we examine in Fig. 2a, b the evolution of the transmission (T_1) in the first output line (red line), the transmission (T_2) in the second output line (blue line), and the reflection (R) in the input line (green line) as a function of the frequency. The geometrical parameters of the system considered play a very important role in finding the Fano resonance. According to Fig. 2a, where $d_1 = d_4 = 1$ m, $d_2 = 0.63$ m, $d_3 = 0.9$ m, $d_5 = 0.57$ m, and $d_6 = 0.8$ m, we notice the existence of two Fano
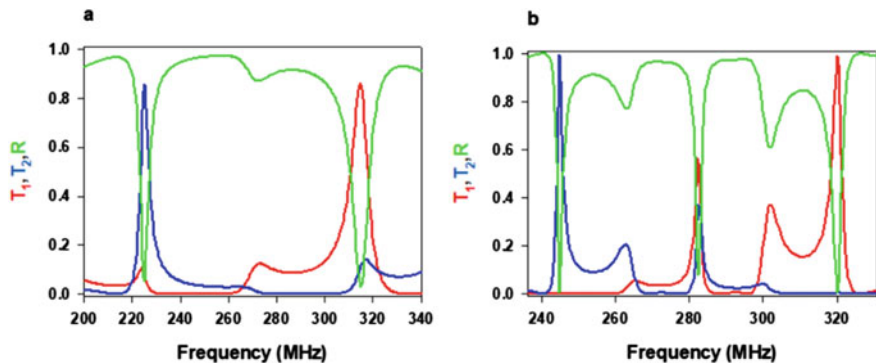
Fig. 2 Variation of the transmission in output 1 (red line), output 2 (blue line), and reflection in the input line (green line) of the photonic demultiplexer as a function of the frequency. We take $d_1 = d_4 = 1$ m, $d_2 = 0.63$ m, $d_3 = 0.9$ m, $d_5 = 0.57$ m, and $d_6 = 0.8$ m for (**a**) and $d_1 = d_4 = 1$ m, $d_2 = 1.63$ m, $d_3 = 1.9$ m, $d_5 = 1.6$ m, and $d_6 = 1.8$ m for (**b**)

resonances with transmission rates reaching $T_1 = T_2 = 0.85$. One Fano resonance is filtered by the first channel (red line) with a frequency around $f_1 = 314.95$ MHz, and the other Fano resonance is filtered by the second channel (blue line) around $f_2 = 226.764$ MHz. The frequency difference between the two Fano resonances is very important, which allowed us to avoid the problem of overlapping between these two waves. According to Fig. 2b, where $d_1 = d_4 = 1$ m, $d_2 = 1.63$ m, $d_3 = 1.9$ m, $d_5 = 1.6$ m, and $d_6 = 1.8$ m, we obtain two Fano resonances of the different frequencies by each channel, with a very significant frequency difference between the two Fano resonances. The two Fano resonances guided by the first channel have the transmission rate equal to $T_1 = 1$ and $T_1' = 0.6$ with a specific frequency $f_1 = 320$ MHz and $f_1' = 285$ MHz, respectively. Moreover, the two Fano resonances obtained by the second channel have the transmission rate equal to $T_2 = 1$ and $T_2' = 0.4$ with frequency values, namely $f_2 = 244.086$ MHz and $f_2' = 285$ MHz, respectively.

3.2   Color Maps of the Transmissions T_1 and T_2

In this part, we report in Fig. 3a, b the variation of the maximum transmission as a function of the frequency and the loop length d_3. The red and dark blue colors indicate the high and low values of transmission T_1 and T_2. For each point of the plane (frequency and d_3), there are resonance modes of transmission T_1 and T_2.

We can see that some regions exist where the transmissions are maximal. For example, a maximum transmission T_1 obtained in the frequency interval varied between 299.202 and 330.697 MHz for $d_3 = 1.9$ m. Moreover, a maximum transmission T_2 obtained in the frequency interval varied between 236.212 and 251.96 MHz for $d_3 = 1.9$ m. Some regions exist where the transmission is average (e.g., when the
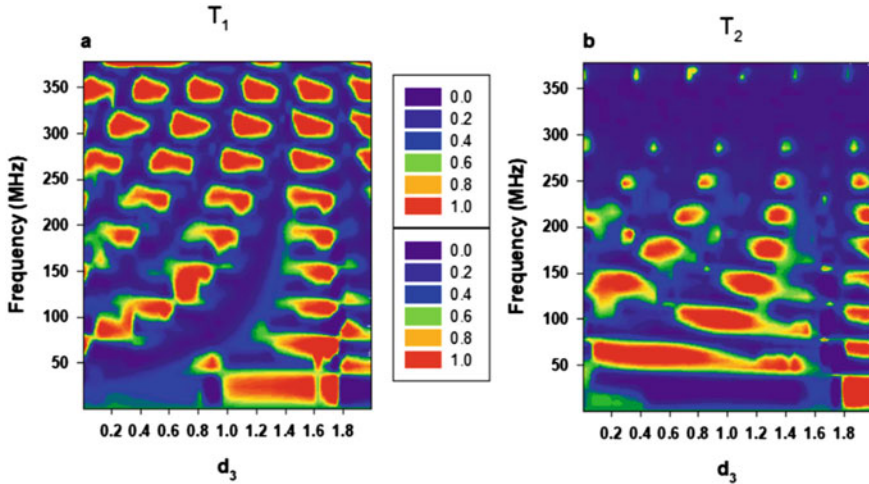
Fig. 3 a, b Represent the color maps of the transmission rates T_1 and T_2 as a function of the frequency and d_3. The geometric parameter values as $d_1 = d_4 = 1$ m, $d_2 = 0.63$ m, $d_5 = 0.57$ m, and $d_6 = 0.8$

frequency varied between 280 and 285 MHz for $d_3 = 1.9$ m). Also, we remark that specific frequency varied between 230 and 240 MHz for $d_3 = 1.9$ m where we have a reflection total $R = 1$ ($T_1 = T_2 = 0$).

3.3 Asymmetrical Loops in the Two Channels of the Electromagnetic Demultiplexer: Fano Resonance

In this paragraph, we show in Fig. 4 the evolution of T_1 (red line), T_2 (blue line), and R (green line) as a function of the frequency for asymmetric loops in two output lines of the photonic demultiplexer, and we take $d_1 = d_4 = 1$ m, $d_2 = 0.88$ m, $d_3 = 1.5$ m, $d_5 = 1.6$ m, and $d_6 = 1.8$ m. We notice through Fig. 4a that the transmission in the first output line reaches unity ($T_1 = 1$ around frequency $f_1 = 74.643$ MHz) while the transmission in the second output line and the reflection are canceled ($T_2 = R = 0$). This mode is near to one transmission zero (Fano resonance). According to Fig. 4b, we notice that the transmission in the second output line reaches unity ($T_2 = 1$ around frequency $f_2 = 245$ MHz), while the transmission in the first output line and the reflection is equal to zero ($T_1 = R = 0$). These types of resonances in (a) and (b) are called Fano resonances. The possibility of realizing a Fano resonance has been the subject of several previous works [13, 14]. The Fano resonance is used in different applications for communication [15, 16], the drop filter [17, 18], ultrafast switching [19], and detection glucose concentration sensing [20].
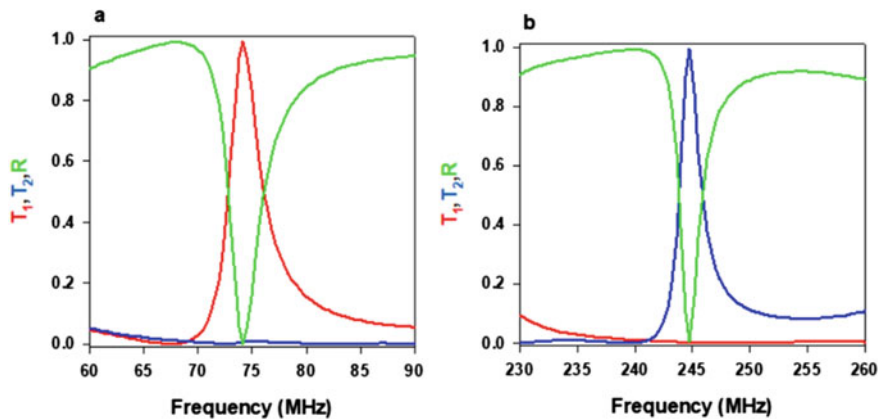
Fig. 4 Variation of the transmission in output 1 (red line) and output 2 (blue line) and reflection in the input line (green line) of the photonic demultiplexer as a function of the frequency. We take $d_1 = d_4 = 1$ m, $d_2 = 0.88$ m, $d_3 = 1.5$ m, $d_5 = 1.6$ m, and $d_6 = 1.8$ m

3.4 Asymmetrical and Symmetrical Loops in the Two Channels

In this section, we represent in Fig. 5 the variation of T_1 (red line), T_2 (blue line), and R (green line) as a function of the frequency for two cases. In the first case, we consider our system composed of the symmetrical loop in the first channel and the asymmetrical loop in the second channel. We realize a condition between the geometrical parameters of the loops by $d_2 = d_3 = 2d_5$ and $d_6 = 3d_5$. We obtain the results indicated in Fig. 5a, c, e with three different values of d_5 ($d_5 = 0.25$ m, $d_5 = 0.36$ m, and $d_5 = 0.6$ m). According to Fig. 5a ($d_5 = 0.25$ m), we notice the existence of a single mode with a maximum transmission ($T_1 = 1$), which is filtering by the first channel with a frequency $f_1 = 132.279$ MHz. At this frequency ($f_1 = 132.279$), the transmission in the second channel and the reflection at the input line are canceled ($T_2 = R = 0$). Taking $d_5 = 0.36$ m (Fig. 5c), we observe the apparition of two modes with a maximum transmission ($T_1 = T'_1 = 1$); the first mode is situated at the frequency $f_1 = 94.485$ MHz and the second mode at a frequency $f'_1 = 274$ MHz. These two modes are guided by the first channel while $T_2 = R = 0$. For $d_5 = 0.6$ m (Fig. 5e), we obtain three maximum transmission modes with different frequencies ($f_1 = 59.84$ MHz, $f'_1 = 160.624$ MHz, and $f''_1 = 261.408$ MHz). These modes are filtered by the first channel, while $T_2 = R = 0$. In the second case, we consider our system which is composed of an asymmetrical loop in the first channel and a symmetrical loop in the second channel; we realize a condition between the geometrical parameters of the loops by $d_3 = 3d_2$ and $d_6 = d_5 = 2d_2$. According to Fig. 5b, d, f, the transmission in the first channel and the reflection in the input line are canceled ($T_1 = R = 0$) when the transmission in the second channel is maximal ($T_2 = 1$). The resonance modes depend on the waveguide length d_2. At the
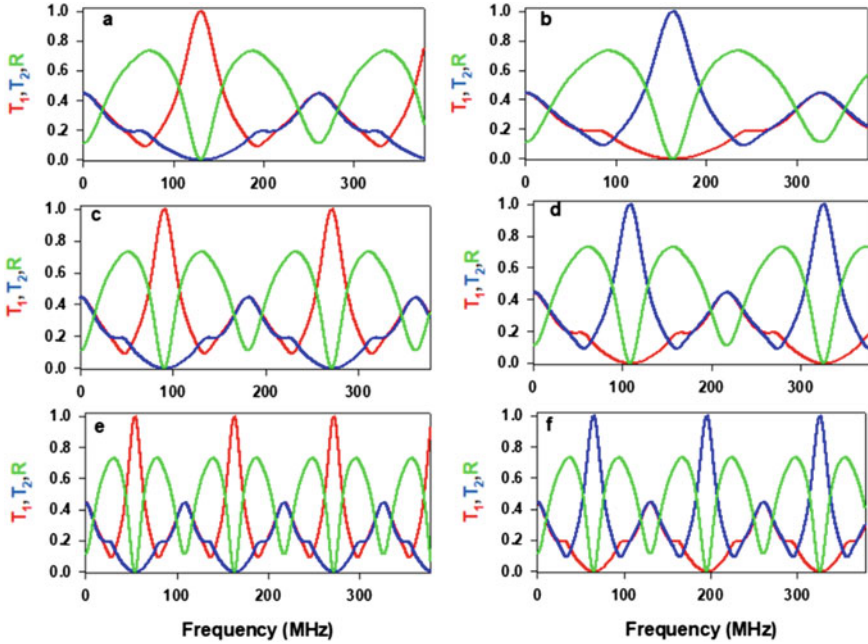
Fig. 5 Variation of the transmission in output 1 (red line) and output 2 (blue line) and reflection in the input line (green line) for symmetric and asymmetric loops of the demultiplexer as a function of the frequency. We take $d_2 = d_3 = 2d_5$ and $d_6 = 3d_5$ ($d_5 = 0.25$ m, $d_5 = 0.36$ m, and $d_5 = 0.6$ m) for (**a, c, e**) and $d_3 = 3d_2$ and $d_6 = d_5 = 2d_2$ ($d_2 = 0.2$ m, $d_2 = 0.3$ m, and $d_2 = 0.5$ m) for (**b, d, f**)

value $d_2 = 0.2$ m (Fig. 5b), we notice the existence of one mode with a maximal transmission ($T_2 = 1$) around a frequency $f_2 = 163.774$ MHz. This mode is guided by the second channel while $T_1 = R = 0$. Then, we take $d_2 = 0.3$ m (Fig. 5d), and we obtain two modes with a maximum transmission ($T_2 = T_2' = 1$) around frequency $f_2 = 110.232$ MHz and $f_2' = 321.249$ MHz, respectively, while $T_1 = R = 0$. For the last value of $d_2 = 0.5$ m (Fig. 5f), we observe the existence of three modes with a maximum transmission ($T_2 = T_2' = T_2'' = 1$), the first mode exists at a frequency $f_2 = 62.99$ MHz, the second mode exists at $f_2' = 192.119$ MHz, and the third mode exists at a frequency $f_2'' = 321.249$ MHz.

4 Conclusion

In this work, we have indicated the possibility of realizing a simple photonic demultiplexer based on Fano resonances. The demultiplexer is a Y-shaped waveguides with one input line and two output lines (two transmission channels). In the first output line, an asymmetrical loop of length d_2 and d_3 is contact with a segment of length d_1. In the second output line, an asymmetrical loop of length d_5 and d_6 is contact

with a segment of length d_4. This device presents very high transmission rates and can separate four electromagnetic incident mixed signals of the different frequencies, namely $f_1 = 244.086\,\text{MHz}$ and $f_1' = 285\,\text{MHz}$ through a first channel and $f_2 = 321.25\,\text{MHz}$ and $f_2' = 285\,\text{MHz}$ through a second channel. The existence of a complete transmission near to a transmission zero (Fano resonance) makes it possible to select a given frequency on one channel, by canceling the transmission in the second channel and the reflection in the input line. The positions of the Fano resonances depend strongly on the variation of the loop lengths that must be chosen appropriately.

References

1. Niemi T, Frandsen LH, Hede KK, Harpoth A, Borel PI, Kristensen M (2005) Wavelength-division demultiplexing using photonic crystal waveguides. IEEE Photonics Technol Lett 18(1):226–228
2. Mouadili A, Khattou S, Amrani M et al (2021) Y-shaped demultiplexer photonic circuits based on detuned stubs: application to radiofrequency domain. In: Photonics, vol 8, no 9. Multidisciplinary Digital Publishing Institute, p 386
3. Masilamani S, Punniakodi S (2020) Photonic crystal ring resonator based optical MUX/DEMUX design structures: a survey and comparison study. J Opt 49(2):168–177
4. Gupta ND, Janyani V (2014) Dense wavelength division demultiplexing using photonic crystal waveguides based on cavity resonance. Optik 125(19):5833–5836
5. Akosman AE, Mutlu M, Kurt H, Ozbay E (2012) Dual-frequency division de-multiplexer based on cascaded photonic crystal waveguides. Phys B 407(20):4043–4047
6. Ben-Ali Y, El Kadmiri I, Ghadban A, Ghoumid K, Mazari A, Bria D (2021) Two-channel demultiplexer based on 1D photonic star waveguides using defect resonators modes. Prog Electromagnet Res B 93:131–150
7. Mouadili A, El Boudouti EH, Soltani A, Talbi A, Haddadi K, Akjouj A, Djafari-Rouhani B (2018) Photonic demultiplexer based on electromagnetically induced transparency resonances. J Phys D Appl Phys 52(7):075101–075125
8. El Kadmiri I, Ben-Ali Y, Khaled A, Bria D (2020) Y-shaped branch structure using asymmetric resonators for phononic demultiplexing. Mater Today Proc 27:3033–3041
9. Mouadili A, Djafari-Rouhani B (2020) Acoustic demultiplexer based on Fano and induced transparency resonances in slender tubes. Eur Phys J Appl Phys 90(1):10902–10909
10. Dobrzynski L (1987) Interface response theory of continuous composite materials. Surf Sci 180(2–3):489–504
11. Ben-Ali Y, Elamri FZ, Ouariach A, Falyouni F, Tahri Z, Bria D (2020) A high sensitivity hydrostatic pressure and temperature based on a defective 1D photonic crystal. J Electromagnet Waves Appl 34(15):2030–2050
12. Ben-Ali Y, El Kadmiri I, Tahri Z, Bria D (2020) High quality factor microwave multichannel filter based on multi-defectives resonators inserted in periodic star waveguides structure. Prog Electromagnet Res C 104:253–268
13. Limonov MF, Rybin MV, Poddubny AN, Kivshar YS (2017) Fano resonances in photonics. Nat Photonics 11(9):543–554
14. Bekele DA, Yu Y, Hu H, Guan P, Galili M, Ottaviano L, Mork J (2018) Signal reshaping and noise suppression using photonic crystal Fano structures. Opt Express 26(15):19596–19605
15. Rostami A, Nazari F, Banaei HA, Bahrami A (2010) A novel proposal for DWDM demultiplexer design using modified-T photonic crystal structure. Photonics Nanostruct Fundam Appl 8(1):14–22

16. Naghizade S, Sattari-Esfahlan SM (2020) An optical five-channel demultiplexer-based simple photonic crystal ring resonator for WDM applications. J Opt Commun 41(1):37–43
17. Ren H, Jiang C, Hu W, Gao M, Wang J (2006) Photonic crystal channel drop filter with a wavelength-selective reflection micro-cavity. Opt Express 14(6):2446–2458
18. Li L, Liu GQ, Chen YH, Tang FL, Huang K, Gong LX (2013) Photonic crystal multi-channel drop filters with Fabry-Pérot micro-cavity reflection feedback. Optik-Int J Light Electron Opt 124(17):2608–2611
19. Yu Y, Heuck M, Hu H, Xue W, Peucheret C, Chen Y, Mørk J (2014) Fano resonance control in a photonic crystal structure and its application to ultrafast switching. Appl Phys Lett 105(6):061117–061122
20. Rakhshani MR (2019) Fano resonances based on plasmonic square resonator with high figure of merits and its application in glucose concentrations sensing. Opt Quant Electron 51(9):1–16

Narrow Localized Electronic States Induced by Defective Electronic Comb-Like Quantum Wires

Siham Machichi, Fatima Zahra Elamri, Youssef Ben-Ali, Farid Falyouni, and Driss Bria

Abstract A theoretical study is done using the transfer matrix method (TMM) to study the propagation of electrons waves in a one-dimensional electronic comb-like waveguides (ECWGs) quantum wires structure composed of a semiconductor segment (GaAs) grafted at its extremity by one semiconductor resonator (GaAlAs). For the first time, we have proposed a simple system with a single GaAlAs resonator embedded between two semi-infinite electronic waveguides. As a result, the energies of the corresponding resonator eigenmodes shift to lower energies as the resonator length increases. The resulting eigenmodes are characterized by high transmission rates but very low-quality factor values, which highlight the importance of the periodicity of the system. Further, in order to create localized electronic defect states with higher quality factors in the bandgaps, a defective system is exanimated. Indeed, a defective segment and resonator are inserted into a perfect system. This defective system can create localized electronic defect states in the electronic bandgaps. The analysis of the transmission spectra and the maximum transmission of these defects states shows a shift to lower energies by increasing the length of the defects. These localized defect states can interact with each other by exchanging energy. Furthermore, it has been shown that the number of localized states induced by the presence of these defects, depends on the defect lengths parameters.

S. Machichi (✉) · F. Z. Elamri · Y. Ben-Ali · F. Falyouni · D. Bria
Laboratory of Materials, Waves, Energy and Environment, Team of Acoustics, Photonics and Materials, Faculty of Sciences, Mohamed First University, Oujda, Morocco
e-mail: siham.machichi@ump.ac.ma

Y. Ben-Ali
Engineering Sciences Laboratory (LSI), Multidisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, Taza, Morocco

203

1   Introduction

Low-dimensional semiconductor materials have provoked great interest at the beginning of the nineteenth century, either theoretically or experimentally because of their frequent applications such as optoelectronics [1, 2], quantum electronics [3], etc. The use of semiconductor materials in different artificial periodic systems (multi-quantum wires, electronic waveguides based on the loops or the resonators, etc.) has been evolving very rapidly for a long time due to their extraordinary properties, in particular the passbands and band gaps.

Recently, our team has concentrated their attention to study comb-like (1D) waveguide systems due to their very important properties. One-dimensional (1D) electronic comb-like waveguides based on quantum wires are extremely narrow structures, where electron transport is only possible in a very small number of transverse modes. This electronic system presents passbands and band gaps, where the transfer of electrons is forbidden through a very wide energy range (gap bands). The insertion of defects within the comb-like electronic waveguides structure allows the creation of localized defect states in the electronic gaps, similarly to the defect modes photonic or phononic localized in the photonic or acoustic band gaps [4, 5]. Akjouj et al. [6] using the green function showed the existence of electronic band gaps by the ECWGs, due to the periodicity of the system and the resonance states of the grafted resonators. Mrabti et al. [7] using the green function have shown the existence of electronic-induced transparency (EIT) and electronic-induced reflection (EIR) resonances in a magnetic structure composed of an Aharonov-Bohm ring attached vertically to two semi-infinite wires (waveguides). The main objective of this work is to study the propagation of electron waves in one-dimensional periodic structure made of materials with different electronic properties. We will study how this propagation can be modified by the insertion of one or more defects in the periodic structure considered. It conducts the filter and separate the localized defect states.

In this paper, we use the transfer matrix method [8] to study the behavior of electrons waves in a 1D defective ECWGs system. The perfect system is composed of the periodicity of the semiconductor segment (GaAs type) characterized by a length d_1 and the effective mass m_1 and grafted at its end by a semiconductor resonator (GaAlAs type) characterized by a length d_2 and the effective mass m_2. These segments and resonators are considered quantum wires. Our objective is to insert two defects, one at the segment of the length d_{01} (with an effective mass is m_1) and the other at the resonator of the length d_{02} (with an effective mass is m_2) located in the middle of the structure. We investigate the behavior of the localized defect states versus the different parameters of the system, in particular the parameters of the defects. The numerical results can be useful for the design of guiding, filtering the electrons waves.

2 Result and Discussion

In this section, we examine the numerical results found from the analytical calculations based on the transfer matrix method. We study the transmission and the band structure for an electronic comb-like waveguide structure containing defects placed at the position $j = 5$ at the resonator and the other at the guide in the middle of the structure (Fig. 1). We set the perfect structure parameters, GaAs concentration at $x_1 = 0$ and length $d_1 = 10$ nm for the segments and for the resonators with concentration $x_2 = 0.35$ and length $d_2 = 10$ nm (Ga$_{0.65}$Al$_{0.35}$As). The energies of the two unperturbed materials are $E_1 = 0$ meV and $E_2 = 330.4$ meV. The effective electron mass corresponds to GaAs is $m_1 = 0.067m_0$ and Ga$_{0.65}$Al$_{0.35}$As is $m_2 = 0.096m_0$ with $m_0 = 9.11 \times 10^{-31}$. The creation of defects in the structure of perfect comb electron waveguides gives rise to localized states within the band gaps.

2.1 Dispersion Relation and the Transmission Rate Passbands

Figure 2a shows the transmission spectrum as a function of energy for finite ECWGs consisting of $N = 9$. This curve shows very clearly that there are seven pass bands separated by six electronic gap bands (white areas) and two mini-gaps bands (white areas where the transmission is different to zero). The appearance of these bandgaps is due to the periodicity of the system and the eigenmodes of the resonator. Figure 2b represents the variation of the real part of Kd_1 versus energy for the case of an infinite structure. The black branches indicate the pass bands which are separated by electronic gap bands (white regions). This study of the dispersal relation is done for an electron, photonic, and acoustic star waveguide structure [5, 9, 10].
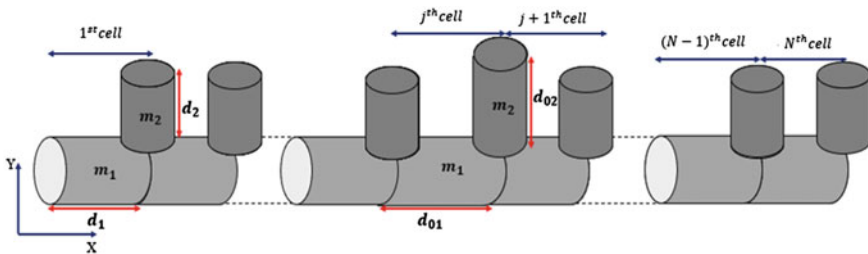
Fig. 1 Electronic comb-like system based on quantum wires containing N segment and resonator cells including two defects. Resonator defect is localized at the j site, and the segment defect is situated between the sites j and $j + 1$
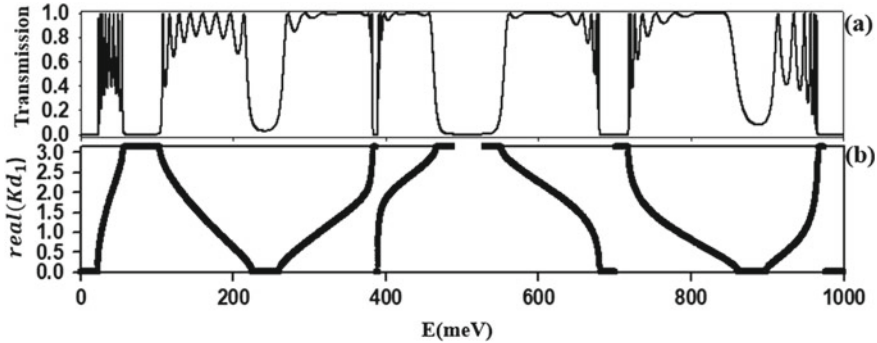
Fig. 2 **a** Transmission rate as a function of energy, when the cells number $N = 9$, **b** variation of the energy as a function of the real part of Kd_1

2.2 Transmission as a Function of the Energy for Different Lengths d_{01} and d_{02}

Figure 3 shows the variation of transmission rate as a function of the energy of the incident electronic wave for different values of lengths of the two defects d_{01} and d_{02}. According to the first case (a) where $d_{01} = d_{02} = 20$ nm, we observe that there is only one electronic defect state in the second bandgap with a transmission rate is $T = 0.75$, while the absence of the localized defect state in the other gaps. For the second case (b) where $d_{01} = d_{02} = 40$ nm, we notice that there are two localized states inside the fifth bandgap and two localized states inside the sixth band gap. In the same time, we remark a single localized state inside the second band gap. For the third case (c) where $d_{01} = d_{02} = 60$ nm, we notice two electronic states in the second band gap, a single state in the first, third and sixth band gap with different transmission rates. On the other hand, we have one state and two states respectively inside the fourth and fifth band gaps, these states associated with the defective resonator. For the last case (d) where $d_{01} = d_{02} = 80$ nm, we observe the appearance of a single electronic defect state in sixth band gap, four electronic states in the second, and fifth band gaps. From these results, we found that the insertion of both defects in a perfect periodic ECWGs structure leads to new localized defect states (electronic states) inside the band gaps with higher transmission rate and quality factor (narrow electronic states).

2.3 Variation of the Electronic Defect States Energy as a Function of the Defect Lengths d_{02} and d_{01}

The trace of the color map is allowed to obtain the geometrical parameters that lead to a maximum transmission rate of the localized defect states. We plot in Fig. 4 the color map of the transmission rate versus both the segment length defect d_{01} and of
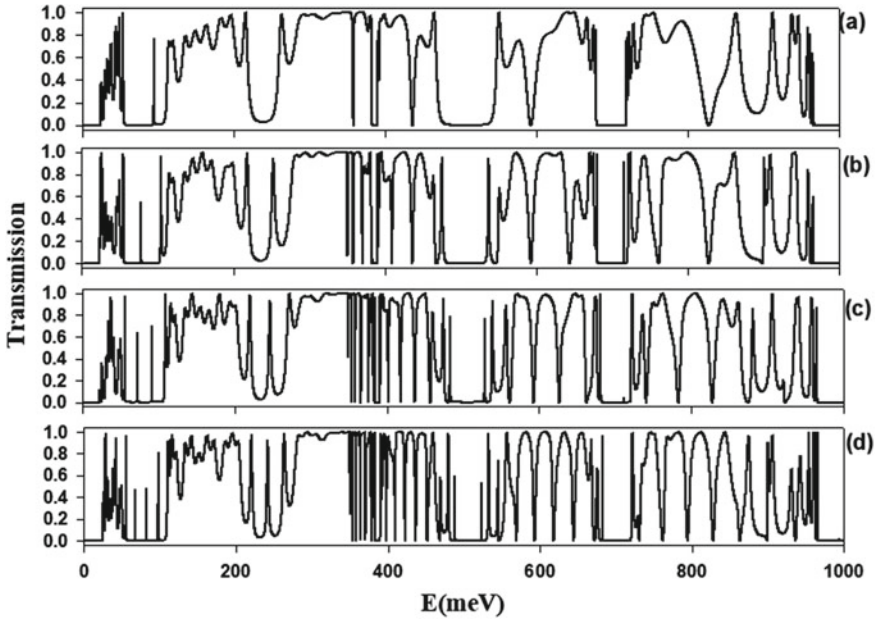
Fig. 3 Transmission spectrum as a function of the energy of the incident electronic wave (meV) for different defect segment length values d_{01} and resonator length d_{02} namely: **a** $d_{01} = d_{02} = 20$ nm, **b** $d_{01} = d_{02} = 40$ nm, **c** $d_{01} = d_{02} = 60$ nm, and **d** $d_{01} = d_{02} = 80$ nm, with $N = 9$ and $j = 5$

resonator length defect d_{02} of the fifth bandgap in interval of energy $E = [480-485]$, with $x_{01} = 0$, $x_{02} = 0.35$, $N = 9$ and $j = 5$. The red and purple areas corresponding respectively to the highest values ($d_{01} = 45.47$ nm and $d_{02} = 24.8$ nm) and lowest values ($d_{01} = 13$ nm and $d_{02} = 36$ nm) of transmission rate, and the green areas represent the means values ($d_{01} = 38.5$ nm and $d_{02} = 47$ nm) of transmission rate. Similar behavior is manifested in the photonic system of symmetric and asymmetric comb-like waveguide [11]. These maximum values of transmission rate (the purple zones) could be used in the electronic domain for electron filtering.

2.4 Energy of the Localized Defect States Versus $d_{01} = d_{02}$

In this part, we investigate through Fig. 5 the variation of energy of the localized electronic defect states as a function of the defect lengths d_{01} and d_{02} which $d_{01} = d_{02}$. The gray areas represent the pass bands of the infinite ECWGs, while the white areas present the gaps where the localized electronic defect states exist. We notice that the localized electronic states emerge from the high pass bands or high band gaps, decrease toward lower energies with increasing the lengths $d_{01} = d_{02}$ and finally enter inside the lower pass bands and become the resonant states of the perfect system.
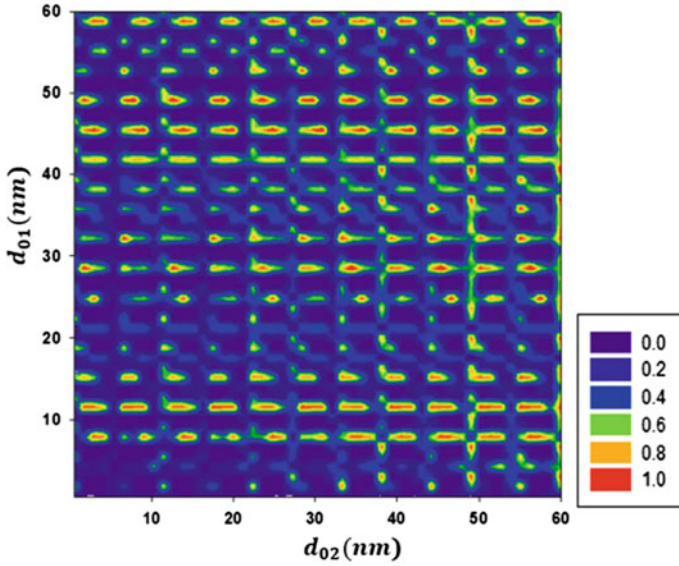
Fig. 4 Color map of the transmission rate versus both d_{01} and d_{02}

Noted that the states located in the first three bandgaps are associated to the segment defect states, while the localized electronic states of the resonator to appear at energy $E = 375$ meV. We notice also that there is also an overlap between the localized states of the segment and the localized states of the resonator for the particular values of lengths d_{01} and d_{02} (fourth gap).

Fig. 5 Variation of the energy as a function of the length $d_{01} = d_{02}$
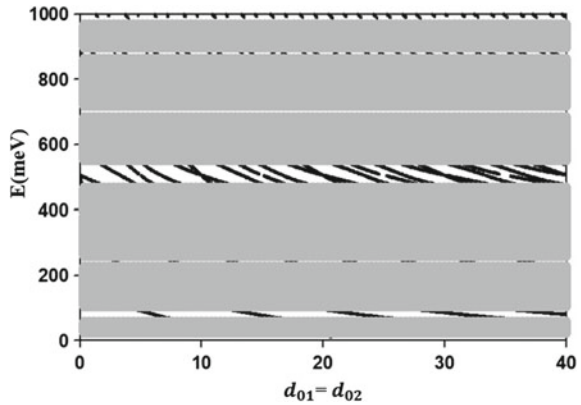
Fig. 6 Behavior of localized defect states as a function of the resonator defect length d_{02} for $d_{01} = 40$ nm
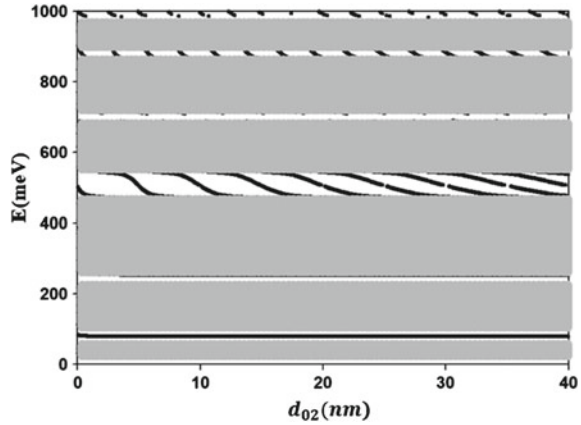
2.5 Maximums Transmissions as a Function of the Length d_{02}

In Fig. 6, we study the influence of the variation of the defect resonator length (keeping the defect segment length at $d_{01} = 40$ nm) on the energy of the localized electronic defect states in the band gaps. The gray areas indicate the pass bands while the white areas represent the band gaps. We note also that the localized electronic resonator defect states appear at the fourth band gap, in contrast to the localized electronic segment defect states which appear at the first band gap. Note that the defect segment localized states are independent of the variation of d_{02}, in contrast to the defect resonator localized states which move to lower energies with the increase of d_{02}. This behavior is recently observed in the 1D photonic CWGs containing two resonator defects located in the same site [11].

2.6 Maximum Transmissions as a Function of Lengths d_{01} and d_{02}

Now, we focus our attention on the fourth band gap observed in Fig. 6, and we study the effect of changing the defect segment length d_{01} (in red $d_{01} = 80$ nm and in black $d_{01} = 40$ nm) on the behavior of the electronic states located this fourth gap. Note that the states that are almost independent of d_{02} correspond to the localized segment defect states, while the localized resonator states move to lower energies with the increasing d_{02}. According this Fig. 7, we observe when $d_{01} = 80$ nm and $d_{01} = 40$ nm, that the resonator electronic states interact with the segment defect states in crossing point around a well determined value of energy $E = 552$ meV. After this interaction, we obtain a change in the behavior of these localized states with the
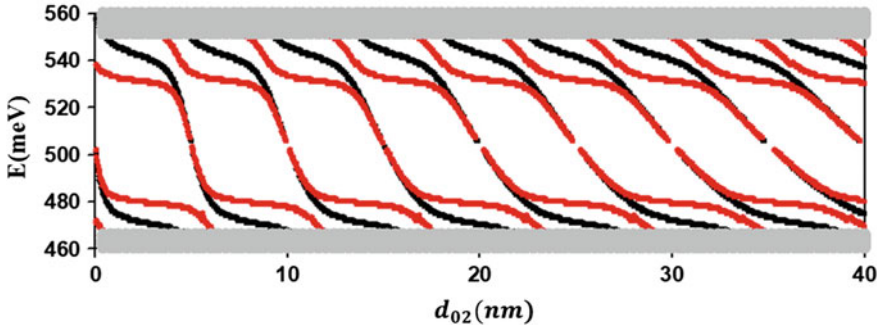
Fig. 7 Defect dependence on the defect length of the d_{02} resonator with of $d_{01} = 40$ nm (black line) and $d_{01} = 80$ nm (red)

increase of d_{02} that is to say that the resonator electronic states become the segment electronic states and vice versa.

3 Conclusion

In this present paper, we have presented a detailed study on the effect of defect insertion inside a 1D periodic electronic comb-like waveguides (ECWGs) system based on quantum wires. The perfect system is composed of a periodicity of semiconductor segments (GaAs) and grafted at its extremity by a semiconductor resonator (GaAlAs). We have shown that the number of localized electronic defect states in the bandgaps are very sensitive to the different parameters of the system, in particular to the lengths of the defects in this perfect structure. The results of this work show that when we insert two defects at the segment and resonator level, we can obtain the phenomenon of change of behavior between the localized defect states (energy exchange phenomenon between the two states). In addition, the number of these localized states increases with increasing the length of the resonator defect.

References

1. Levine BF (1993) Quantum-well infrared photodetectors. Appl Phys 74(8):R1–R81
2. Rosencher E (1992) Quantum-wells and infrared detection. Recherche 23(248):1270–1278
3. Marcuse D (2012) Principles of quantum electronics. Elsevier, Amsterdam
4. Ben-Ali Y, Tahri Z, Bria D (2019) Electromagnetic filters based on a single negative photonic comb-like structure. Prog Electromagnet Res C 92:41–56
5. Elmalki M, Khettabi A (2017) Study of various periodic acoustic lattices by two methods: transfer matrix and Green's method. In: International conference on electrical and information technologies (ICEIT). IEEE, pp 1–3

6. Akjouj A, Dobrzynski L, Djafari-Rouhani B, Vasseur JO, Kushwaha MS (1998) Giant electronic stop bands in one-dimensional comblike structures. EPL (Europhys Lett) 41(3):321–326
7. Mrabti T, Labdouti Z, Mouadili A, El Boudouti EH, Djafari-Rouhani B (2020) Aharonov-Bohm-effect induced transparency and reflection in mesoscopic rings side coupled to a quantum wire. Physica E Low-dimensional Syst Nanostruct 116:113770–113784
8. Li ZY, Lin LL (2003) Photonic band structures solved by a plane-wave-based transfer-matrix method. Phys Rev E 67(4):046607–046617
9. Rouhani BD, Vasseur JO, Akjouj A, Dobrzynski L, Kushwaha MS, Deymier PA, Zemmouri J (1998) Giant stop bands and defect modes in one-dimensional waveguide with dangling side branches. Prog Surf Sci 59(1–4):255–264
10. Errouas Y, Ben-Ali Y, Ouariach A, Tahri Z, Bria D (2020) Propagation of the electromagnetic waves in one-dimensional asymmetric photonic comb-like structure based on Fibonacci chains of grafted resonators. Mater Today Proc 27:3058–3064
11. Errouas Y, Ben-Ali Y, Tahri Z, Bria D (2020) Propagation of electromagnetic waves in one dimensional symmetric and asymmetric comb-like photonic structure containing defects. Mater Today Proc 31:S16–S23

Zak Phase and Topological Tamm States Between Two Photonic Comb Structures

Soufyane Khattou, Yamina Rezzouk, Madiha Amrani, Mohamed El Ghafiani, El Houssaine El Boudouti, and Bahram Djafari-Rouhani

Abstract We investigate the existence of Tamm states at the interface between two one-dimensional (1D) photonic crystals (PCs) through an analysis of local density of states (LDOS) using the Green's function method. The PCs are made of a comb-like structure consisting of stubs grafted periodically along a waveguide with different geometrical parameters. The Tamm states appear as maxima in the LDOS inside the common bandgaps of the periodic PCs. In addition, we prove the existence of such Tamm states using a topological invariant based on the Zak phase of the bulk band for each PC. The Zak phase is calculated using two different arguments, namely (i) the symmetry of the electric field at the band edge states and (ii) the sign of the reflection phases between each PC and a given waveguide. The Tamm state appears as well-defined resonance in the bandgap frequency area of the transmission spectra through two finite PCs in tandem. Our proposed design can be used as filter and sensor.

Keywords Tamm states · Local density of states · Zak phase · Photonic crystal · Comb-like structure

1 Introduction

Tamm states are localized states at the surface of a crystal, discovered by Igor Tamm for electrons in solid crystals [1]. Since then, Tamm states have been extended to various fields of wave physics such as acoustics [2, 3], optics [4, 5] and plasmonics [6] due to their interest in sensing applications. In optics, Tamm states are known as

S. Khattou (✉) · Y. Rezzouk · M. Amrani · M. El Ghafiani · E. H. El Boudouti
LPMR, Département de Physique, Faculté des Sciences, Université Mohammed I, Oujda, Morocco
e-mail: s.khattou@ump.ac.ma

B. Djafari-Rouhani
Département de Physique, Institut d'Electronique, de Microélectronique et de Nanotechnologie (IEMN), UMR CNRS 8520, Université de Lille, Villeneuve d'Ascq 59655, France

localized states that can be excited at the interface of two connected photonic crystals (PCs) [4], or between a photonic crystal and a metalic film, called Tamm plasmon (TP) [5]. Optical Tamm states (OTS) can be employed in numerous applications such as controlling spontaneous optical emissions [7], polariton lasers [4] and optical sensing technology [8].

Recently, topological invariance has shown a great deal of attention in modern physics due to fundamental interest in topological insulators and their potential applications [9–11]. In one-dimensional (1D) systems, there exist different ways that can precisely predict the existence of topological interface states at the interface between two PCs. One of the most important ways is to use topological properties through the Zak phase analysis, a special kind of Berry phase [12]. Different methods have been proposed to determine the Zak phase of each bulk band such as integrating the Berry phase in the Brillouin zone [10], making a symmetry analysis of the modes at the band edges [11, 13], or calculating the sign of the reflection phase of the two bandgaps sandwiching this band [13, 14].

In literature, Tamm states are generally studied through an analysis of transmission coefficient at the interface between two finite systems. In this paper, we show the existence of Tamm sates at the interface between two photonic comb-like structures by (i) directly using the dispersion relation of the two semi-infinite PCs, (ii) the calculation of the reflection phases at the two interfaces PC1/WG and PC2/WG (WG means waveguide) and (iii) using topological invariants through an analysis of the Zak phase θ_{Zak} of each bulk band in the 1D periodic system. The topological invariants are used to define the topological phase of each band and deduce the existence of interface states within specified bandgaps. The PC considered here is a 1D periodic comb-like structure composed by stubs grafted periodically along a waveguide. The analytical calculations of dispersion relation of the infinite and semi-infinite PCs, local density of states (LDOS) as well as transmission and reflection coefficients are obtained by using the Green's function formalism [15]. The theoretical predictions presented here can be verified experimentally in the radiofrequency domain using coaxial cables [15]. The experimental work is in progress. In addition, the results presented in this work can be transposed to nanometric plasmonic MIM waveguides with different applications for filtering and sensing in the THz domain [6, 16].

The rest of the paper is organized as follows: In Sect. 2, we study the existence of Tamm states through an analysis of the LDOS at the interface between two semi-infinite PCs. In Sect. 3, we show the existence of Tamm states at the connection point of two PCs using the topological invariants based on the Zak phase of each bulk band of the 1D periodic system and the reflection phases at the PC1/WG and PC2/WG interfaces. In Sect. 4, we summarize the main results of this work.

2   Tamm States at the Interface Between Two Semi-infinite Comb-Like Structures

In this section, we show the existence of interface Tamm states between two different semi-infinite PCs. The PCs are both based on a comb-like structure composed of stubs grafted periodically along a waveguide. The study is performed through an analysis of the dispersion relation of the two semi-infinite PCs and the LDOS at the interface between the two semi-infinite systems. This latter quantity is essential to define Tamm states at the interface which appear as maxima in the LDOS inside the common bandgaps of the two PCs.

We consider two semi-infinite PCs connected at the interface point ($x = 0$) (Fig. 1a). The first one (at the left-hand side) is a comb-like structure composed of stubs of length d_2 grafted periodically along a waveguide and separated by segments of length d_1, while the second one (at the right-hand side) is composed of stubs of length d_4 and segments of length d_3. The semi-infinite crystals are terminated by
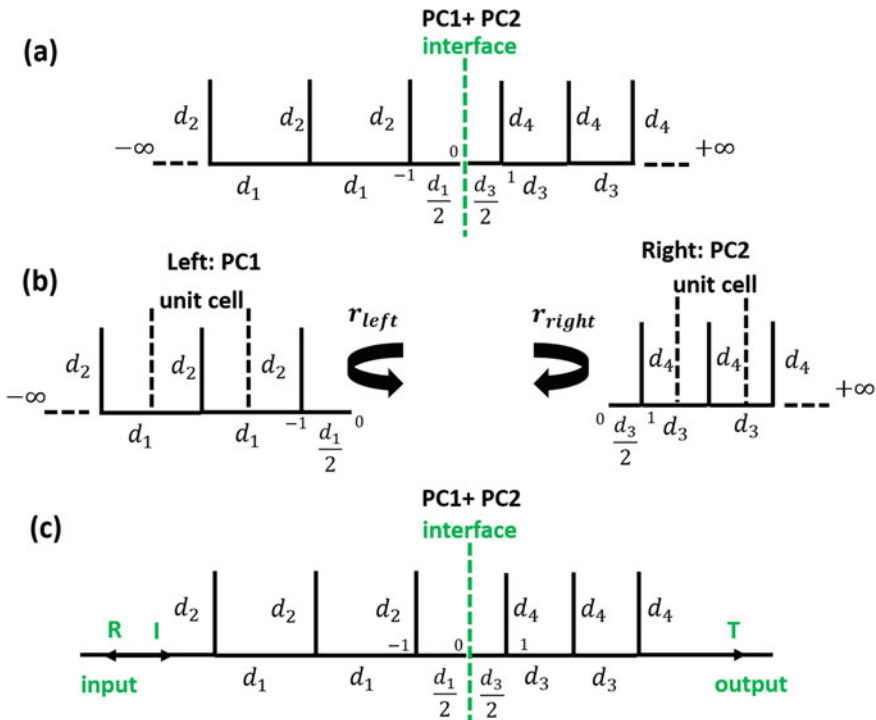
Fig. 1 **a** Interface between two semi-infinite PCs terminated by segments of lengths $\frac{d_1}{2}$ and $\frac{d_3}{2}$. **b** Semi-infinite PCs in contact with a waveguide. **c** Connection between two finite PCs composed each one by $N = 3$ stubs. I, R and T denote the incident, reflected and transmitted waves, respectively

segments of lengths $\frac{d_1}{2}$ and $\frac{d_3}{2}$, respectively. The LDOS at the interface between the two semi-infinite systems is given by,

$$n(\omega) = \frac{2\omega}{\pi} Im[\frac{g_1(\omega^2)g_3(\omega^2)}{g_1(\omega^2) + g_3(\omega^2)}], \tag{1}$$

where $g_i(\omega^2)$ $(i = 1, 3)$ is the Green's function element at the surface of each PC. For $H = 0$ boundary condition (i.e., vanishing magnetic field) at the surface, its expression is given by,

$$g_i(\omega^2) = \frac{1}{\Delta_i}\left[\frac{S_i}{F} - \frac{S_i'}{FC_i'}\left(C_i - \frac{1}{t_i}\right)\right], \tag{2}$$

where

$$\Delta_i = C_i - \frac{S_i S_i'}{C_i'} - \frac{1}{t_i}, \tag{3}$$

with $C_i = \cos(kd_i)$, $S_i = -j\sin(kd_i)$, $C_i' = \cos(k\frac{d_i}{2})$ and $S_i' = -j\sin(k\frac{d_i}{2})$ $(i = 1, 3)$, $k = \frac{\omega}{c}\sqrt{\varepsilon}$, $F = \frac{-j\omega}{Z}$ and $j = \sqrt{-1}$. ω is the angular frequency, c the speed of light in vacuum, ε is the dielectric permittivity and Z the impedance of the waveguide. All the waveguides are characterized by the same impedance. The parameter t_i is defined as $t_i = \exp(k_B^i d_i)$, where d_i is the period of each PC (Fig. 1a and b) and k_B^i is the Bloch wavevector of each PC given by

$$\cos(k_B^i d_i) = C_i + \frac{1}{2}\frac{S_i S_{i+1}}{C_{i+1}}. \tag{4}$$

   In order to give an overview about the existence of localized Tamm states at the interface between two semi-infinite PCs (Fig. 1a), we present in Fig. 2a the projected band structure, i.e., the dimensionless frequency $\Omega = \frac{\omega d_1}{c}\sqrt{\varepsilon}$ as function of the period d_3 of the second PC for $d_1 = 1$. The stubs of the two PCs are taken identical such that $d_2 = d_4 = 0.6$. All the lengths are given in units of d_1. The shaded areas represent the bulk bands, while white areas show the common bandgaps of the two infinite PCs. Inside the common bandgaps, we present the Tamm states obtained by the maxima of the LDOS at the interface between the two systems as function of d_3 using Eq. (1). One can remark that the frequency positions of the localized states depend on the length of the segment d_3. The frequencies of the localized Tamm states decrease as d_3 increases until the corresponding branches merge into the bulk band and become resonant states. In Fig. 2b and c, we give two examples of the LDOS spectra versus Ω for two values of d_3. One can see that the Tamm states appear as well-defined peaks in the LDOS inside the common bandgaps of the two PCs, their positions depending on the length d_3. Let us mention here that the experimental measurement of the density of states can be realized through an analysis of the determinant of the scattering matrix which is related to the density of states [17].
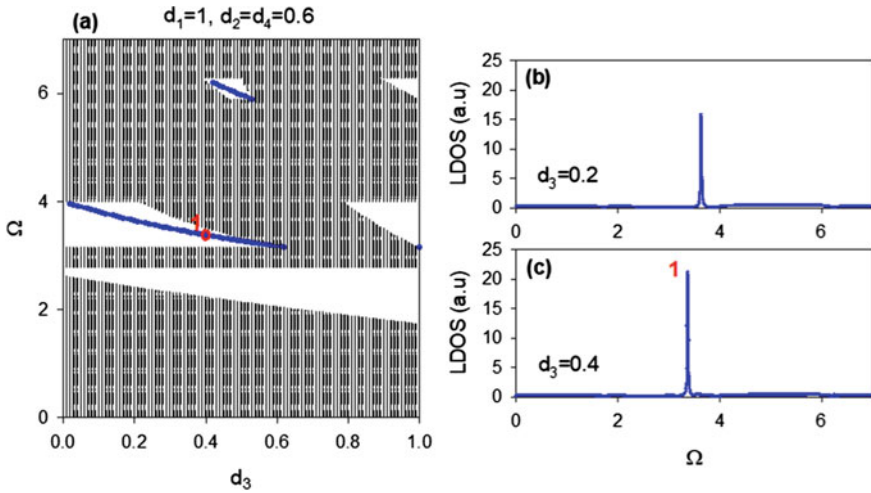
Fig. 2 **a** Frequencies of Tamm interface states between two semi-infinite PCs as function of d_3 for $d_1 = 1$, $d_2 = d_4 = 0.6$. Shaded areas represent bulk bands, while white areas show common bandgaps of the two PCs. **b** and **c** LDOS spectra versus Ω for $d_3 = 0.2$ and $d_3 = 0.4$, respectively

3 Zak Phase and Topological Tamm States at the Comb/Comb Interface

The results presented in the previous section about the existence of Tamm states at the interface between two semi-infinite PCs can be obtained by different complementary arguments. It is well known that the existence of localized Tamm states can be predicted by (i) using a topological invariant through the Zak phase θ_{Zak} of each bulk band of the 1D periodic systems [12] and (ii) calculating the reflection phases ϕ_{PC1} and ϕ_{PC2} at the two interfaces PC1/WG and PC2/WG [5]. These two quantities enable us to deduce the existence and the position of interface states by calculating the sign of ϕ_{PC1} and ϕ_{PC2} in the bandgaps surrounding the band or directly by their intersections when the condition $\phi_{PC1} + \phi_{PC2} = 0$ is fulfilled. In what follows, we will focus on the mode labeled 1 in Fig. 2a and c.

In order to illustrate to above approaches in our photonic device, we consider the interface between two different PCs consisting of alternating symmetric cells (Fig. 1b and c). The two PCs are made of the same material but with different geometrical parameters (PC1: $d_1 = 1$, $d_2 = 0.6$) and (PC2: $d_3 = 0.4$, $d_4 = 0.6$). Figure 3a gives the band structure of the infinite PC1 (red curves) and PC2 (green curves). As the two PCs possess a mirror symmetry, the corresponding Zak phase in each band can be obtained by examining the symmetry of the modulus of the electric field at the band edges states [11, 13]. The Zak phase is 0 if the two band edges have the same symmetry, otherwise it is π. For PC1 (red curves in Fig. 3a), the Zak phase of the second band is 0 since the surrounding band edges C and D have the same symmetry

with respect to the symmetry planes indicated by vertical dashed lines in Fig. 3f. On the other hand, the Zak phase of the third band is π because of the opposite symmetries of the corresponding band edges F and G (Fig. 3g). Similarly for PC2 (green curves in Fig. 3a), the Zak phase of its second band is π as the corresponding edges present different symmetries (Fig. 3i). Based on the band edges symmetries, one can deduce the specified gaps that support an interface state. It is known [11] that if two states at the lower/upper edges of the common bandgap belong to opposite types, an interface state must exist in this gap. Indeed, from the symmetries of band edges A and B (or C and E), the first common bandgap (labeled G_1 in Fig. 3a–d) does not support any interface state, whereas from the symmetries of band edges B and D (or E and F), the second common bandgap (labeled G_2 in Fig. 3a–d) must support an interface state.

The other way to determine the Zak phase of a given band is based on the sign of the reflection phase in the bandgaps surrounding this band. If the sign of the reflection phases in the bandgaps surrounding a given band shows the same sign, then the Zak phase is π, otherwise it is 0.

Figure 3b and c gives the reflection amplitudes and the corresponding phases at the PC1/WG (red curves) and PC2/WG (green curves) interfaces, respectively. For lossless system, the reflection amplitudes show unity in the gaps. The sign of the reflection phases and the obtained Zak phase of our photonic device are summarized in Table 1. As mentioned above, the position of Tamm interface states between two connected PCs can be also obtained by the condition [5] $r_{PC1}r_{PC2} = 1$ with the phase matching condition $\phi_{PC1} + \phi_{PC2} = 0$ where $r_{PC1}, r_{PC2}, \phi_{PC1}$ and ϕ_{PC2} are the reflection coefficients and their corresponding phases at the PC1/WG and PC2/WG interfaces, respectively. In Fig. 3c, we plotted the reflection phases ϕ_{PC1} and $-\phi_{PC2}$. The intersection between these two quantities in the second common bandgap (labeled G_2) gives rise to the interface state at $\Omega = 3.37$ (indicated by filled circle) since the condition $\phi_{PC1} + \phi_{PC2} = 0$ is satisfied. Figure 3d gives the transmission coefficient for two connected finite PCs, each one being composed of N=3 stubs (Fig. 1c). One can see the existence of a transmission peak in the second common bandgap at $\Omega = 3.37$ which gives a clear signature of the Tamm state at the PC1/PC2 interface.

4 Conclusion

In this paper, we have shown the existence of Tamm states at the interface of two PCs. The proposed PCs are both comb-like structures but with different geometrical parameters. First, we have found the interface states between two semi-infinite PCs by using the LDOS at the interface thanks to the Green's function element at this point. The interface states appear as maxima in the LDOS inside the common bandgaps of the two PCs. Then, we have proven the existence of such Tamm states through an analysis of a topological invariant which is based on the Zak phase of each bulk band of the 1D periodic system. The Zak phase is calculated using different methods such as the symmetry of the electric field at the band edge states, or from the sign of
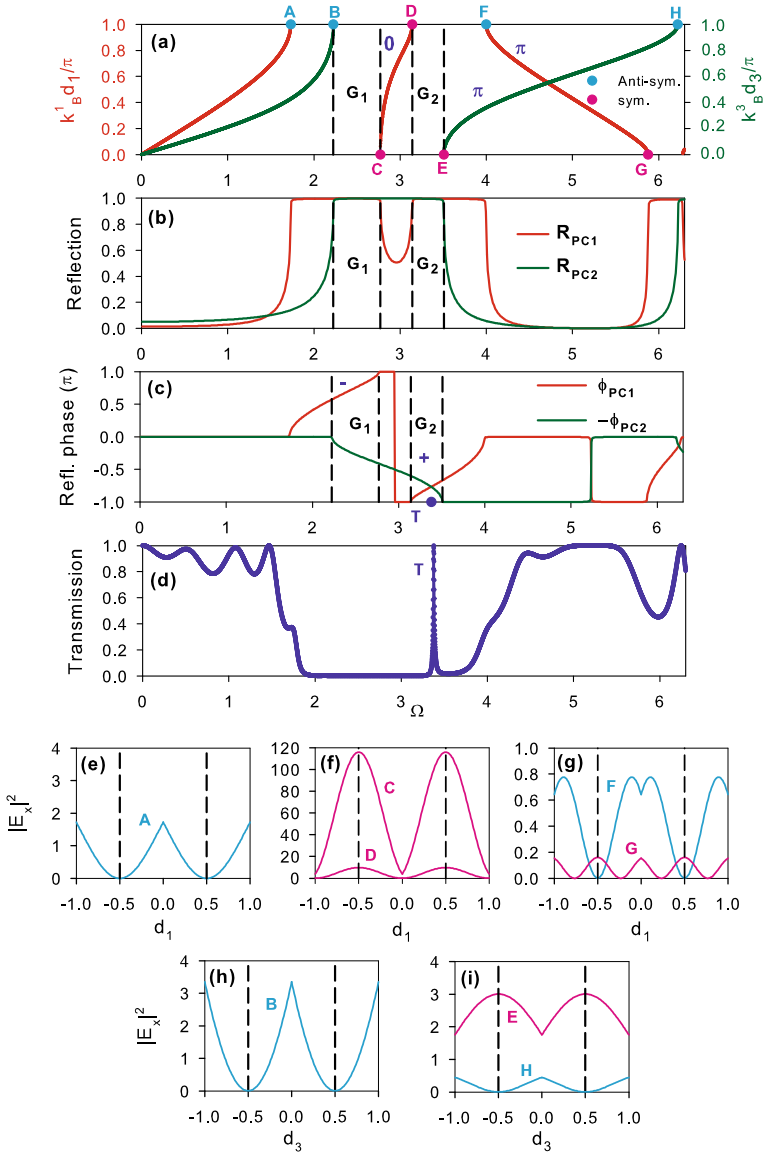
Fig. 3 **a** Band structure of the first infinite PC with a symmetric unit cell of length d_1 (red curves) and second PC with a symmetric unit cell of length d_3 (green curves). The geometric parameters of the unit cell are $d_1 = 1$, $d_2 = 0.6$ for PC1 and $d_3 = 0.4$, $d_4 = 0.6$ for PC2. G_1 and G_2 indicate the first and second common bandgaps, respectively. The vertical dashed lines in **a** indicate the limits of the common bandgaps of the two PCs. **b**, **c** Reflection amplitudes and the corresponding phases at the PC1/WG (red curves) and PC2/WG (green curves) interfaces. **d** Transmission spectra of the two connected finite PCs each one composed of $N = 3$ stubs. T indicates the interface state in the common bandgaps. **e–i** Square modulus of the electric field of the band edge states labeled A–G in (**a**)

Table 1 Zak phase of PC1 and PC2 deduced from Fig. 3c

Bandgap/band	PC1		PC2		
	Sign of reflection phase	Zak phase	Sign of reflection phase	Zak phase	Interface state
Band 3		π		−	
Bandgap 2	−		+		Yes
Band 2		0		π	
Bandgap 1	+		+		No

the reflection phases of the PC1/WG and PC2/WG. In addition, the position of the interface states can be directly obtained by the intersection of the reflection phases of the PC1/WG and PC2/WG when the condition $\phi_{PC1} + \phi_{PC2} = 0$ is fulfilled. Let us mention that it is impossible to find interface states by changing the lengths of the stubs instead of the lengths of the periods. The theoretical predictions presented here can be verified experimentally in the radiofrequency domain using coaxial cables and transposed to plasmonic nanostructures operating in terahertz frequencies [6, 15, 16]. The proposed design can be used as filter or gas sensor. This work is in progress and will be presented elsewhere.

References

1. Tamm IY (1932) Rukovoyashchie idei v tvorchestve Faradeya. Phys Z Sowjetunion 1:733
2. Khattou S, Amrani M, Mouadili A, Djafari-Rouhani B (2021) Acoustic Tamm states in slender tubes. Mater Today: Proc 45:7394–7398
3. Mei X, Ke M, He Z, Yu Z, Yu L, Liu Z (2012) Acoustic Tamm states in double 1D phononic crystals. J Wuhan Univ Technol Mater 27:374–376
4. Kavokin A, Shelykh I, Malpuech G (2005) Optical Tamm states for the fabrication of polariton lasers. Appl Phys Lett 87:261105
5. Kaliteevski M, Iorsh I, Brand S, Abram R, Chamberlain J, Kavokin A, Shelykh I (2007) Tamm plasmon-polaritons: possible electromagnetic states at the interface of a metal and a dielectric Bragg mirror. Phys Rev B 76:165415
6. Rezzouk Y, Amrani M, Khattou S, Djafari-Rouhani B (2022) Plasmonic Tamm states in periodic stubbed MIM waveguides: analytical and numerical study. J Opt Soc Am B 39:600–610
7. Braun T, Baumann V, Iff O, Hofling S, Schneider C, Kamp M (2015) Enhanced single photon emission from positioned InP/GaInP quantum dots coupled to a confined Tamm-plasmon mode. Appl Phys Lett 106:041113
8. Zaky ZA, Ahmed AM, Shalaby AS, Aly AH (2020) Refractive index gas sensor based on the Tamm state in a one-dimensional photonic crystal: theoretical optimisation. Sci Rep 10:1–9
9. Xie BY, Wang HF, Zhu XY, Lu MH, Wang ZD, Chen YF (2018) Photonics meets topology. Opt Express 26:24531–24550
10. Wang Q, Xiao M, Liu H, Zhu S, Chan CT (2016) Measurement of the Zak phase of photonic bands through the interface states of a metasurface/photonic crystal. Phys Rev B 93:041415
11. Xiao M, Zhang ZQ, Chan CT (2014) Surface impedance and bulk band geometric phases in one-dimensional systems. Phys Rev X 4:021017

12. Zak J (1989) Berry's phase for energy bands in solids. Phys Rev Lett 62:2747
13. Juneau-Fecteau A, Fréchette LG (2018) Tamm plasmon-polaritons in a metal coated porous silicon photonic crystal. Opt Mater Express 8:2774–2781
14. Gao WS, Xiao M, Chan CT, Tam WY (2015) Determination of Zak phase by reflection phase in 1D photonic crystals. Opt Lett 40:5259–5262
15. Dobrzynski L, Akjouj A, Boudouti EHE, Lêveque G, Al-Wahsh H, Pennec Y, Ghouila-Houri C, Talbi A, Djafari-Rouhani B, Jin Y (2020) Photonics. Elsevier, Amsterdam, The Netherlands
16. Khattou S, Amrani M, Mouadili A, El Boudouti EH, Talbi A, Akjouj A, Djafari-Rouhani B (2022) Three port photonic and plasmonic demultiplexers based on Cross and U-shaped stub structures: application for filtering and sensing. J Appl Phys 131:153102
17. Khattou S, Amrani M, Mouadili A, Talbi A, Akjouj A, Djafari-Rouhani B (2020) Comparison of density of states and scattering parameters in coaxial photonic crystals: theory and experiment. Phys Rev B 102:165310

High-Q Resonant Modes in Periodic Stubbed Structure

Yamina Rezzouk, Mohamed El Ghafiani, Soufyane Khattou, Madiha Amrani, El Houssaine El Boudouti, Abdelkrim Talbi, and Bahram Djafari-Rouhani

Abstract Bound states in the continuum (BICs) have unique properties and significant applications in photonics. In this paper, we show analytically and experimentally the existence of $(N-1)$ BICs (multi-BICs) in the flat band of a periodic photonic comb made of N stubs of length d_2 separated by segments of length d_1. These BICs occur when d_1 and d_2 are taken commensurate at a given frequency, which turn to quasi-BICs (multi-Fano resonances) when d_1 and d_2 are taken slightly different from the BIC position. The signature of BICs and quasi-BICs can be observed in the transmission and density of states (DOS) spectra. We show that BICs are characterized by an infinite Q factor resonances, while quasi-BICs give rise to high-Q factor resonances that grow cubically with N. Our study improves the theoretical comprehension of BICs in stubbed structure and provides useful guidelines for future applications.

Keywords BIC · Flat band · Fano resonance · High-Q factor

1 Introduction

Bound states in the continuum (BICs) were first proposed by Von Neumann and Wigner in 1929, who constructed an artificial quantum potential extending to infinity to support a BIC electronic state [1]. In 2008, the term BICs has been appeared in the optical field [2]. Moreover, optical BICs were achieved experimentally only in 2011 [3]. Principally, there are three main types of BICs [4]: BIC protected by symmetry

Y. Rezzouk (✉) · M. El Ghafiani · S. Khattou · M. Amrani · E. H. El Boudouti
LPMR, Département de Physique, Faculté des Sciences, Université Mohammed I, Oujda, Morocco
e-mail: rezzouk.yamina@ump.ac.ma

A. Talbi · B. Djafari-Rouhani
Univ. Lille, CNRS, Centrale Lille, ISEN, Univ. Valenciennes, UMR 8520 -IEMN - LIA LICS/LEMAC, Université de Lille, F-59000 Lille, France

or separability, Friedrich Wintgen BIC and Fabry Perot BIC. The unique properties of BICs have led to great interesting applications including lasers [5], sensors [6] and filters [7].

A genuine BIC is a resonant mode with an infinite radiative quality factor (Q factor) [8]. Indeed, the resonant states with high-Q factors can be achieved by perturbing slightly the structure [9] or modifying a physical parameter [10]. The phenomenon of enhancement of the Q factor was studied in a variety of periodic photonic systems such as in the arrays of rods [11], cylinders [12] and in a chain of dielectric spheres or disks [13, 14]. Moreover, the Q factor of the symmetry-protected (SP) quasi-BICs in these structures increases quadratically with the number N of cells [13] and cubically for the accidental BICs [14, 15]. Indeed, the dependence $Q(N) \sim N^2$ was established for the SP BIC theoretically [15] and experimentally [13], while the cubic dependence $Q(N) \sim N^3$ was shown for the accidental BICs. All these works have been demonstrated for the SP or accidental quasi-BICs corresponding to a finite system made of N particles. However, the study of true BICs (multi-BICs) in the flat band which transform to quasi-BICs in the nearly flat band has not been studied previously in periodic stubbed structures, and this is the aim of this article.

A few years earlier, some of us have shown that a structure made of two identical stubs of length d_2 and separated by a distance d_1 (i.e., $N = 2$) [16] can present a Fano resonance; when $d_2 = \frac{d_1}{2} + \varepsilon$ ($\varepsilon \ll 1$), this resonance transforms into a BIC when $\varepsilon = 0$. In this work, we show analytically and experimentally that when this stubbed structure is periodically repeated in N stubs, the BIC turns into $(N - 1)$ times degenerate multi-BICs. Moreover, when slightly deviating from $d_2 = \frac{d_1}{2}$, the Fano resonance transforms into multi-Fano resonances (quasi-BICs) whose Q factor is high. Furthermore, in our finite comb structure composed of N stubs, the Q factor grows cubically with N. The photonic system studied here is composed of an infinite 1D waveguide along which stubs of length d_2 are grafted periodically with spacing period d_1 (Fig. 1a). First, an infinite periodic comb is studied, where we see the existence of a flat band composed of multi-BICs for $d_2 = \frac{d_1}{2}$ and a nearly flat band by slightly detuning the stub d_2 from $\frac{d_1}{2}$. The nearly flat bands contain $(N - 1)$ Fano resonances in the transmission and the density of states (DOS) spectra of the finite comb made of N stubs (Fig. 1b). The multi-BICs appear as resonances with zero width in the transmission and DOS spectra for $d_2 = \frac{d_1}{2}$. Then, we focused on the $Q(N)$ dependence of the resonant modes as a function of the number N of stubs. The dispersion relation of bulk modes of an infinite comb structure as well as the transmission coefficient and the DOS of the finite structure are obtained in a closed form using the Green's function method [17]. The analytical results are confirmed experimentally using coaxial cables in the radio frequency regime.

The paper is organized as follows. In Sect. 2, we give analytical calculations of the dispersion relation for infinite photonic comb (Fig. 1a) and the transmission through the finite comb. In Sect. 3, we present an analytical and experimental comparative study of the transmission coefficient, as well as the DOS for the finite comb attached horizontally along a waveguide between two semi-infinite waveguides (Fig. 1b). The dependence $Q(N)$ is shown in Sect. 4. The conclusion is presented in Sect. 5.
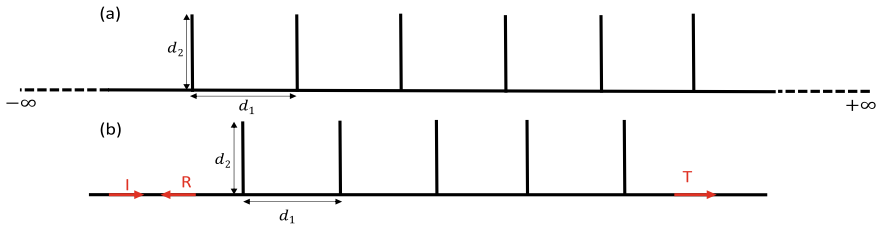
Fig. 1 **a** Infinite comb composed of stubs of length d_2 attached to a periodic waveguide of period d_1. **b** Finite comb containing $N = 5$ stubs sandwiched horizontally between two semi-infinite waveguides. I, R and T indicate the incident, reflected and transmitted waves, respectively

2 Theoretical Results

The studied comb-like structure is made of an infinite 1D waveguide along which stubs of length d_2 are grafted periodically, and separated from each other by segments of length d_1 (see Fig. 1a). The finite structure is obtained by cutting the infinite one and inserting it between two semi-infinite waveguides (Fig. 1b). Moreover, the boundary condition at the end of stubs is a vanishing magnetic field ($H = 0$). The length d_1 is chosen to be 1 m, and d_2 is taken to be the variable length for our work. The waveguides are composed of coaxial cables characterized by the complex permittivity $\varepsilon_r = 2.3 + j\varepsilon_r''$, where ε_r'' is the imaginary part of ε_r which takes into account losses in the cables, and the impedance $Z = 50\,\Omega$.

   In this section, we give an analytical study of the modes in the infinite comb (Fig. 1a) as well as in the finite one composed of N stubs. The dispersion relation of the infinite comb, as well as the transmission coefficient through a finite comb, were obtained in closed form using the Green's function method [17]. The dispersion relation for an infinite comb is given by

$$\cos(k_B d_1) = C_1 - \frac{1}{2}\frac{S_1 S_2}{C_2}, \tag{1}$$

where k_B is the Bloch wave vector, $C_i = \cos(k d_i)$, $S_i = \sin(k d_i)$ ($i = 1, 2$) and $k = \frac{\omega\sqrt{\varepsilon_r}}{c}$. ω is the angular frequency, and c is the speed of light in vacuum.

   The expression of the transmission coefficient through the finite comb of N stubs is given as follows,

$$t = \frac{2\sqrt{\xi}(2\Delta)^{N+1}}{w}C_2^N, \tag{2}$$

where

$$\Delta = 2C_1 C_2 - S_1 S_2 + j\sqrt{\xi} \quad \text{and} \quad \xi = -S_1^2 S_2^2 + 4S_1 C_2(S_1 C_2 + C_1 S_2). \tag{3}$$

$$w = j\Delta^{2N}\left\{S_1(2C_2 + S_2)^2 + 2\sqrt{\xi}(2C_2 + jS_2) + 4C_2(S_1C_2 + C_1S_2) - S_1S_2^2\right\}$$
$$- j(2C_2)^{2N}(jS_1 - C_1)^2\left\{S_1(jS_2 - 2C_2)^2 + 2\sqrt{\xi}(jS_2 - 2C_2) + 4C_2(S_1C_2 + C_1S_2) - S_1S_2^2\right\}.$$
$$\tag{4}$$

It is clear from Eq. (2) that the presence of N stubs gives rise to N transmission zeros given by $C_2 = 0$. It is well known that the eigenmodes of the finite system are given by the poles of the transmission coefficient, namely $w = 0$. However, this quantity (Eq. (4)) is complex, so it becomes difficult to simultaneously cancel its real and imaginary parts at the same frequency. However, when d_1 and d_2 are commensurate (e.g., $d_1 = 2d_2$), then w can be factored by C_2^{N-1} as follows

$$w = w'C_2^{N-1}, \tag{5}$$

where

$$w' = j(\Delta')^{2N}\left\{S_2(2C_2 + jS_2)^2 + 2\sqrt{\xi'}(2C_2 + jS_2) + S_2(-S_2^2 + 4C_2^2 + 2C_1)\right\}$$
$$- j(2jS_2C_2 - C_1)^2\left\{S_2(jS_2 - 2C_2)^2 + 2\sqrt{\xi'}(jS_2 - 2C_2) + S_2(-S_2^2 + 4C_2^2 + 2C_1)\right\} \tag{6}$$

$$\Delta' = C_1 - S_1^2 + j\sqrt{\xi'} \text{ and } \xi' = -S_2^4 + 4S_2^2C_2^2 + 2C_1S_2^2. \tag{7}$$

In this case, the transmission coefficient becomes,

$$t = \frac{8\sqrt{\xi'}(\Delta')^{N+1}C_2}{w'}. \tag{8}$$

Equation (5) clearly shows the existence of $(N-1)$ BICs which are degenerate and given by $C_2 = 0$. Further, Eq. (8) shows that at the BIC position, the transmission vanishes.

3   Numerical and Experimental Results

Figure 2 gives the projected band of the bulk modes of the infinite comb (Fig. 1a) as a function of the stub of length d_2 (i.e., the dimensionless frequency $\Omega = \frac{\omega\sqrt{\varepsilon_r}d_1}{c}$ versus $\frac{d_2}{d_1}$). The shaded areas represent bulk bands in which waves propagate into the comb, and these areas are separated by gaps (white areas). The dotted blue curves inside the passbands represent the dispersion curves obtained from the maxima of the transmission given by Eq. (2) for the finite comb composed of $N = 5$ stubs. One can see clearly a shrinking of the discrete branches and the whole band at $d_2 = 0.5d_1$ when they intercept the zero transmission branches (blue curves). Moreover, at the crossing point, there is a degeneracy of the modes (in this case $N - 1 = 4$ times)

Fig. 2 Projected band structure (the dimensionless frequency Ω vs. d_2) for the infinite comb. The blue curves show the dispersion curves obtained from the maxima of the transmission. The shaded areas represent the bulk bands
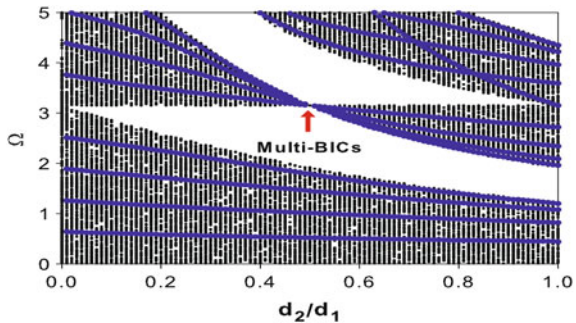
giving rise to a dispersionless flat band around $\Omega = \pi$ at $\frac{d_2}{d_1} = 0.5$ where we have multi-BICs. Indeed, by slightly deviating from the position $d_2 = 0.5d_1$, the multi-BICs transform into $(N-1)$ multi-resonances.

To give a better overview of the behavior of the discrete modes in the transmission and DOS spectra around the position $d_2 = 0.5d_1$, we have plotted in Fig. 3 the dispersion curves of the infinite comb, the transmission coefficients theoretically in the presence of loss (green dashed lines) and without loss (solid lines) as well as experimentally (red open circles) and the DOS of the finite system for some values of d_2. In Fig. 3a–c, we present the dispersion curves using Eq. (1). The red branch in Fig. 3b represents the dispersionless flat band with zero group velocity ($v_g = \frac{d\omega}{dk_B} = 0$), which occurs at $\Omega = \pi$ (i.e., $C_2 = 0$) for $d_2 = 0.5d_1$, whereas for $\frac{d_2}{d_1} \neq 0.5$ this band transforms into a dispersive nearly flat band with negative group velocity ($v_g < 0$) for $d_2 = 0.3d_1$ (Fig. 3a) and positive group velocity ($v_g > 0$) for $d_2 = 0.7d_1$ (Fig. 3c); where the slope of this nearly flat band changes sign around $d_2 = 0.5d_1$. In the finite comb made of $N = 5$ stubs (Fig. 1b), these nearly flat bands give rise to $(N-1)$ Fano resonances with high-Q factor in the transmission (Fig. 3d–f) and the DOS (Fig. 3g–i) spectra. Furthermore, the multi-BICs in the flat band (Fig. 3b) appear as resonances with zero width (i.e., infinite Q factor) in the transmission and DOS spectra for $\frac{d_2}{d_1} = 0.5$ (Fig. 3e and h). One can notice that there is a good agreement between the band structure of the infinite comb and the transmission/DOS spectra despite the finite number of stubs ($N = 5$). Also, the experimental results (red circles) agree well with the numerical simulations (dashed lines) in the transmission spectra. The details of the experimental procedure using coaxial cables and a vector analyzer operating in the radio frequency domain are given in [17].

4 $Q(N)$ Factor Dependence in the Periodic Finite Comb

In this section, we will study the dependence of the Q factor on the number N of stubs for the quasi-BICs in the nearly flat bands. This dependence can be achieved from the dependence of the resonant width γ as a function of $k_B d_1$ of the infinite comb. First,
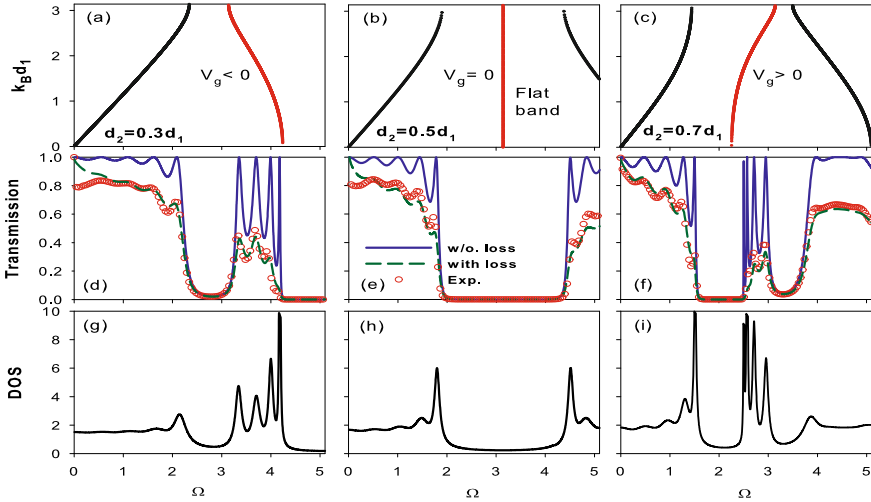
Fig. 3 Dispersion curves of the infinite comb depicted in Fig. 1a for $d_2 = 0.3d_1$ (**a**), $d_2 = 0.5d_1$ (**b**), $d_2 = 0.7d_1$ (**c**). **d–f** Transmission through the finite system made of $N = 5$ stubs in Fig. 1b. The blue solid and green dashed curves correspond to the theoretical results in a lossless and lossy system, whereas red circles correspond to the experimental results. **g–i** Same as in **d–f** but for the DOS.

we have calculated the resonant width γ corresponding to the half-width of each resonant peak in the transmission spectrum for $d_2 = 0.8d_1$ and $N = 10$. We have found that $\gamma = 0.0392(k_B d_1)^3$. Then, from the slope of $\gamma(k_B d_1)$, one can deduce the Q(N) dependence in the finite comb. Substitution of $k_B d_1$ into $\gamma(k_B d_1)$ gives rise to a simple estimation for the dependence of the Q factor of the BIC on the number N of stubs as,

$$Q = \frac{\Omega_{BIC}}{\Delta\Omega} = \frac{2.26}{0.0392(k_B d_1)^3} \simeq 1.8623N^3. \tag{9}$$

This estimation presented by the black dashed line in Fig. 4a agrees well with the results of the numerical simulation in the lossless finite comb (red circles). Furthermore, the Q factor follows the N^3 law. Therefore, these modes can be qualified as accidental BICs of Perot Fabry type. In Fig. 4b, we have plotted the Q factor of the first four resonance modes for the finite comb composed of $N = 10$ stubs and for $d_2 = 0.8d_1$, corresponding to the nearly flat band in the infinite system. As a result, we obtain the dependence $Q(N) \sim N^3$ for the various Fano resonances. Moreover, the resonance corresponding to the Γ point (i.e., $k_B d_1 = 0$) in the nearly flat band has the highest Q factor characterized by the narrowest width and the highest slope ($\alpha_1 = 1.8696$). Indeed, those resonances which are proportional to N^3 have the largest Q factor compared to other passband modes which have low Q factors due to their wide resonance widths.
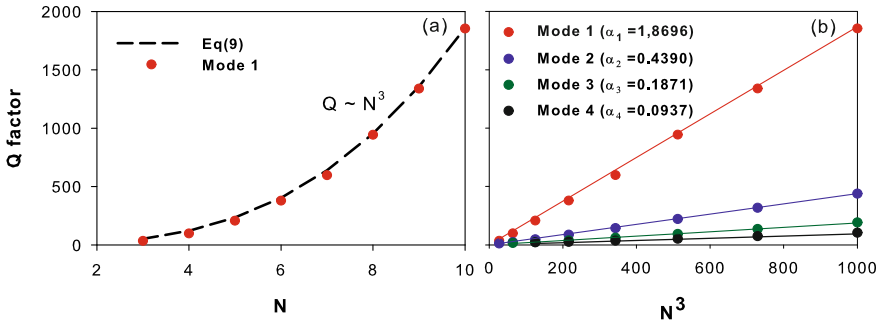
Fig. 4 **a** Dependence of the Q factor as a function of the number N of stubs for $d_2 = 0.8d_1$. The dashed line is the cubically approximation given by Eq. (9). **b** Behavior of the Q factor for various resonant modes as a function of N^3. α_i ($i = 1$–4) represents the slope of the first four resonances

5 Conclusion

In this paper, we investigated analytically and experimentally the existence of $(N-1)$ times degenerate multi-BICs in the flat band of a periodic stubbed structure made of N stubs of length d_2 separated by segments of length d_1 such that $\frac{d_2}{d_1} = 0.5$. This band transforms into a nearly flat band with $(N-1)$ quasi-BICs (multi-Fano resonances), when the stub d_2 is slightly detuned from $\frac{d_1}{2}$. In addition, the $(N-1)$ Fano resonances in the finite comb present a high-Q factor (quasi-BICs) in the nearly flat bands and become infinite (BICs) in the totally flat bands. The dependence $Q(N)$ of the quasi-BICs is cubic with N, and the proportionality between Q and N^3 depends on the mode number inside the nearly flat band. The analytical results of the DOS can be confirmed by experimental measurement of the determinant of the scattering matrix. Also, the work presented here can be transposed to the plasmonic nanoscale waveguides operating in the THz domain. These results will be presented elsewhere.

References

1. Von Neumann J, Wigner EP (1929) Über merkwürdige diskrete Eigenwerte. Z Phys 30:465–467
2. Marinica DC, Borisov AG, Shabanov SV (2008) Bound states in the continuum in photonics. Phys Rev Lett 100:183902
3. Plotnik Y, Peleg O, Dreisow F, Heinrich M, Nolte S, Szameit A, Segev M (2011) Experimental observation of optical bound states in the continuum. Phys Rev Lett 107:183901
4. Hsu CW, Zhen B, Stone AD, Joannopoulos JD, Soljacic M (2016) Bound states in the continuum. Nat Rev Mater 1:1–13
5. Hwang MS, Lee HC, Kim KH, Jeong KY, Kwon SH, Koshelev K, Kivshar Y, Park HG (2021) Ultralow-threshold laser using super-bound states in the continuum. Nat Commun 12:1–9
6. Liu Y, Zhou W, Sun Y (2017) Optical refractive index sensing based on high-Q bound states in the continuum in free-space coupled photonic crystal slabs. Sensors 17:1861

7. Foley JM, Young SM, Phillips JD (2014) Symmetry-protected mode coupling near normal incidence for narrow-band transmission filtering in a dielectric grating. Phys Rev B 89:165111

8. Hsu CW, Zhen B, Lee J, Chua SL, Johnson SG, Joannopoulos JD, Soljacic M (2013) Observation of trapped light within the radiation continuum. Nature 499:188–191

9. Koshelev K, Lepeshov S, Liu M, Bogdanov A, Kivshar Y (2018) Asymmetric metasurfaces with high-Q resonances governed by bound states in the continuum. Phys Rev Lett 121:193903

10. Yuan L, Lu YY (2018) Bound states in the continuum on periodic structures surrounded by strong resonances. Phys Rev A 97:043828

11. Taghizadeh A, Chung I-S (2017) Quasi bound states in the continuum with few unit cells of photonic crystal slab. Appl Phys Lett 111:031114

12. Yuan L, Lu YY (2017) Strong resonances on periodic arrays of cylinders and optical bistability with weak incident waves. Phys Rev A 95:023834

13. Sadrieva ZF, Belyakov MA, Balezin MA, Kapitanova PV, Nenasheva EA, Sadreev AF, Bogdanov AA (2019) Experimental observation of a symmetry-protected bound state in the continuum in a chain of dielectric disks. Phys Rev A 99:053804

14. Sidorenko MS, Sergaeva ON, Sadrieva ZF, Roques-Carmes C, Muraev PS, Maksimov DN, Bogdanov AA (2021) Observation of an accidental bound state in the continuum in a chain of dielectric disks. Phys Rev Appl 15:034041

15. Bulgakov EN, Maksimov DN (2017) Light enhancement by quasi-bound states in the continuum in dielectric arrays. Opt Express 25:14134–14147

16. Mouadili A, El Boudouti EH, Soltani A, Talbi A, Akjouj A, Djafari-Rouhani B (2013) Theoretical and experimental evidence of Fano-like resonances in simple monomode photonic circuits. J Appl Phys 113:164101

17. Dobrzynski L, Akjouj A, El Boudouti EH, Lévêque G, Al-Wahsh H, Pennec Y, Ghouila-Houri C, Talbi A, Djafari-Rouhani B, Jin Y (2020) Photonics. Elsevier

Theoretical Study of the Sensitivity of the Localized Electronic States Induced by the Presence of Defects in a ZnO/Zn$_{1-x}$Mg$_x$O MQWs Under Hydrostatic Pressure and Temperature

Abdelkader Baidri, Fatima Zahra Elamri, Farid Falyouni, Youssef Ben-Ali, and Driss Bria

Abstract In this work, we study the impact of hydrostatic pressure and temperature a multi-quantum wells (MQWs) perturbed by a barrier defect. Our study is based on the Green function method, which allows us to determine the physical properties of ZnO/Zn$_{1-x}$Mg$_x$O multi-quantum wells. The insertion of defects inside the MQWs produces localized electronic states, also called defect states which are sensitive the hydrostatic pressure and temperature. Firstly, we perturb the MQWs with a material defect where the concentration of the defect layer equals 0.15. Then we perturbed the MQWs by a geometrical defect with a thickness equal to 55 A°. Finally, we perturbed our system by a geo-material defect with a concentration and thickness equal to 0.15 and 60 A°. We found that for the case of a material and geo-material defect, the average variation of the central energy of the defect state with the hydrostatic pressure and the temperature is higher than that of a geometric defect.

Keywords MQWs · ZnO/Zn$_{1-x}$Mg$_x$O · Pressure · Temperature · Defect barrier

1 Introduction

Recently, the transport of electrons in heterostructures has gained considerable interest, making the researchers more interested in studying the impact of introducing defects in multi-quantum wells (MQWs). They found that the introduction of defects favors the transport of electrons through the multi-quantum wells using energy levels located in the bandgap, which are corresponding to the localized electronic states [1, 2]. In one of these studies, Ezzarfi et al. [3] have studied using

A. Baidri (✉) · F. Z. Elamri · F. Falyouni · Y. Ben-Ali · D. Bria
Laboratoire des Matériaux, Ondes, Energie et Environnement, Université Mohammed Premier, 60000 Oujda, Morocco
e-mail: amrou2018baidri@gmail.com

Y. Ben-Ali
Engineering Sciences Laboratory (LSI), Multidisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, B.P. 1223, Taza Gare, Morocco

the transfer matrix method the impact of the presence of a material defect on the behavior of the electronic states on the transmission spectrum of the system at the barrier on a CdTe/MnCdTe multi-quantum wells. Elamri et al. [4] have studied also the impact of a geometrical defect introduced in a GaAs/AlGaAs multi-quantum wells (MQWs), but they used the Green function method to determine the transmission of the system. These MQW structures are suitable for many applications, such as sensors, based on the displacement of the electronic states due to external perturbations (pressure, temperature, magnetic and electric fields). These external perturbations can these localized defect states moving from higher or lower energies. Segovia-Chaves et al. [5] have observed that the increase of the applied value of hydrostatic pressure decreases the dielectric constant of GaAs, which gives an additional shift of the spectrum at short wavelengths. In addition, we find Baraket et al. [6] have found that symmetric application of linear strain improves the system's temperature sensitivity, which acts as a low-temperature sensor, a transmission peak, which shifts to higher wavelength. Mehaney et al. [7] have proposed a sensor developed to monitor many gases such as CO_2 in industrial and biomedical applications. They found that the position of the resonance peak is invariant, while its intensity changes of the concentration of the gas obtained.

Our work is a theoretical study in the framework of the structure of MQWs. The principal property of these systems is the appearance of band gaps in the transmission spectrum, which allows us to direct, control and manipulate the propagation of electronic waves. The introduction of defects in periodic systems gives us the presence of localized states; these states can be used in several applications in different domains; such the design of sensors, which is based on the change of positions or intensities of localized electronic states that appear in the bandgap with the variation of physical quantities (hydrostatic pressure and temperature). In recent years, the realization of sensors is developed thanks to periodic structures; we find in particular the multi-quantum wells based on semiconductor materials. This work is organized as follows, a general introduction describing the studied structure. Then we proceed to the theoretical model based on the Green function formalism to determine the transmission of the studied system, move on to the physical parameters of the structure such the effective mass and the thicknesses of the materials which constitute the wells and the barrier, their dependence on pressure and temperature. We will then discuss the results obtained, in which we study the transmission and the band structure in the case of a defect (barrier). Finally, we end this work with a conclusion.

2 Theoretical Model

The structure studied in this work is $ZnO/Zn_{1-x}Mg_xO$ as a square periodic multi-quantum wells (Fig. 1), consisting of 10 layers ($N = 10$), wherein each layer contains two type II/VI semiconductor materials, ZnO as the wells and $Zn_{1-x}Mg_xO$ as the barrier, with equal thicknesses $d_1 = d_2 = 40\,\text{A}°$. The concentration of the magnesium used in this work is 0.25, where both materials crystallize in the same crystal structure
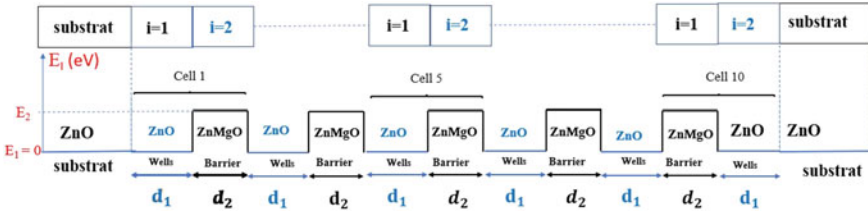
Fig. 1 Geometry of a finite MQWs limited by two substrates of the same type (ZnO)

(wurtzite) [8]. The potential energy of ZnO is given as the energy reference ($E_1 = 0$ eV), and the potential energy of $Zn_{1-x}Mg_xO$ is equal to $E_2 = Q$ ($E_g(Zn_{1-x}Mg_xO)$-$E_g(ZnO)$), where $Q = 0.7$ represents the band offset [8] and $E_g(Zn_{1-x}Mg_xO) = E_g(ZnO) + 2.51.x$ [9].

The application of the response theory, also called the Green function, has already been reviewed in elasticity and electromagnetism. Here we work with electrons in the effective mass approximation. This method is used to determine the physical properties of a finite system. In our work, we are interested in the transmission coefficient. The general approach of this method is to build finite defective MQWs limited by two substrates (semi-finite structure). This approach requires the determination of the Green function of a multi-quantum wells perturbed by the presence of defects at the structure.

The cleavage operator $\overleftrightarrow{V}(M_s, M_s)$ is written in the following form [4]:

$$\overleftrightarrow{V}(M_s, M_s) = \overrightarrow{d}^{-1}(M_m, M_m) - \overleftarrow{g}^{-1}(M_m, M_m) \tag{1}$$

where $\overleftarrow{g}^{-1}(M_m, M_m)$ is the inverse of the Green function of the infinite system, and $\overrightarrow{d}^{-1}(M_m, M_m)$ is the inverse of the Green function of a finite MQWs perturbed by a defect.

After calculating the cleavage $\overleftrightarrow{V}(M_s, M_s)$, we now pass to determining the interface response operator $\overleftrightarrow{A}(M_s, M_s)$ which is written in the following form [4]:

$$\overleftrightarrow{A}(M_s, M_s) = \sum_{M_s} \overleftrightarrow{V}(M_s, M_s).\overleftarrow{g}(M_s, M_s) \tag{2}$$

After determining the coupling operator and the interface response operator, in the interface space M_S, we are only interested in the elements of these operators in the interface space: $M_0 = (n = 0, i = 1, -\frac{d_1}{2})$, $(n = J, i = 02, -\frac{d_0}{2})$, $(n = J, i = 02, \frac{d_0}{2})$, $(n = N, i = 1, \frac{d_1}{2})$. The elements of the response function $\overleftrightarrow{d}(M_0, M_0)$ in the interface space are obtained from equation

$$\overleftrightarrow{d}(M_0, M_0) = \overleftarrow{g}(M_0, M_0)\overleftrightarrow{\Delta}^{-1}(M_0, M_0) \tag{3}$$

where: $\overleftrightarrow{\Delta}^{-1}(M_0, M_0)$ is the inverse of the operator $\overleftrightarrow{\Delta}(M_0, M_0)$:

$$\overleftrightarrow{\Delta}(M_0, M_0) = I + \overleftrightarrow{A}(M_s, M_s) \tag{4}$$

Deduce the inverse of the truncated matrix $\overleftrightarrow{d}_a(M'_0, M'_0)$ in the space $M'_0 = (0, N)$.

$$\overleftrightarrow{d}_a^{-1}(M'_0, M'_0) = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \tag{5}$$

With $B_{11}, B_{12}, B_{21}, B_{22}$ are the elements $(1, 1), (1, 4), (4, 1), (4, 4)$ of the inverse of the matrix $\overleftrightarrow{d}(M_0, M_0)$.

We then deduce the final green function of a finite MQWs containing a barrier defect, located between two substrates of the same type, which is written as follows

$$\overleftrightarrow{d}_h(M'_0, M'_0) = \frac{1}{(B_{11} - F_S)(B_{22} - F_S) - B_{21}B_{12}} \begin{bmatrix} B_{22} - F_{S2} & -B_{21} \\ -B_{12} & B_{11} - F_{S1} \end{bmatrix} \tag{6}$$

The following equation gives the transmission coefficient through the structure:

$$T = -2F_S \overleftrightarrow{d}_h(s, e) = 2F_S \frac{1}{(B_{11} - F_S)(B_{22} - F_S) - B_{21}B_{12}} B_{12} \tag{7}$$

It is well known that the application of hydrostatic pressure affects the thickness of the layers that make up the MQWs so that the variation of thickness as a function of pressure is written in the following form:

$$d(P) = d_0(1 - (S_{11} + 2S_{12})P) \tag{8}$$

where $S_{11} = \frac{(C_{11} + C_{12})}{(C_{11} - C_{12})(C_{11} + 2C_{12})}$ and $S_{12} = \frac{-C_{12}}{(C_{11} - C_{12})(C_{11} + 2C_{12})}$, d_0 is the initial thickness at $P = 0$ Kbar and P, is the hydrostatic pressure (Kbar).

Moreover, the application of hydrostatic pressure induces a decrease in the lattice parameter of the materials that constitute our system [10]. This also influences the mismatch between the two semiconductors ZnO–ZnMgO.

Similarly, the effective mass of ZnO depends on hydrostatic pressure and temperature [11, 12]:

$$\frac{m_0}{m_e^*} = 1 + \frac{E_P(E_g(P, T) + \frac{2}{3}\Delta_0)}{(E_g(P, T)((E_g(P, T) + \Delta_0))} + 2F \tag{9}$$

Here m_0 is the mass of free electrons, E_P is the energy related to the momentum matrix element, Δ_0 is the spin–orbit fractionation, F is the Kane variable, and $E_g(P, T)$ is the gap energy of ZnO which depends on temperature and pressure given by the following equation [13, 14]:

Table 1 Parameters used in our calculations for ZnO

Parameter (ZnO)	C_{11} (GPa)	C_{12} (GPa)	Δ_0 (eV)	F	E_p (eV)
Values	209	120	0.013	−2.51	28.2

$$E_g(P, T) = 3.44 + 0.0024P - 0.28 \times 10^{-5} P^2 - \frac{7.76 \times 10^{-4} T^2}{(T + 700)} \qquad (10)$$

The effective mass of the electrons in the barrier is written as [14]:

$$m_e^*(\text{ZnMgO}) = m_e^*(\text{ZnO}) + 0.05.x.m_0 \qquad (11)$$

All values for ZnO adopted in our calculation are summarized in Table 1 [15, 16].

3 Numerical Results and Discussion

3.1 The Sensitivity of the Electronic States Induced by a Material Defect to the Pressure and the Temperature Variation

The introduction of a material defect in the middle of the perfect ZnO/Zn$_{1-x}$Mg$_x$O MQWs allows us to create well localized electronic states at the 4th and the 5th gaps. As a result of this introduction, Fig. 2 represents the transmission spectrum as a function of the energy of the incident electronic wave for different values of pressure for the case at fixed parameters $d_{02} = 40$ A° and $x_{02} = 0.15$. We can notice for $P = 0$ Kbar, $T = 300$ K, the appearance of two electronic states inside the gaps, one around 483.22 meV (4th gap), with a transmission equal to $T = 0.57$, and the other one located in the middle of the bandgap around 649.65 meV (5th gap), with a transmission equal to $T = 0.7$. Moreover, we can notice that the appeared electronic states are narrow, which gives a good quality factor. Then we fix the temperature at $T = 300$ K and study the pressure effect on the electronic states that appear in the gaps. We notice that the electronic states move toward lower energies thanks to the decrease in electronic mobility, when the pressure increases. The pressure applied to the structure has a notable effect on the transmission of the electronic state, which appears around 483.22 meV as when the pressure increases, the transmission rate of the peak decreases. On the other hand, for the electronic state, which appears around 649.65 meV, its transmission remains almost stable when the pressure varies from [0–10 Kbar]. Then we deduce the sensitivity of these states; we find that the peak that appears around 649.65 meV has a sensitivity (of 0.37 meV/Kbar), which is higher than the one appeared around 483.22 meV (0.24 meV/Kbar). Similarly in Fig. 3, we fix the pressure at $P = 0$ Kbar while varying the temperature from [260–300 K]. We notice that the temperature causes a shift of the electronic state to higher energies.
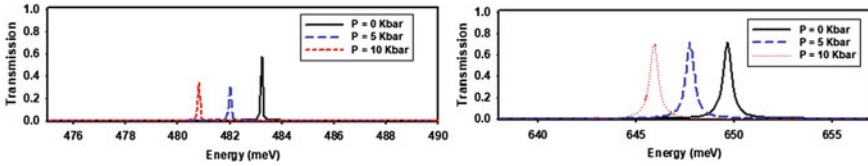
Fig. 2 Transmission spectrum of the electronic localized state induced by the presence of a material defect as a function of the energy of the electronic incident wave for different values of hydrostatic pressure
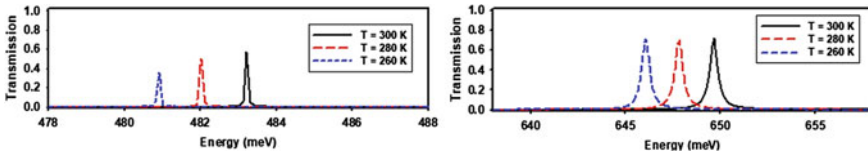
Fig. 3 Transmission spectrum of the electronic localized state induced by the presence of a material defect as a function of the energy of the electronic incident wave for different values of temperature

Moreover, we determine that the average variation of the central electronic energy with the temperature of the electronic state, which appears around 483.22 meV equals 0.0575 meV/K in a temperature range of 260–300 K. We also find that the electronic state that appears around 649.65 meV keeps the same behavior as the one that appears around 483.22 meV with an almost stable transmission rate during the temperature variation from 260 to 300 K, with the variation of the center of the electronic state energy with the temperature equal to 0.09 meV/K. Moreover, we found that for $T = 300$ K and $P = 5$ Kbar, we obtain an electronic state that is localized around an energy 647.75 meV, and with a decrease of $\Delta T = 20$ K and $\Delta P = 5$ Kbar ($T = 280$ K, $P = 0$ Kbar), the defect state remains almost in the same position around 647.85 meV. We have found that the variation of pressure and temperature gives us the possibility to control and adjust the displacement of the state inside the gap. Furthermore, applying a variation of $\Delta T = 20$ K and $\Delta P = 5$ Kbar simultaneously gives us the possibility to nearly keep the position of the state inside the gap.

3.2 The Sensitivity of the Electronic States Induced by a Geometrical Defect to the Pressure and the Temperature Variation

In the second part, we perturb our system by a geometrical defect at the 5th position ($J = 5$). We have taken the case where the proportion of magnesium in the defect layer is equal to $x_{02} = x = 0.25$, which makes $E_2 = E_{02}$, and the thickness of the defect layer is equal to $d_{02} = 55$ A°. Figure 4 represents the transmission spectrum as

a function of the energy of the incident electron wave for different pressure values. For a temperature $T = 300$ K and a zero hydrostatic pressure ($P = 0$ Kbar). We observe two defect states, one around 504.1 meV (4th gap) with a transmission rate equal to $T = 0.65$ and the other around 645.55 meV (5th gap), with a transmission rate equal to $T = 0.70$ located in the middle of the gap. Then we study the sensitivity of these electronic localized states that appear inside the gaps. It is clear from Fig. 4 that the electronic state can be controlled by the temperature and the pressure variation. This result is similarly founded for the case of a material defect. We can also observe that external perturbations (hydrostatic pressure and temperature) influence the transmission spectrum. The electronic state corresponding to the defect, around 645.55 meV (5th gap), is more sensitive to hydrostatic pressure and temperature than the state, around 504.1 meV (4th gap). We find that the sensitivity of the electronic states to the pressure variation respectively equal to 0.298 meV/Kbar and 0.12 meV/Kbar. In Fig. 5, we show the effect of temperature on the behavior of the electronic state. We notice that the electronic states keep the same behavior as they move toward higher energies. Then we can determine the sensitivity of the states that appear inside the gap for a variation of 260–300 K of temperature. The electronic localized state that appears around 645.55 meV has a sensitivity equal to 0.074 meV/K. Furthermore, around 504.1 meV (4th gap), the electronic state has a sensitivity equal to 0.03 meV/K. Therefore, we can deduce that the electronic state that appears around 645.55 meV has a higher sensitivity than the state that appears around 504.1 meV (4th gap). This defective structure can be used as a basic pressure or temperature sensor.
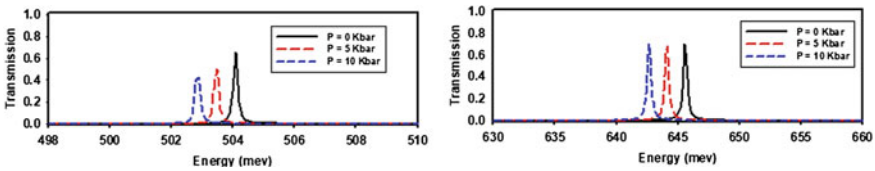
Fig. 4 Transmission spectrum of the electronic localized state induced by the presence of a geomatrical defect as a function of the energy of the electronic incident wave for different values of hydrostatic pressure
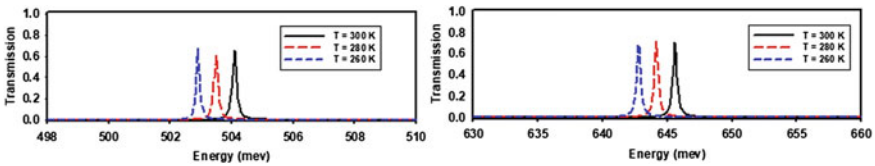
Fig. 5 Transmission spectrum of the electronic localized state induced by the presence of a geomatrical defect as a function of the energy of the electronic incident wave for different values of temperature

3.3 The Sensitivity of the Electronic States Induced by a Geo-material Defect to the Pressure and the Temperature

In the 3rd part, we study the impact of the hydrostatic pressure and the temperature on the behavior in the electronic states induced by a geo-material defect. Considering the previous results, we are interested in the 5th band gap, where an electronic state appears to be more sensitive to other states. Figure 6a represents a color map to determine the thickness and the concentration of the defect layer that allows us to obtain a state that has good transmission rate. We observe that for a value $d_{02} <$ 20 A°, the transmission rate is maximal (red regions), while it is minimal for ($x_{02} >$ 0.3 and $d_{02} > 20$ A°) (purple regions). We also observe that for the case where (d_{02} = 55 A°, x_{02} = 0.15), which corresponds to the parameters of the previous defect, there is a state with a transmission rate ($T = 0.8$). Still, it is close to the allowed band (Fig. 6b). Following this path, we seek to find the thickness and the concentration of the defect layer that allows us to obtain a state localized in the middle of the gap and which has higher transmission rate. For this, we set the concentration of the defect layer to $x_{02} = 0.15$, so that E_2 is greater than E_{02} and we search the thickness of the defect layer that corresponds to x_{02}. For this, we plot in Fig. 6b the electronic energy as a function of the thickness of the defect layer d_{02}. We observe that increasing the defect layer thickness creates a state moving towards lower energies. We also observe that for the cases, where $d_{02} = 20$ A°, $d_{02} = 40$ A°, $d_{02} = 60$ A° and d_{02} = 80 A°, the state is located in the middle of the gap. Then, we study in Fig. 7, the impact of pressure and temperature on electronic state appeared for the case (d_{02} = 60 A°, x_{02} = 0.15) around 650.55 meV (5th gap), with a transmission rate equal to 0.74. Using the optimal physical parameters of the geo-material defect (d_{02}, x_{02}), which gives us a well-defined peek inside the gap with a higher transmission rate. We observe that the central energy of the electronic state shifts to higher energies due to the decrease of the electron mobility. This result is consistent with the computational results presented in the literature [14]. Similarly, we find that the electron mobility increases with the increasing of the temperature, which leads to a noticeable shift of the localized electronic state towards higher energies (Fig. 7b). The average variation of the central energy of the electronic state with the hydrostatic pressure and temperature respectively equal to 0.37 meV/Kbar and 0.087 meV/K. We also notice that for a variation of $\Delta T = 20$ K and $\Delta P = 5$ Kbar, the electronic states are distinguished from each other, which allows us to use our structure that contains a geo-material defect as a detection device for hydrostatic pressure and temperature.

4 Conclusion

In our work, we investigate the impact of the hydrostatic pressure and the temperature on the electronic state induced by the presence of defects in a ZnO MQWs. The
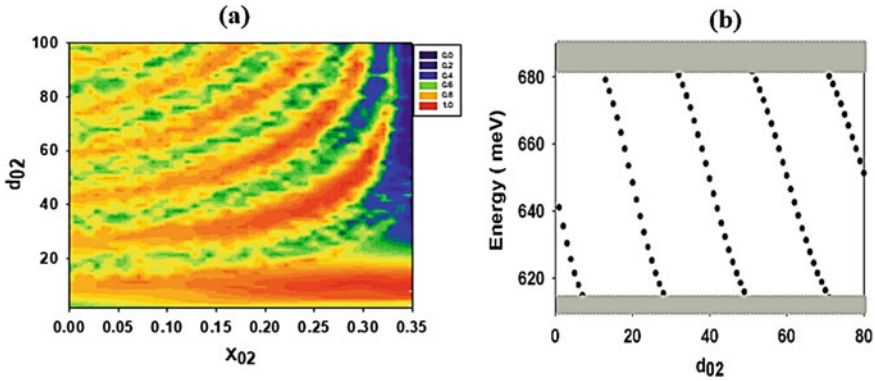
Fig. 6 **a** Color map of the transmission coefficient as functions of both x_0 and x concentrations, **b** electronic energy versus defect thickness d_{02} for a concentration equal to $x_{02} = 0.15$
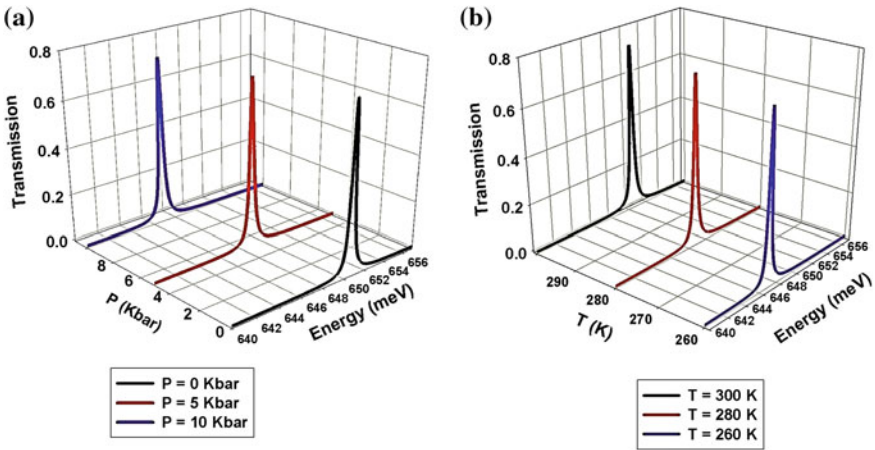
Fig. 7 Transmission spectrum of the electronic localized: **a** as a function of energy and hydrostatic pressure, **b** as a function of energy and temperature

localized electronic states appeared can be tuned and controlled by varying hydrostatic pressure and temperature. Our study is based on the Green function method, which allows us to determine the physical properties of the multi-quantum wells ZnO/Zn$_{1-x}$Mg$_x$O. In the first part, we perturb the system by material, geometrical and geo-material defects. For the three types of defects studied, we found that for a variation of 5 Kbar, the localized electronic states are directed to lower energies with a large energy shift, and for a variation of 20 K to higher energies. Furthermore, we determined that the average variation of the central energy of the defect state with the hydrostatic pressure for a material, geometric and geo-material defect is equal to 0.37 meV/Kbar, 0.298 meV/Kbar and 0.37 meV/Kbar respectively. For temperature

variation, we found central energy variation equal to 0.09 meV/K, 0.03 meV/K and 0.08 meV/K, respectively. We determined a decrease of $\Delta T = 20$ K and an increase of $\Delta P = 5$ Kbar, while the electronic state remains almost in the same position.

References

1. Ezzarfi A, Elamri FZ, Safi FZ, Bouchafra Y, Ben-Ali Y, Sali A, Bria D (2021) High quality factor multichannel filter of electrons based on defective CdMnTe/CdTe multi-quantum wells. Phys Scr 96(12):125811
2. Elamri FZ, Falyouni F, Kerkour-El Miad A, Bria D (2019) Effect of defect layer on the creation of electronic states in GaAs/GaAlAs multi-quantum wells. Appl Phys A 125(10):1–12
3. Ezzarfi A, Elsamri FZ, Bouchafra Y, Ben-Ali Y, Sali A, Bria D (2021) Localized defect states based on defective multi-quantum wells. Mater Today Proc 45:7388–7393
4. Elamri FZ, Falyouni F, Bria D (2020) Effect of a barrier defect in the creation of localized states in the structure of multi-quantum wells. Mater Today Proc 27:3101–3107
5. Segovia-Chaves F, Vinck-Posada H (2018) Simultaneous effects of the hydrostatic pressure and the angle of incidence on the defect state of a one-dimensional photonic crystal of GaAs/Ga0.7Al0.3As. Optik 164:686–690
6. Baraket Z, Zaghdoudi J, Kanzari M (2017) Investigation of the 1D symmetrical linear graded superconductor-dielectric photonic crystals and its potential applications as an optimized low temperature sensor. Opt Mater 64:147–151
7. Mehaney A, Ahmed AM (2020) Theoretical design of porous phononic crystal sensor for detecting CO_2 pollutions in air. Phys E 124:114353
8. Zan YH, Ban SL (2021) Electronic mobility limited by optical phonons in symmetric $Mg_xZn_{1-x}O$/ZnO quantum wells with mixed phases. Superlattices Microstruct 150:106782
9. Koike K, Hama K, Nakashima I, Takada GY, Ogata KI, Sasa S, Yano M (2005) Molecular beam epitaxial growth of wide band gap ZnMgO alloy films on (1 1 1)-oriented Si substrate toward UV-detector applications. J Cryst Growth 278(1–4):288–292
10. Mommadi O, El Moussaouy A, El Hadi M, Nougaoui A (2019) Excitonic properties in an asymmetric quantum dot nanostructure under combined influence of temperature and lateral hydrostatic pressure. Mater Today Proc 13:1023–1032
11. Vurgaftman I, Meyer JÁ, Ram-Mohan LÁ (2001) Band parameters for III–V compound semiconductors and their alloys. J Appl Phys 89(11):5815–5875
12. Brahim T, Bouazra A, Said M (2021) Temperature, hydrostatic pression and composition x effects on intersubband energy levels in $ZnSe/ZnS_xSe_{1-x}$ core–shell quantum dot. Optik 225:165860
13. Mang A, Reimann K (1995) Band gaps, crystal-field splitting, spin-orbit coupling, and exciton binding energies in ZnO under hydrostatic pressure. Solid State Commun 94(4):251–254
14. Jolivet A (2019) Dispositifs infrarouges à cascade quantique à base de semiconducteurs GaN/AlGaN et ZnO/ZnMgO. Thèse de doctorat. Université Paris Saclay (COmUE)
15. Chiaria S, Goano M, Bellotti E (2011) Numerical study of ZnO-based LEDs. IEEE J Quantum Electron 47(5):661–671
16. Dallali L, Jaziri S, El Haskouri J, Amorós P (2009) Optical properties of exciton confinement in spherical ZnO quantum dots embedded in SiO_2 matrix. Superlattices Microstruct 46(6):907–916

Simultaneous Effects of Hydrostatic Pressure and Temperature on the Transport of an Electron in AlGaAs Cylindrical Quantum Wire Sandwiched Between Two GaAs Cylindrical Quantum Well Wires

Mohammed Rida Qasem, Youssef Ben-Ali, Farid Falyouni, and Driss Bria

Abstract In this work, we use the Green function method to investigate the effect of hydrostatic pressure and temperature applied on the $Ga_{1-x}Al_xAs$ finite barrier cylindrical semiconductor quantum wire (CSQWR) sandwiched between two GaAs semi-infinite cylindrical semiconductor quantum well wires (CSQWWRs). This study is performed in the framework of the effective mass approximation for a finite confinement potential. The results show that the hydrostatic pressure and temperature are of significant effect on the behavior of the energy levels of the CSQWR. As result, we found that the energy levels of electrons in the band conduction decrease as the radius of the CSQWR increases. In addition, the results demonstrate that their energy levels decrease as the pressure increases while increase as the temperature increases. These variations in the energy levels of electrons, as a function of structural and external parameters, allow the production of tunable semiconductor devices.

Keywords Hydrostatic pressure · Cylindrical semiconductor quantum wire · Green function

1 Introduction

The one-dimensional (1D) nano-heterosystems consisting of cylindrical semiconductor quantum wires (CSQWRs) are interesting because they allow guiding, controlling and manipulating the electron density along the one-dimensional quantum wires. For this reason, the study of the behavior of the electrons transport in one-dimensional

M. R. Qasem (✉) · F. Falyouni · D. Bria
Laboratory of Materials, Waves, Energy and Environment, Team of Acoustics, Photonics and Materials, Faculty of Sciences, Mohamed First University, Oujda, Morocco
e-mail: qasem.mohammedrida@ump.ac.ma

Y. Ben-Ali
Engineering Sciences Laboratory (LSI), Multidisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, B.P. 1223, Taza Gare, Morocco

CSQWRs structures during the application of hydrostatic pressure and temperature is interesting in terms of the production of new tunable semiconductor devices [1, 2]. The advantages of axially combined semiconductors materials allow for countless potential applications, such as the nanoelectronics, the medical field, the mechanism of one-dimensional transport of carriers, etc. [3]. In recent years, the study of the behavior of different types of waves (light or electrons waves) in photonic or electronic systems has been the subject of several works. For example, Elamri et al. have studied the effect of hydrostatic pressure and temperature on the localized electronic defect states situated in the electronic band gaps, which are induced by a geometrical defect in a multi-quantum well. This study shows that the localized defect states are very sensitive to the pressure variation, while almost independent of the temperature variation [4]. In addition, Başer [5] has studied the effect of hydrostatic pressure and temperature on the behavior of the semiconductor structure. The results demonstrate that when the pressure applied on the system, the lattice parameters of the system, the band gap and the periodic potential of the structure change. Similarly, the electronic structure of the semiconductor changes with temperature. In the other work, Ben Ali et al. have studied the transmission spectrum and band structure of a one-dimensional photonic crystal and found that the insertion of the GaAs layer defect inside the perfect structure creates a defect mode (photonic state) in the gaps which are very sensitive to the temperature and hydrostatic pressure [6]. In addition, Safarpour et al. have investigated the effect of pressure and temperature on the electronic structure of spherical InAs quantum dots located at the center of cylindrical GaAs nanowires. This study shows that for a given value of the pressure and temperature, the energy levels of the electrons and the transition energy decrease when the dot radius increases [7]. In this paper, we present a detailed study of a one-dimensional structure consisting of a $Ga_{1-x}Al_xAs$ finite potential barrier of CSQWR located between two GaAs CSQWWRs (Fig. 1).

2 The Theory of Cylindrical Semiconductor Quantum Wire

In the present paper, we consider the one-dimensional cylindrical semiconductor quantum wire (CSQWR) characterized by the radius $R(P)$, the thickness $d_B(P)$ and the effective mass $m_{e,B}^*(P, T, x)$ along Oz axis; this system is embedded between two semi-infinite CSQWWRs.

The potential along the nanowire is given by:

$$E_i(\rho, z, P, T, x) = \begin{cases} E_B, & \rho \leq R(P) \text{ and } \frac{-d_B(P)}{2} \leq z \leq \frac{+d_B(P)}{2} \\ E_W, & \rho \leq R(P) \text{ and } -\infty \leq z \leq \frac{-d_B(P)}{2} \\ E_W, & \rho \leq R(P) \text{ and } \frac{+d_B(P)}{2} \leq z \leq +\infty \\ \infty, & \rho \geq R(P) \end{cases} \quad (1)$$
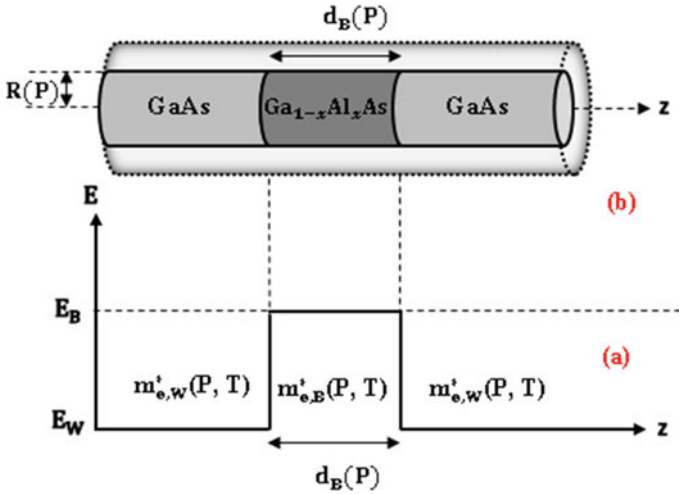
Fig. 1 Potential and geometrical representation of finite barrier CSQWR $Ga_{1-x}Al_xAs$ sandwiched between two semi-infinite GaAs nanowires

where $d_B(P)$ is the thickness of the barrier $Ga_{1-x}Al_xAs$, $R(P)$ is the radius of the nanowire. In our model, we approach the potential energy at infinity in the radial direction when $\rho \geq R(P)$. To investigate the electron quantum states in semiconductor CQWR, it is necessary to solve the stationary Schrodinger equation with the Hamiltonian in cylindrical system coordinates:

$$[E - H_i]G_i(\vec{r}, \vec{r}') = \delta(r - r') \tag{2}$$

$$H_i = -\frac{\hbar^2}{2m_i}\left(\frac{\partial^2}{\partial^2\rho} + \frac{1}{\rho}\frac{\partial}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial^2\theta} + \frac{\partial^2}{\partial^2 z}\right) + E_i(\rho, z) \tag{3}$$

where $G_i(\vec{r}, \vec{r}')$, the response function associated with the infinite semiconductor i, E is the energy and $\delta(r - r')$ is Dirac's function. The effective mass of an electron along the nanowire is given by:

$$m_i(\rho, z, P, T) = \begin{cases} m_{e,B}^*(P, T, x), & \text{for } \frac{-d_B(P)}{2} \leq z \leq \frac{+d_B(P)}{2} \\ m_{e,W}^*(P, T), & \text{for } -\infty \leq z \leq \frac{-d_B(P)}{2} \\ m_{e,W}^*(P, T), & \text{for } \frac{+d_B(P)}{2} \leq z \leq +\infty \end{cases} \tag{4}$$

The application of pressure and temperature changes the radius $R(P)$, thickness $d_B(P)$ and effective masses of finite barrier CSQWR. $m_{e,W}^*(P, T)$ and $m_{e,B}^*(P, T, x)$ correspond to conduction effective mass in the CSQWWR layer and barrier layer as a function of pressure and temperature, respectively. These are defined by Refs. [8, 9]:

$$m_{e,W}^*(P, T) = \left[1 + 7.51\left(\frac{2}{E_g(P, T)} + \frac{1}{E_g(P, T) + 0.341}\right)\right]^{-1} m_0 \qquad (5)$$

$$m_{e,B}^*(P, T, x) = m_{e,W}^*(P, T) + 0.083xm_0 \qquad (6)$$

where m_0 is free electron mass, x is the mole fraction of aluminum in $Ga_{1-x}Al_xAs$ barrier and $E_g(P, T)$ is the variation of the energy band gap for a GaAs CSQWR at Γ-point with hydrostatic pressure and temperature. This $E_g(P, T)$ takes a form as follows [9]:

$$E_g(P, T) = \left[1.519 - 5.405 \times 10^{-4}\frac{T^2}{T + 204}\right] + 1.26 \times 10^{-2}P - 3.77 \times 10^{-5}P^2 \qquad (7)$$

The potential height $E_B(\rho, z, P, T, x)$, which confines electrons in the quantum wire region, is written using band offset parameters as:

$$E_B(\rho, z, P, T, x) = Q\Delta E_g^\Gamma(x, P, T) \qquad (8)$$

where $Q = 0.658$ is the conduction band offset parameter and $\Delta E_g^\Gamma(x, P, T)$ represent the temperature and pressure difference in the band-gap energy of GaAs and $Ga_{1-x}Al_xAs$ at the Γ point. This $\Delta E_g^\Gamma(x, P, T)$ is given by this formal:

$$\Delta E_g^\Gamma(x, P, T) = 1.155x + 0.37x^2 + D(x)P + G(x)T \qquad (9)$$

The coefficients $D(x)$ and $G(x)$ are $D(x) = -1.3 \times 10^{-3}x$ eV Kbar^{-1} and $G(x) = -1.15 \times 10^{-4}x$ eV K^{-1} [10]. When the hydrostatic pressure is applied on the CSQWR system, the structure parameters such as the thickness of the barrier layer and the radius of the layers in CSQWR change [9–11]. In this situation, the radius and the thickness of CSQWR are given by:

$$R(P) = R_0[1 - 3(S_{11} + 2S_{12})P]^{1/2} \qquad (10)$$

$$d_B(P) = d_0[1 - (S_{21} + 2S_{22})P] \qquad (11)$$

where R_0 and d_0 are respectively the radius and the thickness of the material barrier CSQWR when the applied hydrostatic pressure is $P = 0$ Kbar. The elastic constants are $S_{11} = 1.16 \times 10^{-3}$ Kbar^{-1} and $S_{12} = -3.7 \times 10^{-4}$ Kbar^{-1} for GaAs, $S_{21} = (1.16 + 0.03x) \times 10^{-3}$ Kbar^{-1} and $S_{22} = (-3.7 - 0.02x) \times 10^{-4}$ Kbar^{-1} for $Ga_{1-x}Al_xAs$ [11].

3 Results and Discussion

For numerical calculations, we use the aluminum mole fraction $x = 0.3$, and the hydrostatic pressure is applied in the radial direction, where the values of P are in the range from 0 to 50 Kbar and the selected temperature range is from 0 to 450 K. To understand clearly the effect of wire radius, their thickness, the temperature and applied pressure on the finite barrier, we plot the energy levels of electrons and electronic transmission spectrum of the one-dimensional structure described in (Fig. 2).

In Fig. 2, we plot the electronic transmission spectrum as a function of the incoming electron wave energy for different pressure values ($P = 0$, 15, 30 Kbar) the temperature $T = 300$ K, the barrier thickness $d_0 = 100$ Å and the radius of the nanowire $R_0 = 50$ Å. The results show that the energy levels move to low energy by increasing the applied pressure on the system. The energy shift between the two consecutive energy levels is 120 meV when the pressure is varied between 0 and 15 Kbar, while this energy shift equals 213 meV when the pressure is varied between 15 and 30 Kbar. From our analysis, when hydrostatic pressure is applied, the following effects appear; the radius of the CSQWR and the potential height of the barrier are decreased, while the effective mass increases. Because of the latter effect, the electronic transmission spectrum shifts toward the low energies. This behavior is a direct consequence of the strong dependence of the effective mass of the electrons on the pressure.

In Fig. 3, we plot the electron transmission spectrum as a function of incoming electron wave energy for different temperatures at a fixed pressure value $P = 15$ Kbar, with barrier thickness $d_0 = 100$ Å and the radius of the quantum wire without hydrostatic pressure $R_0 = 50$ Å. By increasing the temperature between [0 and 450 K], the consecutive electronic of the energy levels shift to higher energy when the temperature increases; this energy shift decreases with increasing the temperature. We
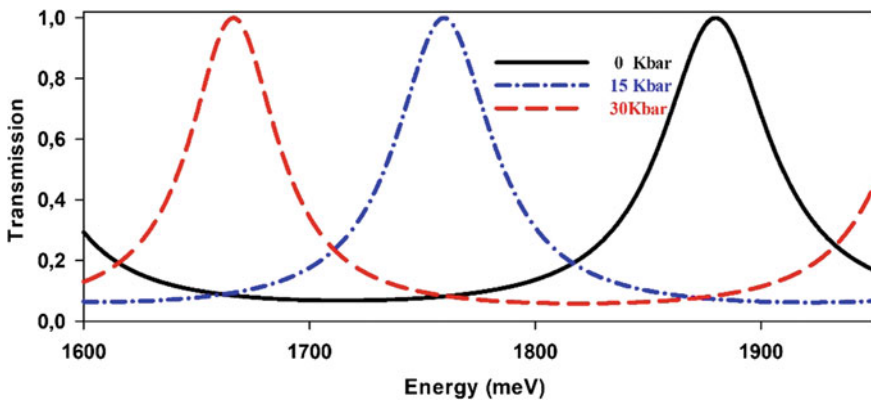
Fig. 2 Electronic transmission spectrum as a function of the energy of incoming electronic wave for different pressure values with fixed temperature $T = 300$ K, $d_0 = 100$ Å and $R_0 = 50$ Å

observe that the electronic transmission spectra overlap at 150 K. Furthermore, it is clear that as the temperature increases, the following effects appear; the effective mass of the electrons decreases, the electron potential barrier decreases, which increases the electron mobility. This behavior is a direct consequence of the strong dependence of the temperature and the electron energy levels. This last result indicates clearly that when the temperature increases, the potential barrier decreases, which shifts the electronic energy levels toward high energies with temperature. However, the increase of the effective mass is also due to the increase of the gap energy, the decrease of temperature or by the decrease of the electron mobility with the applied pressure (see Fig. 4). It is also clear that the confinement potential in the barrier decreases with increasing temperature or hydrostatic pressure.
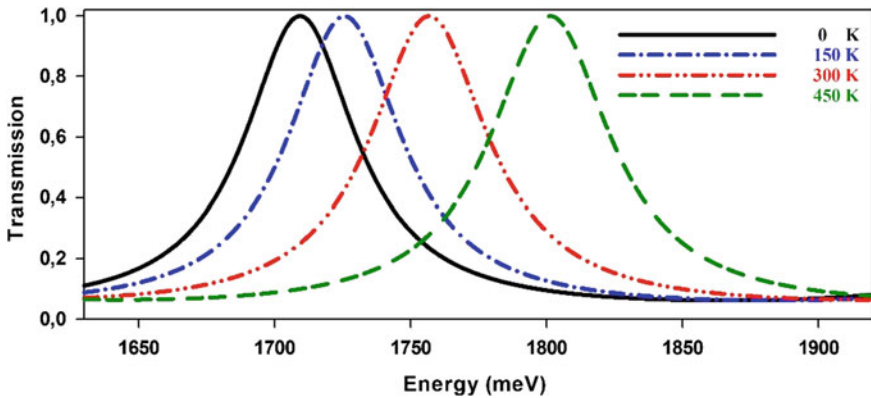
Fig. 3 Electronic transmission spectrum as a function of the energy of incoming electronic wave for different temperature at fixed $P = 15$ Kbar, $d_0 = 100$ Å and $R_0 = 50$ Å
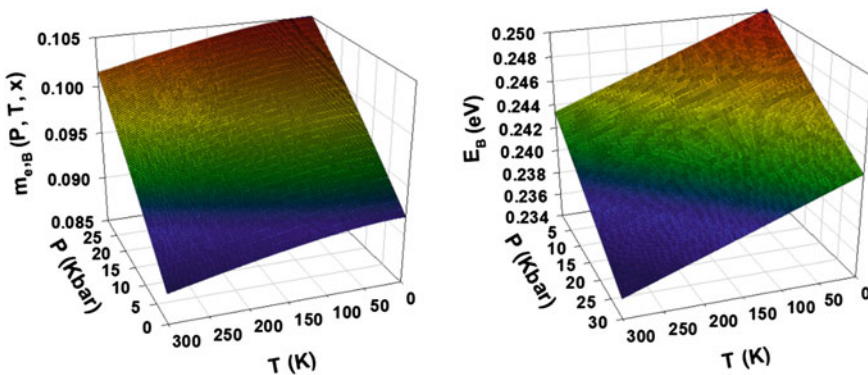
Fig. 4 Variation of effective mass $m^*_{e,B}(P, T, x)$ and barrier potential height $E_B(P, T, x)$ with hydrostatic pressure P and temperature T, at $x = 0.3$

In Fig. 5, we represent the variation of the energy levels as a function of the radius $R(P)$ of CSQWR for different hydrostatic pressure values P with fixed temperature $T = 300$ K and the finite potential barrier $d_0 = 100$ Å. According to this Fig. 5, we observe that the electron energy levels decrease with increasing pressure, due to the increasing effect of pressure on the energy gap (see Eq. 5). We note that the energy levels of the incoming electron wave increase strongly with decreasing radius of the CSQWR until a critical radius $R_c = 20$ Å where the energy levels tend to infinity when $R < R_c$. It is interesting to note that the decrease of the values of the energy levels in the conduction band with the increase of P shows that exist a strong influence of the variation of the mass of the electron and the radius of the CSQWR with P (see Eqs. 3 and 4) and the decrease of the radius of the CSQWR. This effect is very robust for quantum confinement. Moreover, from our analysis, the application of hydrostatic pressure changes the parameters of CSQWR such as the effective mass and the potential height of the barrier, which changes the energy levels of the electron in the conduction band. Then, the effect of pressure on the energy levels becomes more important in high energies. After these results, we conclude that the hydrostatic pressure confinement effect is more important than the geometric confinement effect.

Figure 6 represents the variation of the energy levels as a function of the radius $R(P)$ of the CSQWR for different temperature values T with fixed hydrostatic pressure $P = 15$ Kbar and $d_0 = 100$ Å. According to this Fig. 6, we observe that the energy levels increase with increasing temperature. From our analysis, it appears that when the temperature increases, the energy levels of electron in conduction band increase. This phenomenon is due to the dependence of the energy on the effective mass and potential height of the barrier. Then, as the radius increases, the effect of temperature on the energy levels becomes more important. At the same time, we conclude
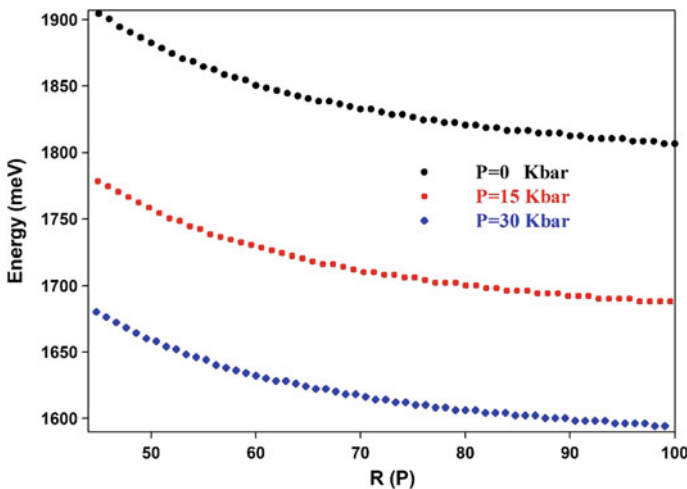
Fig. 5 Variation of the energy levels as a function of the radius $R(P)$ for three different hydrostatic pressure values P, at fixed temperature $T = 300$ K and barrier thickness $d_0 = 100$ Å
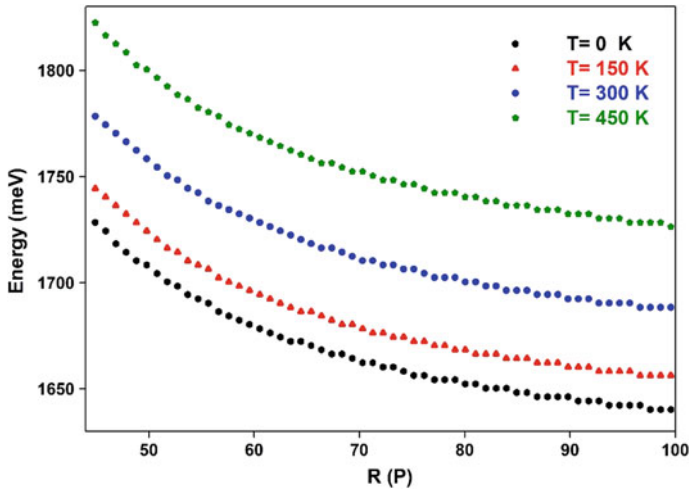
Fig. 6 Variation of the energy levels as a function of the radius of the quantum wire $R(P)$ for different temperature values T with fixed hydrostatic pressure $P = 15$ Kbar and $d_0 = 100$ Å

that when the temperature increases, the effective mass of finite barrier CSQWR decreased, which increases the mobility of the electron inside the CSQWR.

4 Conclusions

In this work, we have studied the effect of hydrostatic pressure and temperature on the behavior of the propagation of electron waves in a finite barrier of the CSQWR located between two semi-infinite CSQWWRs. We have obtained that the energy of the electron waves is inversely proportional with the hydrostatic pressure while it is proportional with the temperature. This behavior is in fact a direct consequence of the strong dependence of the electron mass on the pressure and temperature, thus increasing the electron mobility. In addition, we have demonstrated that there is a critical radius $R_c = 20$ Å of the CSQWR for quantum confinement, below which the energy levels tend to infinity. Hydrostatic pressure, temperature, aluminum mole fraction, thickness and the wire radius of the finite barrier and wire radius of the system are the most important parameters for the optimization of the CSQWR nanostructure.

References

1. Kasapoglu ESİN, Ungan FATİH, Sari HÜSEYİN, Sökmen I (2010) The hydrostatic pressure and temperature effects on donor impurities in cylindrical quantum wire under the magnetic field. Phys E 42(5):1623–1626
2. Tshipa M, Winkoun DP, Nijegorodov N, Masale M (2018) Donor impurity binding energies of coaxial GaAs/Al$_x$Ga$_{1-x}$As cylindrical quantum wires in a parallel-applied magnetic field. Superlattices Microstruct 116:227–237
3. Crouse D (2006) Phonon modes and electron-phonon interactions in cylindrical quantum wires: macroscopic and microscopic analyses for device applications. J Appl Phys 100(1):014509
4. Elamri FZ, Falyouni F, Bria D (2020) Nonlinear pressure effect on the electronic states induced by a defect layer in a multi-quantum wells structure. Mater Today Proc 31:S109–S113
5. Başer P (2021) Effect of pressure, temperature, and magnetic field on the binding energy of the electron-hole system in III-V group semiconductors. Cumhuriyet Sci J 42(2):403–412
6. Ben-Ali Y, Elamri FZ, Ouariach A, Falyouni F, Tahri Z, Bria D (2020) A high sensitivity hydrostatic pressure and temperature based on a defective 1D photonic crystal. J Electromagn Waves Appl 34(15):2030–2050
7. Safarpour G, Moradi M, Barati M (2012) Hydrostatic pressure and temperature effects on the electronic energy levels of a spherical quantum dot placed at the center of a nanowire. Superlattices Microstruct 52(4):687–696
8. Aspnes DE (1976) GaAs lower conduction-band minima: ordering and properties. Phys Rev B 14(12):5331
9. Baser PINAR, Karki HD, Demir I, Elagoz S (2013) The hydrostatic pressure and temperature effects on the binding energy of magnetoexcitons in cylindrical quantum well wires. Superlattices Microstruct 63:100–109
10. Mommadi O, El Moussaouy A, El Hadi M, Nougaoui A (2019) Excitonic properties in an asymmetric quantum dot nanostructure under combined influence of temperature and lateral hydrostatic pressure. Mater Today Proc 13:1023–1032
11. Segovia-Chaves F, Vinck-Posada H (2018) Simultaneous effects of the hydrostatic pressure and the angle of incidence on the defect mode of a one-dimensional photonic crystal of GaAs/Ga0.7Al0.3As. Optik 164:686–690

Hydrostatic Pressure and Temperature Effects on Linear and Nonlinear Optical Properties in 2D Ultra-Thin Quantum Dot

Mohammed Hbibi, Omar Mommadi, Soufiane Chouef, Reda Boussetta, Laaziz Belamkadem, Abdelaziz El Moussaouy, Juan Alejandro Vinasco, Carlos Alberto Duque, and Farid Falyouni

Abstract   In this paper, the effects of hydrostatic pressure, temperature, and quantum dot (QD) size on the exciton optical absorption coefficients (ACs) and refractive index changes (RIC) in a 2D quantum disk are studied. Our method has been carried out in the framework of effective mass approximation and two band model. The ground and subband states energies have been determined by using a variational method with a suitable test wave function and simultaneously considering the geometrical confinement and external perturbation. The results show that the optical properties depend strongly on size, pressure, and temperature. The calculations also shown that the 2D-QD size has a significant influence on the ACs and RIC and the corresponding amplitude. We have demonstrated that the increasing of the temperature (hydrostatic pressure), the ACs and RI change shift to lower and (higher) energies.

Keywords 2D quantum dot · Absorption coefficient · Refractive index changes · Hydrostatic pressure · Temperature

M. Hbibi · O. Mommadi (✉) · S. Chouef · R. Boussetta · L. Belamkadem · A. El Moussaouy · F. Falyouni
Laboratory of Materials, Waves, Energy and Environment, OAPM Group, Department of Physics, Faculty of Sciences, Mohamed I University, 60000 Oujda, Morocco
e-mail: omommadi@gmail.com

A. El Moussaouy
Laboratory of Innovation in Science, Technology and Education, Regional Centre for the Professions of Education and Training, 60000 Oujda, Morocco

J. A. Vinasco · C. A. Duque
Grupo de Materia Condensada-UdeA, Facultad de Ciencias Exactas y Naturales, Instituto de Física, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia

1 Introduction

Quantum dots (QD) are one of the most studied objects in nanophysics. In recent years, the low-dimensional structures such as QDs have been of great interest for their potential applications in microelectronics and optoelectronic devices. In these structures, electron, hole, and exciton are confined to the artificial potential of any size. To investigate the electronic and optical properties of QDs, many researchers use several methods, such as perturbative [1], variational [2–4], and exact solution methods [5].

The study of excitonic optical properties, in particular linear and nonlinear optical absorption coefficients (AC) and refractive index changes (RIC) in low-dimensional semiconductor quantum systems, have recently been the subject of intensive research. Thus, many researches have studied optical properties in various situations: with and without external disturbances such as magnetic field, electric field, pressure, and temperature. Among the most significant results, we can cite the following: Yildirim and Tomak [6] have shown that the nonlinear part of the AC associated with transitions between subbands in GaAs QD is affected by both geometric asymmetry and electric field effects. El-Bakkari et al. [7] have calculated the exciton 1s and 2s binding energies and their interband emission energy under effects of the temperature and pressure. They have shown that the binding and interband emission energies are influenced by the quantum ring size and external perturbation. In addition, in one interesting works on cylindrical QD, El Moussaouy et al. [8] have studied the effects of pressure and temperature on exciton taking into account the corrections induced by the interactions of charge carriers with both the confined LO. Also, El-Yadri et al. [9] studied the binding energy and optical transitions in quantum disks affected by the hydrostatic pressure. They have shown that the binding energy and optical properties are very sensitive to pressure. Kria et al. [10] theoretically studied the linear and nonlinear AC and RIC related to the 1s-1p transition of impurity in cylindrical core/shell QD. Their results shown that the optical response of core–shell quantum dots can be sensitively changed by changing the thickness of the shell, the impurity location, temperature and pressure.

Experimentally, the optical properties such as the cathodoluminescence and photoluminescence of $In_xGa_{1-x}N/GaN$ quantum disks have been examined by Damilano et al. [11]. They have shown that the light emission of $In_xGa_{1-x}N/GaN$ QD is blue-shifted with QD size. So far, there are no known theoretical and experimental studies on the third-order linear and nonlinear optical properties between correlated and uncorrelated states in a GaAs ultra-thin quantum disk.

In this work, we studied the third-order linear and nonlinear optical responses associated with the excitonic correlated and uncorrelated transition of the ground state in a quantum disk taking into account the influence of hydrostatic pressure, temperature, and 2D-QD size. The manuscript is organized as follows: next section presents our theoretical detail to determine the eigenstates, eigenenergies, and optical properties; our numerical results and discussions are presented in Sect. 3; and a brief conclusion is presented in the last section.

2 Framework

Despite advances in low-dimensional semiconductors, much research remains to be done before they can be used on a large scale, especially with the recent development of surface states and two-dimensional nanomaterial engineering. Ultra-thin quantum dots are characterized by atomic-scale thickness, high flexibility and multi-functionality in optoelectronic applications, which allows us to study an exciton (hole-electron pair) confined in a GaAs ultra-thin quantum disk characterized by the radius R. When its thickness d is very low compared to R ($d << 2R$), we have interested only on the lateral study after neglecting the changes of polarization induced in surface [11]. Within the effective mass approximation, under the effect of hydrostatic pressure and temperature, the basic Hamiltonian of the electron–hole system in 2D ultra-thin QD can be described approximately as:

$$
H_X = -\frac{\hbar}{2m_e^*(P,T)}\nabla_e^2 - \frac{\hbar}{2m_h^*(P,T)}\nabla_h^2 - \frac{e^2}{\varepsilon(P,T)\rho_{eh}}
$$
$$
+ V^e(\rho_e) + V^h(\rho_h), \tag{1}
$$

where $\varepsilon(P, T)$ and $m_i^*(P,T)$ are, respectively, the pressure and temperature-dependent dielectric constant and effective masses of the electron and hole with $i = (e, h)$. $\rho_{eh} = \left|\overrightarrow{\rho_e} - \overrightarrow{\rho_h}\right|$, where ρ_e, ρ_h are the electron and hole the lateral spatial coordinates. The dependence of hydrostatic pressure and temperature, in kbar and kelvin, on $m_i^*(P, T)$ and $\varepsilon(P, T)$ well described by the following relation [4]:

$$
m_e^*(P, T) = m_0\left[1 + 7.51\left(\frac{2}{E_g(P,T)} + \frac{1}{E_g(P,T) + \Delta_0}\right)\right]^{-1}, \tag{2}
$$

$$
m_h^*(P, T) = m_0\left(0.09 - 0.2 \times 10^{-3}P - 3.55 \times 10^{-5}T\right), \tag{3}
$$

and

$$
\varepsilon(P, T)
$$
$$
= \begin{cases} 12.7\exp\left(-1.67 \times 10^{-3}P\right)\exp\left(9.4 \times 10^{-5}(T - 75.6)\right) & \text{for } 0 \le T \le 200\,\text{K} \\ 13.8\exp\left(-1.73 \times 10^{-3}P\right)\exp\left(20.4 \times 10^{-5}(T - 300)\right) & \text{for } T > 200\,\text{K} \end{cases} \tag{4}
$$

where m_0 is the free electron mass, Δ_0 is the spin orbit splitting and $E_g(P, T)$ is the pressure and temperature-related band gap energy in eV, given by the following expression [5]:

$$
E_g(P, T) = 1.519 - \frac{5.4 \times 10^{-4}T^2}{T + 204} + 0.01261P + 3.77 \times 10^{-5}P^2. \tag{5}
$$

The last two terms of Eq. 1, $V_\omega^e(\rho_e)$ and $V_\omega^h(\rho_h)$ represent, respectively, the confinement potential of the electron and hole considered in our nanostructure:

$$V^i = \begin{cases} 0 & \text{if } \rho_i(P,T) \le R(P) \\ \infty & \text{otherwise} \end{cases} \quad i = (e,h) \tag{6}$$

According to the same analysis as in references [11], the orbitals of electron and hole are extended only in the xy-plane, while the Coulomb interaction is only considered in the plane of the disk. This assumption allows the use of an adiabatic approximation and separation of the Hamiltonian in two independent terms $H(\rho)$ and (z) ($[H(\rho), H(z)] = 0$). $H(z)$ represents the kinetic energy of exciton along the z-direction. The volumes with and without pressure are related by the elastic constants. Assuming that the radius of the disk follows the following relationship, the description can be summarized regardless of the nanostructure shape [12]:

$$R(P) = R(0)[1 - (S_{11} + 2S_{12})P]^{1/3} \tag{7}$$

where $R(0)$ is the 2D-QD radius without pressure. We used the reduced atomic units, for the energy we introduced the exciton Rydberg $R_{ex}^* = \hbar^2/2\mu a_X^2$, and the exciton Bohr radius $a_{ex}^* = \varepsilon\hbar^2/e^2\mu$ as unit of length, where μ represents the reduced mass $(1/\mu = 1/m_e^* + 1/m_h^*)$.

The effective Hamiltonian in Hylleraas coordinates is expressed as follows [13, 14]:

$$\begin{aligned} H_{\text{eff}} = &-\frac{1}{1+\sigma}\left[\frac{m_e^*}{m_e^*(P,T)}\right]\left[\frac{\partial^2}{\partial\rho_e^2} + \frac{1}{\rho_e}\frac{\partial}{\partial\rho_e} + \frac{\rho_{eh}^2 + \rho_e^2 - \rho_h^2}{\rho_e\rho_{eh}}\frac{\partial^2}{\partial\rho_e\partial\rho_{eh}} + \frac{1}{\rho_{eh}}\frac{\partial}{\partial\rho_{eh}}\right] \\ &-\frac{\sigma}{1+\sigma}\left[\frac{m_h^*}{m_h^*(P,T)}\right]\left[\frac{\partial^2}{\partial\rho_h^2} + \frac{1}{\rho_h}\frac{\partial}{\partial\rho_h} + \frac{\rho_{eh}^2 + \rho_h^2 - \rho_e^2}{\rho_h\rho_{eh}}\frac{\partial^2}{\partial\rho_h\partial\rho_{eh}} + \frac{1}{\rho_{eh}}\frac{\partial}{\partial\rho_{eh}}\right] \\ &-\frac{2\varepsilon(0,300)}{\varepsilon(P,T)\rho_{eh}} + V^e(\rho_e) + V^h(\rho_h) \end{aligned} \tag{8}$$

where $\sigma = \frac{m_e^*}{m_h^*}$ is the electron to hole mass ratio. In order to calculate the exciton energies and the exciton transition energy between the ground and sub-band states, so we have chosen the following trial wave functions:

$$\Psi_X^{1S}(\rho_e, \rho_h, \rho_{eh}) = J_0\left(\frac{\theta_0}{R}\rho_e\right)J_0\left(\frac{\theta_0}{R}\rho_h\right)\exp(-\alpha\rho_{eh}) \tag{9}$$

$$\Psi_X^{sb}(\rho_e, \rho_h, \rho_{eh}) = J_0\left(\frac{\theta_0}{R}\rho_e\right)J_0\left(\frac{\theta_0}{R}\rho_h\right) \tag{10}$$

where $J_0(X)$ is the zero order Bessel function, with $\theta_0 = 2.4048$. The term $\exp(-\alpha\rho_{eh})$ represents the Coulomb spatial correlation between the electron and the hole, which descript the correlated state. After obtaining the wave functions of

the correlated (Eq. 9) and uncorrelated (Eq. 10) states as well as the calculation of the excitonic energy, the first-order linear and third-order nonlinear AC can be calculated, respectively, as follows [15, 16]:

$$\alpha^{(1)}(\omega) = \omega \sqrt{\frac{\mu}{\varepsilon(P, T)}} \frac{\sigma_D \hbar \Gamma_{fi} |M_{fi}(P, T)|^2}{(E_{fi}(P, T) - \hbar\omega)^2 + (\hbar\Gamma_{fi})^2} \tag{11}$$

$$\alpha^{(3)}(\omega, I) = -\omega \sqrt{\frac{\mu}{\varepsilon(P, T)}} \left(\frac{I}{2\varepsilon_0 n_r c}\right) \frac{4\sigma_D \hbar \Gamma_{fi} |M_{fi}(P, T)|^4}{\left[(E_{fi}(P, T) - \hbar\omega)^2 + (\hbar\Gamma_{fi})^2\right]^2}$$

$$\times \left[1 - \frac{|M_{ff} - M_{ii}|^2}{4|M_{fi}(P, T)|^2} \frac{3E_{fi}(P, T)^2 - 4\hbar\omega E_{fi}(P, T) + \hbar^2(\omega^2 - \Gamma_{fi}^2)}{E_{fi}(P, T)^2 + (\hbar\Gamma_{fi})^2}\right] \tag{12}$$

where ω is the angular frequency of the incident photon, μ_1 is the magnetic permeability, ε_r is the real part of the permittivity, σ_D is the electron density related to the occupied volume $\sigma_D = V(P)$, $\hbar$ is the reduced Planck constant, Γ_{fi} is the non-diagonal matrix element $\Gamma_{fi} = 1/\tau_{fi}$ with τ_{fi} is the relaxation time and known as relaxation rate of initial and final states, $E_{fi} = E_f - E_i$ is the difference energy between the uncorrelated states $\left(E_f = E_e + E_h = 2 \times \left[\frac{\theta_0}{R}\right]^2, \text{ with } \theta_0 = 2.404\right)$ and the correlated state $\left(E_i = \frac{\langle \Psi_X^{1S}|H_X|\Psi_X^{1S}\rangle}{\langle \Psi_X^{1S}|\Psi_X^{1S}\rangle}\right)$, which equivalent to the ground states binding energy $(E_{fi} = E_B)$ that illustrated in Fig. 1a, I is the intensity of the incident electromagnetic field, ε_0 is the permittivity of free space, and n_r is the refractive index. M_{fi} is the dipole matrix element which defined by:

$$M_{fi} = e\langle \Psi_X^{1S}|\vec{\rho_e} - \vec{\rho_h}|\Psi_X^{Sb}\rangle = e\langle \Psi_X^{1S}|\rho_{eh}|\Psi_X^{Sb}\rangle \tag{13}$$

The total AC is calculated by:
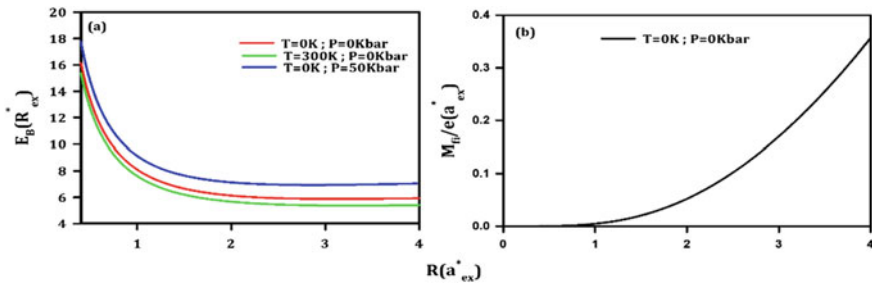
Fig. 1 **a** Energy difference between correlated and uncorrelated states (E_B) versus 2D-QD radius for different sets of hydrostatic pressures and temperatures $[P(\text{kbar}), T(\text{K})] = (0, 0), (0, 300)$ and $(50, 0)$. **b** The electric dipole transition matrix element of an exciton versus the disk radius

$$\alpha(\omega, I) = \alpha^{(1)}(\omega) + \alpha^{(3)}(\omega, I) \tag{14}$$

The linear and nonlinear of RICs are given as follows:

$$\frac{\Delta^{(1)}n(\omega)}{n_r} = \frac{1}{2\varepsilon_0 n_r^2} \frac{\sigma_D |M_{fi}(P, T)|^2 (E_{fi}(P, T) - \hbar\omega)}{(E_{fi}(P, T) - \hbar\omega)^2 + (\hbar\Gamma_{fi})^2} \tag{15}$$

$$\frac{\Delta^{(3)}n(\omega)}{n_r} = -\frac{\mu c I \sigma_D |M_{fi}(P, T)|^4}{\varepsilon_0 n_r^3}$$

$$\times \frac{(E_{fi}(P, T) - \hbar\omega)}{\left[(E_{fi}(P, T) - \hbar\omega)^2 + (\hbar\Gamma_{fi})^2\right]^2} [1$$

$$-\frac{|M_{ff} - M_{ii}|^2}{4|M_{fi}(P, T)|^2 \left(E_{fi}(P, T)^2 + (\hbar\Gamma_{fi})^2\right)}\Bigg]$$

$$\times [\{E_{fi}(P, T)(E_{fi}(P, T) - \hbar\omega)$$

$$-(\hbar\Gamma_{fi})^2 - (\hbar\Gamma_{fi})^2 \frac{(2E_{fi}(P, T) - \hbar\omega)}{(E_{fi}(P, T) - \hbar\omega)}\}\Bigg]\Bigg] \tag{16}$$

3 Results and Discussions

In this work, we investigated the influence of hydrostatic pressure, temperature and quantum confinement on optical properties related to the correlated and uncorrelated states transition of exciton confined in a GaAs ultra-thin QD. The physical parameters that are used in our calculations are presented in Table 1.

To investigate the influence of hydrostatic pressure and temperature on the optical properties, we presented in Fig. 1a the energy difference between correlated and uncorrelated states (E_B) as a function of the radius of the 2D QD for different sets [P (kbar), T (K)] = (0, 0), (0, 300), and (50, 0). We find that the quantum dot size effect for the energy difference between correlated and uncorrelated states is significant. We note that the E_B energy of the exciton increases as the QD size decreases. This is due to the increase of the Coulomb energy caused by the interaction between the electron and the hole and in other way the overlap between the electron and the hole wave functions increases with the decrease of QD radius. Moreover, our results show that for a large QD radius, the effect of hydrostatic pressure P and temperature T on the

Table 1 GaAs parameters

$m_e^* = 0.063m_0$	$m_h^* = 0.079m_0$	$\sigma = 0.77$	$\varepsilon(0, 300) = 13.18$
$E_g(0, 300) = 1.422\,\text{eV}$	$\Delta_0 = 0.341\,\text{eV}$	$I = 300\,\text{MW/m}^2$	$\sigma_D = 5 \times 10^{24}\,\text{m}^{-3}$
$S_{11} = 1.16 \times 10^{-3}\,\text{kbar}^{-1}$	$S_{12} = -3.7 \times 10^{-4}\,\text{kbar}^{-1}$	$a_{ex}^* = 19.6\,\text{nm}$	$R_{ex}^* = 2.78\,\text{meV}$

E_B energy is less sensitive than for small QDs. Moreover, the E_B energy decreases with increasing temperature for the same pressure. This is due to the variation of the mass and dielectric constant with the change of pressure. On the other hand, we find that the E_B energy increases with increasing pressure for the same temperature. Also, in Fig. 1b, we have presented the element of the dipole matrix M_{fi} as a function of the radius of the QD. The curve shows interesting results: M_{fi} increases as the QD size increases. Our results are in good agreement with those obtained in the case of impurity and exciton in the spherical quantum dot [17, 18].

Figures 2 and 3 Illustrate, respectively, the confinement effect on the absorption coefficients and the linear refractive index for three disk sizes $R = 0.8$, 0.9 and $1a_{ex}^*$ with $P = 0$ and $T = 0$. The linear, the third nonlinear and total absorption coefficients have been presented in Fig. 3a–c, respectively. We notice that the AC peak is strongly affected by the ultra-thin QD size, we know that the E_{fi} energy difference is not significant for strong pressure, as shown in Fig. 1a. Moreover, in Eq. 11, we notice that the AC amplitude is inversely proportional to the E_{fi} transition. Our results show that when the QD size decreases, the AC peak shifts to higher energy (blue shift). In Fig. 3, we have traced RIL as a function of incident photon energy, where we notice that RIL follows the same behavior as the CAL. This behavior is caused by an increase in the excitonic binding energy as the QD size is reduced.

We have plotted in Figs. 4 and 5 the ACs $\alpha^{(1)}(\omega)$, $\alpha^{(3)}(\omega, I)$ and $\alpha(\omega, I)$] and RIC, respectively, as a function of the incident photon energy for three pressure values ($P = 0$, 25 and 50 kbar) for a fixed disk radius $R = 1a_{ex}^*$ and temperature value $T = 0$ K. It can be clearly seen from Fig. 4 that the linear, nonlinear, and total terms of the absorption coefficient magnitude increase whenever the pressure increase. It is also seen that the absorption spectrum shows a blue shift with pressure. In Fig. 5a, as the pressure increases, the RIC shifts to higher energy, and their amplitude decreases slightly. The physical origin of this displacement is the decrease of the QD size with pressure and the increase in the transition energy between the correlated and uncorrelated states.
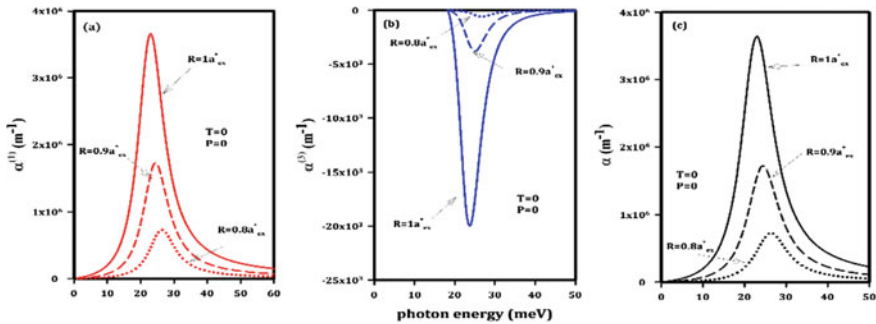
Fig. 2 Linear (**a**), third nonlinear (**b**) and total (**c**) absorption coefficients as functions of incident photon energy for $R = 0.8$, 0.9 and $1a_{ex}^*$ with $P = 0$ and $T = 0$
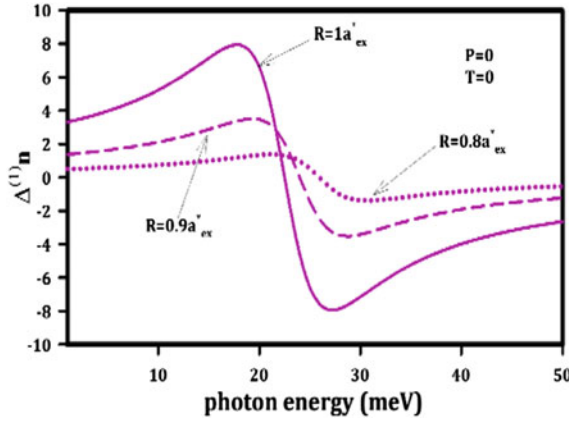
Fig. 3 Linear refractive index changes versus the incident photon energy for $R = 0.8, 0.9$ and $1a_{ex}^*$ with $P = 0$ and $T = 0$
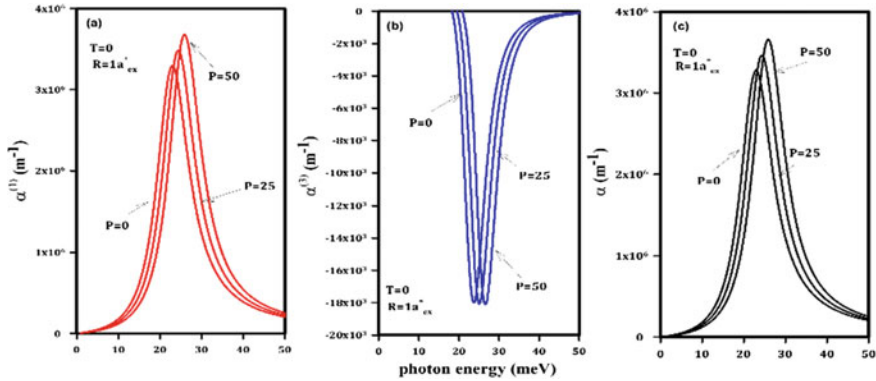
Fig. 4 Linear (**a**), third-order nonlinear (**b**), and total (**c**) optical absorption coefficients as a function of the incident photon energy for three different values of the pressure, with $T = 0$ K and $R = 1a_{ex}^*$

Figures 6 and 7 plot the ACs and RIC, respectively, versus the incident photon energy for three temperature values ($T = 0$, 200 and 300 K) for a given disk radius $R = 1a_{ex}^*$, and hydrostatic pressure value $P = 0$ kbar. Our founding shows that, when the temperature increases, the ACs peaks move to the lower energies (the red shift) with a decrease of their amplitudes. This is due to the decrease of the energy transition. Contrary to Fig. 5, Fig. 7 shown that the RIC shifts to lower energy, and their amplitude decreases slightly as the temperature increases. As the temperature increases, the electron and hole effective masses decrease, which means an increase in kinetic energy of exciton, and the dielectric constant increases, which translates into a decrease in the exciton Coulomb interaction and result in a shift toward lower energies.
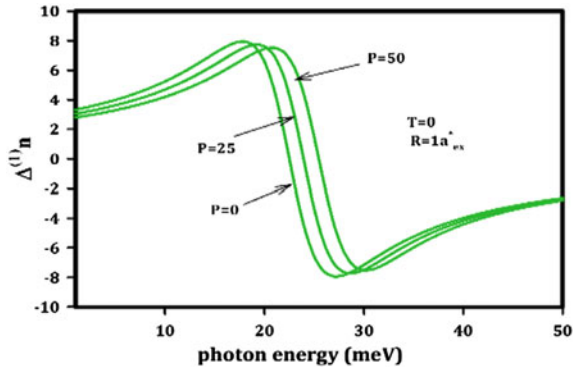
Fig. 5 The linear refractive index changes as a function of incident photon energy for three different values of the pressure ($P = 0$, 25 and 50 kbar) with $R = 1a_{ex}^*$ and $T = 0$ K
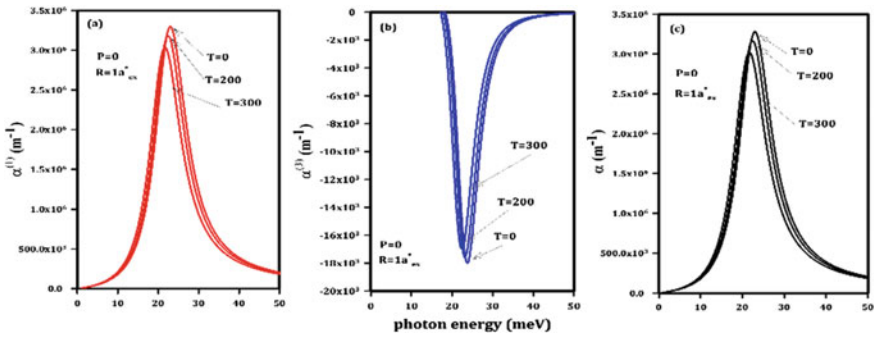
Fig. 6 Linear (**a**), third nonlinear (**b**), and total (**c**) absorption coefficients versus incident photon energy for different temperature $T = 0$, 200 and 300 K, with $P = 0$ kbar and $R = 1a_{ex}^*$
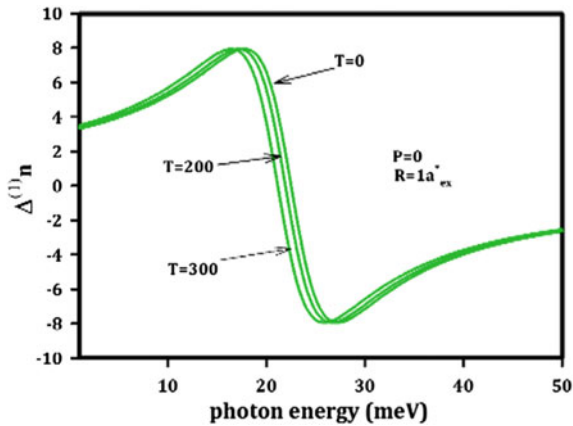
Fig. 7 The linear refractive index changes as a function of incident photon energy for three different values of temperature: $T = 0$, 200 and 300 K, with $P = 0$ kbar and $R = 1a_{ex}^*$

4 Conclusion

In conclusion, we have investigated the optical properties (the linear, nonlinear, and total absorption coefficients as well as the refractive index changes) involving the uncorrelated and correlated states transitions in ultra–thin 2D-QD with considering the influence of QD size, hydrostatic pressure, and temperature. Our results show that when the hydrostatic pressure increases, the AC and RIL curves are shifted towards higher energies (blue shift) but the increase in temperature and QD size induces a shift of the optical peaks (AC and RIL) toward lower energies (red shift). We should note that we studied an ultra-thin nanostructure that follows a circular geometry in our model, i.e., our system consists of a surface taken as a segment of a disk of almost zero thickness. In this way, our model is ideal since in general, layered systems have at least the thickness of an atom. Furthermore, in our model, we considered an infinite confinement potential, which is an excellent model to describe ultra-thin QD surrounded by vacuum or air, resulting in an infinite confinement potential that prevents charge carriers from escaping from the heterostructure to the surrounding atmosphere or vacuum region.

References

1. Gorski G, Kucab K (2019) Influence of assisted hoping interaction on the linear conductance of quantum dot. Physica E 190–200
2. Mommadi O, El Moussaouy A, Chnafi M, El Hadi M, Nougaoui A, Magrez H (2020) Exciton–phonon properties in cylindrical quantum dot with parabolic confinement potential under electric field. Physica E 113903
3. Mommadi O, El Moussaouy A, Belamkadem L (2020) Diamagnetic susceptibility of bound exciton in cylindrical quantum nanodots under hydrostatic pressure and temperature effects. Phys Scr 095809
4. Mommadi O, El Moussaouy A, El Hadi M, Chnafi M, Meziani YM, Duque CA (2020) Stark shift and exciton binding energy in parabolic quantum dots: hydrostatic pressure, temperature, and electric field effects. Philos Mag 753–775
5. Castano-Yepes JD, Amor-Quiroz DA, Ramirez-Gutierrez CF, Gomez EA (2019) Impact of a topological defect and Rashba spin-orbit interaction on the thermomagnetic and optical properties of a 2D semiconductor quantum dot with Gaussian confinement. Physica E 59–66
6. Yildirim H, Tomak M (2006) Optical absorption of a quantum well with an adjustable asymmetry. Eur Phys J B 559–564
7. El-Bakkari K, Sali A, Iqraoun E, Rezzouk A (2018) Effects of the temperature and pressure on the electronic and optical properties of an exciton in GaAs/Ga$_{1-x}$Al$_x$As quantum ring. Phys B Condens Matter 85–94
8. El Moussaouy A, Ouchani N, El Hassouani Y, Abouelaoualim D (2014) Temperature and hydrostatic pressure effects on exciton–phonon coupled states in semiconductor quantum dot. Superlattices Microstruct 22–37
9. El-Yadri M, Aghoutane N, El Aouami A, Feddi E, Zazoui M, Dujardin F, Restrepo RL, Duque CA (2017) optical transitions in strained Wurtzite GaN ultrathin quantum disk under hydrostatic pressure effects. Curr Nanosci 604–609
10. Kria M, El-Yadri M, Aghoutane N, Perez LM, Laroze D, Feddi E (2020) Forecasting and analysis of nonlinear optical responses by tuning the thickness of a doped hollow cylindrical quantum dot. Chin J Phys 444–452

11. Damilano B, Vézian VS, Portail M, Alloing B, Brault J, Courville A, Brändli BV, Leroux M, Massies J (2017) Optical properties of $In_xGa_{1-x}N$/GaN quantum-disks obtained by selective area sublimation. J Cryst Growth 262–266
12. El-Yadri M, Aghoutane N, Feddi E, Dujardin F (2017) Tunable excitonic transitions in strained GaAs ultra-thin quantum disk. Superlattices Microstruct 382–390
13. Ruiz MB (2005) Hylleraas method for many-electron atoms. I. The Hamiltonian. Int J Quantum Chem 246–260
14. Hylleraas EA (1929) Neue Berechnung der Energie des Heliums im Grundzustande, sowie des tiefsten Terms von Ortho-Helium. Z Physik 347–366
15. Gul-Kilica D, Sakiroglub S, Kasapogluc E, Sarid H, Sokmenb I (2020) Impurity-modulated optical response of a disc-shaped quantum dot subjected to laser radiation. Photonics Nanostruct 100748
16. Yang X, Xing Y (2020) The effects of built-in electric field and donor impurity on the linear and nonlinear optical properties of a wurtzite. Chin Phys B 087802
17. El Ghazi H, Jorio A, Zorkani I (2014) Linear and nonlinear intra-conduction band optical absorption in (In,Ga)N/GaN spherical QD under hydrostatic pressure. Opt Commun 73–76
18. Aghoutane N, El-Yadri M, El Aouami A, Feddi E, Dujardin F, El Haouari M, Duque CA, Nguyen CV, Phuc HV (2019) Refractive index changes and optical absorption involving 1s–1p excitonic transitions in quantum dot under pressure and temperature effects. Appl Phys A 12517

Size Effect of Spindle Toroidal Quantum Dot on Electronic Properties

Reda Boussetta, Omar Mommadi, Laaziz Belamkadem, Soufiane Chouef, Mohammed Hbibi, Abdelaziz El Moussaouy, Juan Alejandro Vinasco, Carlos Alberto Duque, and Abdelhamid Kerkour El Miad

Abstract In this work, we presented a theoretical study of the quantum confinement effects on the ground state energy of an electron trapped in a spindle toroidal quantum dot (STQD). Based on the effective mass approximation and the infinite potential, we have solved analytically the Schrödinger equation by the 3D finite difference method. Our results show that the electron energy is influenced by the confinement produced by decreasing the four geometrical parameters that describe the STQD. Furthermore, we have shown the predominance of quantum confinement is produced by small values of φ_0 and θ_0 angles on the variation of the curvature radius. On the other hand, our calculations show that the electronic energy stabilizes for large values of both φ_0 and θ_0 angles.

Keywords Spindle toroidal quantum dot · Electronic energy · 3D finite difference method

1 Introduction

At the end of the 1980s, advances in semiconductor technology made the growth of 0D nanostructures, named quantum dots (QDs). The singularity of these structures is that they are around ten nanometers in size and that the carriers are limited in three directions. The confinement quasi-particles in a nanoscale volume result in a discrete

R. Boussetta · O. Mommadi (✉) · L. Belamkadem · S. Chouef · M. Hbibi · A. El Moussaouy · A. K. El Miad
Laboratory of Materials, Waves, Energy and Environment, OAPM Group, Department of Physics, Faculty of Sciences, Mohamed I University, 60000 Oujda, Morocco
e-mail: omommadi@gmail.com

A. El Moussaouy
Laboratory of Innovation in Science, Technology and Education, Regional Centre for the Professions of Education and Training, 60000 Oujda, Morocco

J. A. Vinasco · C. A. Duque
Grupo de Materia Condensada-UdeA, Facultad de Ciencias Exactas y Naturales, Instituto de Física, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellin, Colombia

density of state distribution, which fundamentally affect the relaxation processes of photo-created electron–hole pairs found in large nanostructures. This property has led to the term "artificial atoms" being applied to these 0D nanostructures. The semiconductor QDs have been the subject of many works, both experimental and theoretical [1–5]. Several geometries of QDs (cylindrical and conical) are used in the literature to study their effect on electronic and optical properties [3–6]. In the framework of the effective mass and an infinite potential [7], in V-shaped QD, the electronic ground state energy decreases with the increase of the volume with great sensitivity of the electronic energy in small sizes of the dot [5]. The effects of temperature, external electric field, and hydrostatic pressure on the exciton confined in cylindrical QD have been examined. Their results show that the binding and the interband emission energies of the exciton depend strongly on the decrease of the temperature, the increase of the hydrostatic pressure, and the decrease of the electric field. Using the variational method and considering the effective mass approximation, the effect of temperature on the uncorrelated energy of a negatively charged trion confined in a cylindrical QD has been studied [8]. We have shown that the uncorrelated energy of the negatively charged trion is more important for the small QD size and the augmentation of the temperature leads to an increase in uncorrelated energy. So far, there is no study of the ground state energy of an electron trapped inside a spindle torus with different geometries controlled by the four geometric parameters (R_g, R_c, θ_0, and φ_0). We are therefore interested in studying the electronic energy at different sizes of spindle toroidal quantum dot (STQD).

This study aims to study theoretically the effects of quantum confinement on the electron confined in a STQD defined by two radial parameters R_C and R_g and two angular parameters θ_0 and φ_0 with a condition of $R_g > R_C$. In the first step, we consider an infinite potential and solve the Schrödinger equation analytically using the three-dimensional finite difference method. As a result, the work is structured as follows: The theoretical model is discussed in Sect. 2; the main findings of the study and their discussion are presented in Sect. 3, and the work conclusion is presented in Sect. 4.

2 Theoretical Model

Consider a single electron confined in the STQD, characterized by two internal parameters: the geometric angle θ_0 and the geometric radius R_g. The torus is specified by the radius of curvature R_C measured from the center of the torus, the angle of curvature φ_0, and geometrical angle θ_0. The STQD must always verify the condition $R_g > R_C$ (Fig. 1).

In the effective mass approximation, the Schrödinger equation for an electron can be written:

$$H_e \psi_e(r, \theta, \varphi) = E_e \psi_e(r, \theta, \varphi), \tag{1}$$
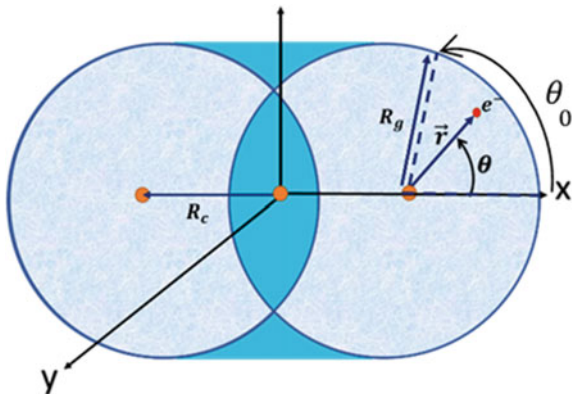
where $H_e = -\frac{\hbar^2}{2m_e^*}\nabla(r, \theta, \varphi) + V_C(r, \theta, \varphi)$ presents the Hamiltonian of the system, E_e represents the electron energy of the system and $\psi_e(r, \theta, \varphi)$ is the electron wave function. The confinement potential $V_C(r, \theta, \varphi)$ of STQD is given by:

$$V_c(r, \theta, \varphi) = \begin{cases} 0 & \text{inside STQD} \\ \infty & \text{outside STQD} \end{cases}. \tag{2}$$

The Laplace operator in toroidal coordinates has the following form [9]:

$$\Delta(r, \theta, \varphi) = \frac{\partial^2}{\partial r^2} + \frac{R_C + 2r\cos\theta}{r(R_C + r\cos\theta)}\frac{\partial}{\partial r} + \frac{1}{(R_C + r\cos\theta)^2}\frac{\partial^2}{\partial \varphi^2} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}$$
$$- \frac{\sin\theta}{r(R_C + r\cos\theta)}\frac{\partial}{\partial \theta} \tag{3}$$

To solve theoretically our system of equations, we used the finite difference method. Equation 1 is a three-dimensional second-order equation, and we discretized it on a mesh containing $N_r, *N_\theta * N_\varphi$ neouds and divide the interval $[0-R_g-R_c]$, $[0-\theta_0]$, and $[0-\varphi_0]$, into $N_r + 1$, $N_\theta + 1$ and $N_\varphi + 1$, respectively. $\frac{\partial}{\partial r}\psi(r, \theta, \varphi)$ and $\frac{\partial}{\partial \theta}\psi(r, \theta, \varphi)$ are approximated by the second-order central difference for the first derivative, and $\frac{\partial^2}{\partial r^2}\psi(r, \theta, \varphi)$, $\frac{\partial^2}{\partial \theta^2}\psi(r, \theta, \varphi)$, and $\frac{\partial^2}{\partial \varphi^2}\psi(r, \theta, \varphi)$ are approximated by the second-order central difference for the second derivative. Then, the discretization of Eq. 1 is written as:

$$-\frac{\hbar^2}{2m_e^*}\left[\left(\frac{1}{\Delta r^2} + \frac{R_C + 2i\,\Delta r\cos\theta}{2i(R_C + i\,\Delta r\cos\theta)\Delta r^2}\right)\psi_{i+1,j,k}\right.$$
$$+ \left(\frac{1}{\Delta r^2} - \frac{R_C + 2i\,\Delta r\cos\theta}{2i(R_C + i\,\Delta r\cos\theta)\Delta r^2}\right)\psi_{i-1,j,k}$$
$$+ \left(\frac{1}{(i\,\Delta r)^2(\Delta\theta)^2} - \frac{\sin\theta}{2i(R_C + i\,\Delta r\cos\theta)\Delta r\,\Delta\theta}\right)\psi_{i,j+1,k}$$

$$+ \left(\frac{1}{(i\,\Delta r)^2 (\Delta\theta)^2} + \frac{\sin\theta}{2i(R_{\mathrm{C}} + i\,\Delta r\cos\theta)\Delta r\,\Delta\theta} \right) \psi_{i,j-1,k}$$

$$+ \left(\frac{1}{(R_{\mathrm{C}} + i\,\Delta r\cos\theta)^2 \Delta\varphi^2} \right) \psi_{i,j,k+1} + \left(\frac{1}{(R_{\mathrm{C}} + i\,\Delta r\cos\theta)\Delta\varphi^2} \right) \psi_{i,j,k-1}$$

$$+ \frac{\hbar^2}{2m_{\mathrm{e}}^*} \left[\frac{2}{(i\,\Delta r)^2} + \frac{1}{(R_{\mathrm{C}} + i\,\Delta r\cos\beta)^2 \Delta\varphi^2} + \frac{1}{(i\,\Delta r)^2 (\Delta\beta)^2} \right] \psi_{i,j,k}$$

$$= E_{\mathrm{e}} \psi_{i,j,k} \tag{4}$$

3 Results and Discussions

This study aims to provide a global description of the variation of the ground state electronic energy as a function of different parameters (R_{g}, R_{c}, θ_0, φ_0). We start by Fig. 2, in which we have illustrated the energy as a function of the geometrical radius R_{g} for several values of the angle and radius of curvature (φ_0 and R_{c}). It is clear that the electronic energy decreases as the geometrical radius increases for a given angle and radius of curvature. On the other hand, an increase in energy is observed when the curvature angle decreases, keeping the constant value of the curvature radius and that of the geometric radius, which is the same behavior found with the decrease of the curvature radius when the angle of curvature and the geometric radius are fixed. These effects are more pronounced when the geometric radius is smaller. In particular, in the range of small geometrical radius, we notice that the energetic shift between two angles $\varphi_0 = 180°$ and $360°$ at $R_{\mathrm{c}} = 10\,\mathrm{nm}$ is more significant than when $R_{\mathrm{c}} = 5\,\mathrm{nm}$, which can be explained by the fact that the predominance of confinement of the curvature radius than the confinement produced by the curvature angle. In general, increasing the degrees of particle freedom produced by the increase of the four geometrical parameters leads to a decrease of the electronic energy with the increase of the STQD volume, which extend their wave function. These results are similar to these obtained in the case of the variant cylindrical QD [7].

In order to show the effect of the curvature radius R_{C}, we have illustrated in Fig. 3 the electronic energy as a function of the curvature radius, for two angle values and two geometrical radii R_{g}. We can observe a similar behavior to that obtained in Fig. 2 such as the electron energy decreases almost linearly with the augmentation of the curvature radius, and this can be explained by the fact that the relief of the electron wave function. On the other hand, for a given radius R_{c} and angle φ_0 of curvature, the augmentation of geometrical radius leads to decrease in the electron energy, and the same results are obtained when both curvature parameters (R_{c} and φ_0) increase. Moreover, we have a very interesting behavior at $R_{\mathrm{c}} = 13.5\,\mathrm{nm}$, such that the energies correspond to a given geometrical parameters ($R_{\mathrm{g}} = 20\,\mathrm{nm}$ and $\varphi_0 = 180°$) and ($R_{\mathrm{g}} = 25\,\mathrm{nm}$ and $\varphi_0 = 90°$) are equal.

Figure 4 shows the electronic energy as a function of the geometrical angle for two geometrical radius ($R_{\mathrm{g}} = 15$ and $20\,\mathrm{nm}$) and two curvature angles ($\varphi_0 = 180°$

Fig. 2 Ground state electronic energy versus the geometrical radius R_g of the STQD for two angle values ($\varphi_0 = 180°$ and $360°$) and two curvature radius values ($R_c = 5$ and 10 nm)
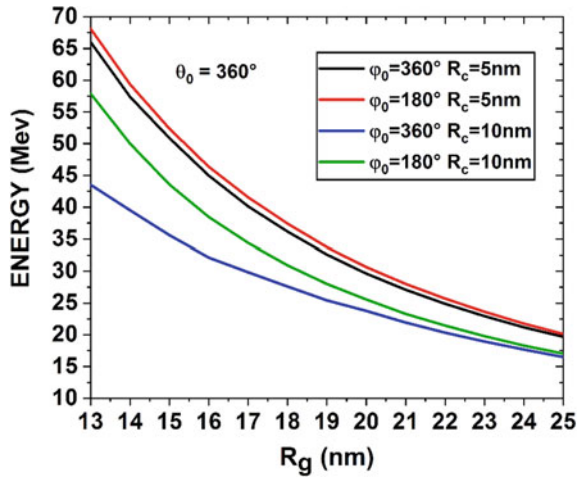
Fig. 3 Ground state electronic energy versus the curvature radius R_C of the STQD for two angle values ($\varphi_0 = 90°$ and $180°$) and two geometrical radius R_g ($R_g = 20$ and 25 nm)
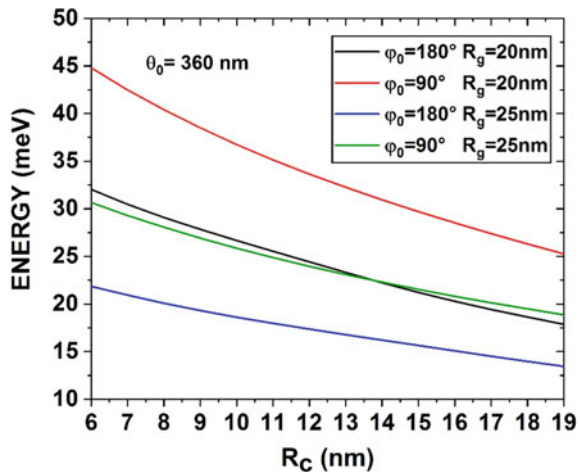
and $360°$) with $R_c = 10$ nm. This figure shows a monotonic decrease of the electronic energy with increasing geometrical angle, and the effect of the second angular parameter φ_0 is more pronounced when the quantum system tends to the QDs of the circular cross section. This is due to the predominance of the confinement effect produced by φ_0 on that produced by the geometrical angle θ_0. In contrast, geometrical radius effect is more important when their angle tend to small value, due to the combined confinement effect produced by geometrical parameters (R_g and θ_0).

In Fig. 5, we have plotted the electronic ground state energy as a function of the curvature angle for two geometrical angle and two curvature radius with $R_g = 15$ nm. The results that we have found confirm the results that obtained in Figs. 2, 3 and 4, such as the decrease of the curvature angle φ_0 leads to a decrease in the STQD volume.

Fig. 4 Electronic energy as a function of the geometrical angle θ_0 for two geometrical radius ($R_g = 15$ and $20\,\text{nm}$) and two curvature angles ($\varphi_0 = 180°$ and $360°$) with $R_c = 10\,\text{nm}$
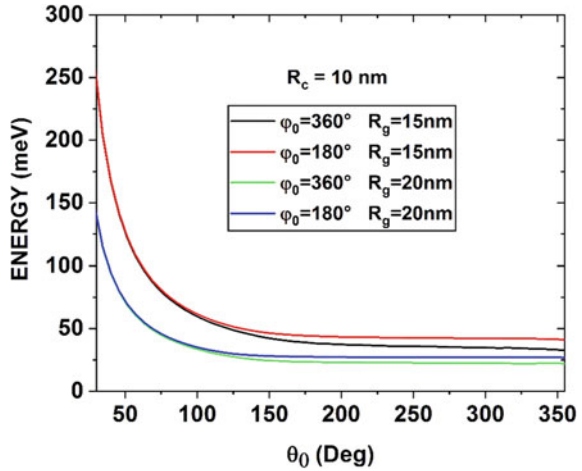
Consequently, the electronic energy grows, this is quite clear in the range [30–100] of φ_0. On the other hand, at a range [100–360], the electronic energy is less sensitive by the variation of curvature angle, which can be explained that the electronic probability density does not affect it more. Moreover, at a fixed curvature angle, the increase of curvature radius or the geometrical angle decreases the electronic energy. In our opinion, the theoretical study of the ground state electronic energy in STQD taking into consideration the impact of different geometrical parameters (R_g, R_c, θ_0 and φ_0) will lead to a better tuning the nanofabrication of nanostructure systems.

Fig. 5 Electron energy as a function of the curvature angle φ_0 for two geometrical angle ($\theta_0 = 180°$ and $360°$) and two curvature radius ($R_c = 15$ and $20\,\text{nm}$). Results are with a fixed value of the geometrical radius at $R_g = 15\,\text{nm}$
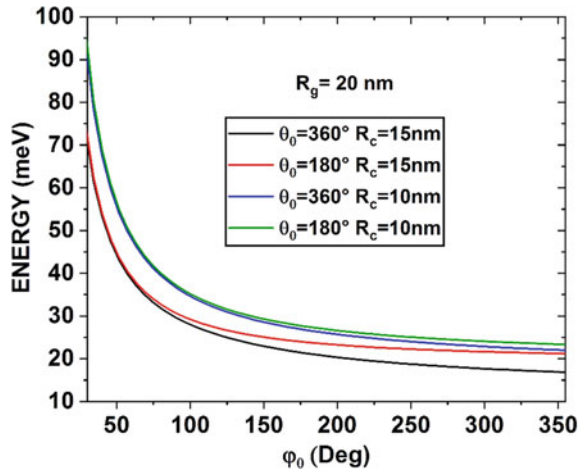
4 Conclusion

By using the effective mass approximation and in infinite confinement potential, we have solved the 3D Schrödinger equation of an electron confined in a GaAs STQD by the FDM. We have shown that a significant sensitivity of the electron ground state energy to the two radial parameters (R_c and R_g) and two angular parameters (θ_0 and φ_0). As we have already seen, the electron energy shows a decreasing behavior when the size of the quantum dot becomes larger, either by increasing both angles and radius. Moreover, we have shown that for large values of the angles, the influence of the radius R_c and R_g is less sensitive than in small values of the angles θ_0 and φ_0. We believe that this paper will contribute to a better understanding of the impact of different geometries on the nanofabrication of these types of nanostructures.

References

1. Fedorova KA (2018) Near-IR InAs/GaAs quantum-dot lasers and their application for efficient frequency conversion. In: International conference laser optics (ICLO), pp 141–141
2. Belamkadem L, Mommadi O, Boussetta R, Chnafi M, Vinasco JA, Laroze D, Pérez LM, El Moussaouy A, Meziani YM, Kasapoglu E, Tulupenko V, Duque CA (2022) First study on the electronic and donor atom properties of the ultra-thin nanoflakes quantum dots. Nanomaterials 12:966
3. Chnafi M, Belamkadem L, Mommadi O, Boussetta R, El Hadi M, El Moussaouy A, Duque CA (2021) Hydrostatic pressure and temperature effects on spectrum of an off-center single dopant in a conical quantum dot with spherical edge. Superlattices Microstruct 159:107052
4. Mommadi O, El Moussaouy A, Belamkadem L (2020) Diamagnetic susceptibility of bound exciton in cylindrical quantum nanodots under hydrostatic pressure and temperature effects. Phys Scr 95(9):095809
5. Mommadi O, El Moussaouy A, El Hadi M, Chnafi M, Meziani YM, Duque CA (2021) Stark shift and exciton binding energy in parabolic quantum dots: hydrostatic pressure, temperature, and electric field effects. Phil Mag 101(6):753–775
6. Mommadi O, El Moussaouy A, Chnafi M, El Hadi M, Nougaoui A, Magrez H (2020) Exciton–phonon properties in cylindrical quantum dot with parabolic confinement potential under electric field. Phys E 118:113903
7. Belamkadem L, Mommadi O, Vinasco JA, Laroze D, El Moussaouy A, Chnafi M, Duque CA (2021) Electronic properties and hydrogenic impurity binding energy of a new variant quantum dot. Phys E 129:114642
8. Chnafi M, El Moussaouy A, Mommadi O, Belamkadem L (2020) Energy and stability of negatively charged trion in cylindrical quantum dot under temperature effect. Phys B 594:412333
9. Boussetta R, Mommadi O, Belamkadem L, Chouef S, Hbibi M, El Moussaouy A, Vinasco JA, Duque CA, Satori H, Kenfack-Sadem C, Keumo Tsiaze RM, Fobasso Mbognou FC, El-Miad AK (2022) Deformation and size effects on electronic properties of toroidal quantum dot in the presence of an off-center donor atom. Micro Nanostruct 207209

Impact of Electric Field Strength on the Binding Energy of an Off-Center Donor in Quantum Ring: Quarter Cross Section Case

Soufiane Chouef, Omar Mommadi, Reda Boussetta, Mohammed Hbibi, Laaziz Belamkadem, Abdelaziz El Moussaouy, Juan Alejandro Vinasco, Carlos Alberto Duque, and Farid Falyouni

Abstract Using the finite difference method within the effective mass approximation, we have calculated the binding energy in a GaAs circular quantum ring with a quarter-circle cross section. In the presence of an off-center donor atom, under the contribution of electric field directed along the *x*-axis and assuming an infinite confinement potential, the binding energy of electron-impurity is reported as a function of the applied electric field and the position of the donor impurity. We have shown that the impurity binding energy depends strongly on the quantum confinement, the applied electric field as well as the displacement of the impurity inside the quantum ring. The binding energy decreases with electric field when the impurity position is in the same direction of electric field and the opposite behavior when it moves in a direction opposite to that of the electric field.

Keywords Quantum ring · Electric field · Off-center impurity · Binding energy

1 Introduction

Nanoscale semiconductor structures such quantum dots (QDs) and quantum rings (QRs) are currently hot topics due to their potential use in optoelectronic device and fabrication design [1]. The invention of recent nanofabrication techniques, including electron beam lithography, chemical vapor deposition of organometallic compounds

S. Chouef · O. Mommadi (✉) · R. Boussetta · M. Hbibi · L. Belamkadem · A. El Moussaouy · F. Falyouni
Laboratory of Materials, Waves, Energy and Environment, OAPM Group, Department of Physics, Faculty of Sciences, Mohammed I University, 60000 Oujda, Morocco
e-mail: omommadi@gmail.com

A. El Moussaouy
Laboratory of Innovation in Science, Technology and Education, Regional Centre for the Professions of Education and Training, 60000 Oujda, Morocco

J. A. Vinasco · C. A. Duque
Grupo de Materia Condensada-UdeA, Facultad de Ciencias Exactas y Naturales, Instituto de Física, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia

[2], and molecular beam epitaxy which are the techniques for producing arrays of semiconductor nanostructures of different geometries [3]. In addition, it is well known that the presence of impurity in nanostructures significantly alters the opto-electronic properties of nanostructures. In Refs. [4–12], the effect of confinement on the excitonic, optical properties and electronic states in low-dimensional nanostructures have been studied. Under the effect of external perturbations, such as hydro-static pressure and temperature on the optoelectronic characteristics of QRs have been studied [13, 14], In addition, the effect of the electric field has been the subject of many experimental and theoretical reports, due to the importance of these systems in the development of new semiconductor devices [5–10]. These works show that the application of an external electric field produces significant changes in the optical absorption response and the stability of the quantum systems.

For this way, we propose a new examination of the spectrum related to a conduction electron coupled to an off-center donor atom confined in a GaAs circular quantum ring (CQR) with a quarter-circle cross section, taking into account the electric field factor. We discuss in detail the effect of this field and the effect of the donor position as well as the geometrical parameters of CQR on the electron-impurity binding energy. The solution of the Schrödinger equation, which depends on the effective mass approximation, is solved by the finite difference method. We have organized this paper as follows: in the next section, we describe the theoretical framework. Section 3 presents the results and discussion, and finally, the remarkable conclusions are summarized in Sect. 4.

2   Theoretical Frameworks

In this work, we consider the electron motion of the conduction band in CQR with a quarter-circle cross section configuration, under the effect of an electric field F directed toward the x-axis (see Fig. 1). The quantum system is described by four parameters: $(i) R_C$: is the torus radius that controls the internal circular hole of our nanostructure along the x–y plane, (ii) R_g is the transversal cross-section radius that controls the internal circular radius of our nanostructure, (iii) $\varphi_0 = 2\pi$ is azimuthal torus angle of CQR, and (iv) θ_0 generate the quarter-circle cross section which is fixed at $\frac{\pi}{2}$.

The electron coordinates (ρ, θ, φ) in CQR with a quarter-circle cross section are limited as follow: $R_C - R_g < \rho < R_C$, $0 < \varphi < 2\pi$ and $0 < \theta < \frac{\pi}{2}$. The cartesian coordinates of electron related to the toroidal coordinates, which are presented in Fig. 1a, b, are given by:

$$\begin{cases} x(\rho, \theta, \varphi) = (R_C - \rho \cos(\theta)) \cos(\varphi) \\ y(\rho, \theta, \varphi) = (R_C - \rho \cos(\theta)) \sin(\varphi) \\ \qquad z(\rho, \theta, \varphi) = \rho \sin(\theta) \end{cases} \tag{1}$$
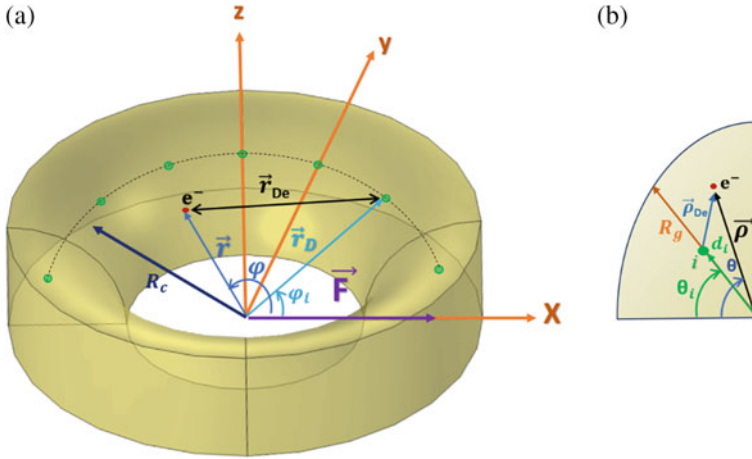
Fig. 1 **a** GaAs quantum ring in the presence of an applied electric field along the *x*-axis, and **b** present the electron and impurity coordinates in *xz*-plan of our CQR system

The impurity cartesian coordinates which moves along the e_φ-direction and fixed at $d_i = \frac{R_g}{2}$ and $\theta_i = \frac{\pi}{2}$ is written as follows:

$$
\begin{cases}
x_i(d_i, \theta_i, \varphi_i) = (R_C - d_i \cos(\theta_i)) \cos(\varphi_i) \\
y_i(d_i, \theta_i, \varphi_i) = (R_C - d_i \cos(\theta_i)) \sin(\varphi_i) \\
\quad z_i(d_i, \theta_i, \varphi_i) = \frac{R_g}{2} \sin(\theta_i)
\end{cases}
\tag{2}
$$

Within the effective mass approximation and in the presence of an applied electric field along the *x*-axis, the Hamiltonian of an electron-donor atom confined in the CQR is expressed as follows:

$$
\begin{aligned}
H_D = &-\frac{\hbar^2}{2m_e^*}\left[\frac{\partial^2}{\partial\rho^2} + \frac{R_C + 2\rho\cos\theta}{\rho(R_C + \rho\cos\theta)}\frac{\partial}{\partial r}\right. \\
&+ \frac{1}{(R_C + \rho\cos\theta)^2}\frac{\partial^2}{\partial\varphi^2} + \frac{1}{\rho^2}\frac{\partial^2}{\partial\theta^2} - \frac{\sin\theta}{\rho(R_C + \rho\cos\theta)}\frac{\partial}{\partial\theta}\right] \\
&+ V_D(\rho, \theta, \varphi) + V_C(\rho, \theta, \varphi) + eF(R_C + \rho\cos\theta)\cos(\varphi)
\end{aligned}
\tag{3}
$$

Concerning Eq. 3, the first term represents the kinetic energy of the electron with $m_e^* = 0.068m_0$ is their effective mass. The second term is the Coulomb interaction between the donor impurity and the electron which is given by:

$$
V_D(\rho, \theta, \varphi) = \frac{-e^2}{4\pi\varepsilon_0\varepsilon_r\left|\overrightarrow{r_{De}}\right|}
\tag{4}
$$

with e, ε_0 and ε_r are, respectively, the electron charge, vacuum permittivity and dielectric constant of GaAs. $\left|\overrightarrow{r_{De}}\right|$ presents the electron-impurity distance which can be expressed by:

$$\left|\overrightarrow{r_{De}}\right| = \sqrt{\begin{array}{l} ((R_c - \rho\cos(\theta))\cos(\varphi) - (R_c - d_i\cos(\theta_i))\cos(\varphi_i))^2 \\ ((R_c - \rho\cos(\theta))\sin(\varphi) - (R_c - d_i\cos(\theta_i))\sin(\varphi_i))^2 \\ + \left(\rho\sin(\theta) - \frac{1}{2}R_g\sin(\theta_i)\right)^2 \end{array}} \tag{5}$$

The quantum confinement potential of our system is written by:

$$V_c(\rho, \theta, \varphi) = \begin{cases} 0 & \text{if } R_C - R_g \leq \rho \leq R_C, 0 \leq \theta \leq \frac{\pi}{2} \text{ and } 0 \leq \varphi \leq 2\pi \\ \infty & \text{otherwise} \end{cases} \tag{6}$$

The last term of Eq. 3 $eF(R_C + \rho\cos\theta)\cos(\varphi)$ describes the electron-impurity electrostatic force due to the applied electric field. The ground state binding energy of electron-impurity is defined by the energy difference between the electronic energy without and with donor atom:

$$E_B = E_e - E_D \tag{7}$$

where E_e and E_D are, respectively, the electron energy with and without donor atom.

To solve our system of equations, we used the finite difference method. Equation 3 is a three-dimensional second-order equation, that we discretized it on a mesh containing $N_\rho * N_\theta * N_\varphi$ nodes, and divide the interval $\left[R_C - R_g, R_C\right]$, $\left[0, \frac{\pi}{2}\right]$ and $[0, 2\pi]$ into $N_\rho + 1$, $N_\theta + 1$ and $N_\varphi + 1$, respectively. $\frac{\partial}{\partial\rho}\psi(\rho, \theta, \varphi)$ and $\frac{\partial}{\partial\theta}\psi(\rho, \theta, \varphi)$ are approximated by the second-order central difference for the first derivative, $\frac{\partial^2}{\partial\rho^2}\psi(\rho, \theta, \varphi)$, $\frac{\partial^2}{\partial\theta^2}\psi(\rho, \theta, \varphi)$, and $\frac{\partial^2}{\partial\varphi^2}\psi(\rho, \theta, \varphi)$ are approximated by the second order central difference for the second derivative. Then, the discretization of Eq. (3) is written as follows:

$$\frac{\hbar^2}{2m_e^*}\left[\left(\frac{1}{\Delta\rho^2} + \frac{R_C + 2i\Delta\rho\cos\theta}{2i(R_C + i\Delta\rho\cos\theta)\Delta r^2}\right)\psi_{i+1,j,k}\right.$$

$$+ \left(\frac{1}{\Delta\rho^2} - \frac{R_C + 2i\Delta\rho\cos\theta}{2i(R_C + i\Delta\rho\cos\theta)\Delta r^2}\right)\psi_{i-1,j,k}$$

$$+ \left(\frac{1}{(i\Delta\rho)^2(\Delta\theta)^2} - \frac{\sin\theta}{2i(R_C + i\Delta\rho\cos\theta)\Delta\rho\Delta\theta}\right)\psi_{i,j+1,k}$$

$$+ \left(\frac{1}{(i\Delta\rho)^2(\Delta\theta)^2} + \frac{\sin\theta}{2i(R_C + i\Delta\rho\cos\theta)\Delta\rho\Delta\theta}\right)\psi_{i,j-1,k}$$

$$+ \left(\frac{1}{(R_C + i\Delta\rho\cos\theta)^2\Delta\varphi^2}\right)\psi_{i,j,k+1} + \left(\frac{1}{(R_C + i\Delta\rho\cos\theta)\Delta\varphi^2}\right)\psi_{i,j,k-1}$$

$$+ \frac{\hbar^2}{2m_e^*}\left[\frac{2}{(i\,\Delta\rho)^2} + \frac{1}{(R_C + i\,\Delta\rho\cos\beta)^2\,\Delta\varphi^2} + \frac{1}{(i\,\Delta\rho)^2(\Delta\beta)^2} + \frac{2m_e^*}{\hbar^2}V_D^{i,j,k}\right.$$
$$\left. + eF(R_C + i\,\Delta\rho\cos(\theta))\cos(\varphi)]\psi_{i,j,k}\right] = E_D\psi_{i,j,k} \tag{8}$$

where: $\Delta\rho = R_g/(N_\rho + 1)$ is the pitch (the distance between two consecutive grid nodes in the $\overrightarrow{e_\rho}$ direction), $\Delta\theta = \frac{\pi}{2(N_\theta+1)}$ is the pitch (the angle between two consecutive grid points in the $\overrightarrow{e_\theta}$ direction), and $\Delta\varphi = \frac{2\pi}{N_\varphi+1}$ is the distance between two consecutive grid points in the $\overrightarrow{e_\varphi}$ direction. $V_D^{i,j,k}$ is the coulombic interaction between the electron and the donor atom at point (i, j, k). The energy eigenvalues are obtained by numerically solving this system of equations.

3 Results and Discussion

To give an overall picture of the effects of electric field, impurity position, and geometrical confinement on the electron-impurity binding energy, we have plotted in Fig. 2 the variation of the electron-impurity binding energy as a function of electric field strength, for different impurity position ($\varphi_i = 0°$ and $180°$) and three quarter-circle cross section radius of CQR $R_g = 4$, 6 and 8 nm with $R_C = 10$ nm. From this figure, the following behaviors can be noted: (i) When the impurity localized in $\varphi_i = 0°$ the application of the electric field produces a decreasing function of the binding energy, especially at the interval $[0, 20\,\text{kV cm}^{-1}]$. The physical reason for this behavior is that the electric field strength forces the electron to move away from the impurity position, which leads to an increase in the average of the electron-impurity distance and accompanied by a reduction in the Coulomb interaction energy, this energy is decreased when the geometric radius of the quarter-circle cross section is increased. However, close the electric field strength $20\,\text{kV cm}^{-1}$, the binding energy will be changing their behavior with the geometrical radius. From the value of electric field intensity $(F > 20\,\text{kV cm}^{-1})$, the electron wave function is very strongly confined away from impurity, which it does not respond to the effects of the electric field (the impurity is localized in the same direction of electric field), consequently, the binding energy remains stable at this electric field interval whatever the geometrical radius $R_g = 4$, 6 and 8 nm. (ii) When the impurity localized at $\varphi_i = 180°$, the impurity is localized in the opposite of electric field direction, the application of this field leads to increases the binding energy, and this is due to the fact that the electric field effect pushes the electron toward the impurity.

In Fig. 3, we have studied the binding energy as a function of the electric field strength for different curvature radius ($R_c = 8$, 10 and 12 nm) and two impurity positions ($\varphi_i = 0$ and $180°$) with $R_g = 5$ nm, $d_i = \frac{R_g}{2}$ and $\theta_i = 45$. We can observe that the effect of the impurity position on the bound electron-impurity energy versus electric field behavior is similar to that obtained in Fig. 2. Also, the decrease of the curvature radius leads to an increase of the binding energy when the impurity
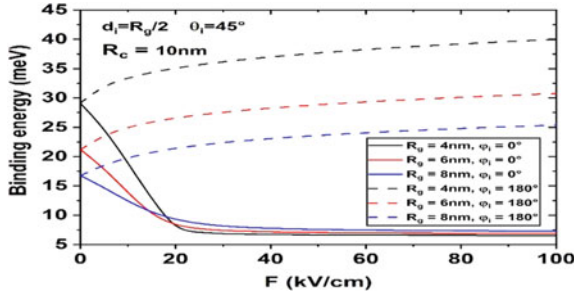
Fig. 2 Binding energy as a function of the applied electric field for different radius geometry $R_g = 4, 6$ and 8 nm and two impurity positions $\varphi_i = 0°$ and $180°$. Results are with a fixed value of the tore radius $R_C = 10$ nm and azimuthal impurity position $d_i = \frac{R_g}{2}$ and $\theta_i = 45°$

is localized at $\varphi_i = 0°$. In this case ($\varphi_i = 0°$), the increase of the electric field intensity leads to decreases the binding energy until reach critical value, and then it saturates, the saturation energy values are dependent on the curvature radius. It is clear that as the curvature radius increases, the saturation of the binding energy is faster. On the other hand, the binding energy increases with the curvature radius when the impurity is located at $\varphi_i = 180°$, especially, when the electric field at the range $F < 30$ kV cm^{-1}. Also, for $F > 30$ kV cm^{-1}, and for all curvature radius $R_c = 8, 10$ and 12 nm and for a given impurity position $\varphi_i = 180°$, the binding energy converges to the same value with electric field.

As it is known, when the electric field is applied, the electrons start to localize towards the opposite direction of the electric field. Depending on the position of the impurity inside the CQR, the binding energy can be increases or decreases. In order to show this behavior, we have presented in Fig. 4 the variation of the binding energy of the donor atom in a GaAs CQR as a function of the electric field for different impurity positions of $\varphi_i = 0$ ($\varphi_i = 0, 90$ and $180°$). We can observe the following
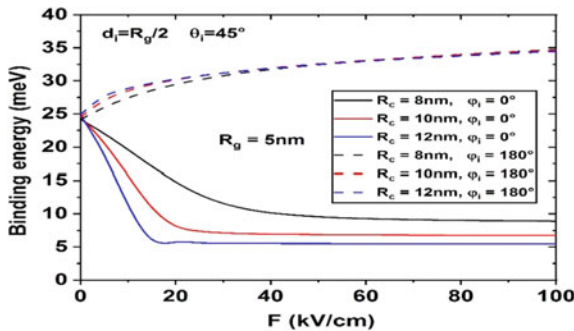
Fig. 3 Binding energy as a function of the applied electric field for different curvature radius of $R_c = 8, 10$ and 12 nm and two azimuthal impurity positions $\varphi_i = 0$ and $180°$, with a fixed value of the geometrical radius $R_g = 5$ nm and impurity position $d_i = \frac{R_g}{2}$ and $\theta_i = 45°$
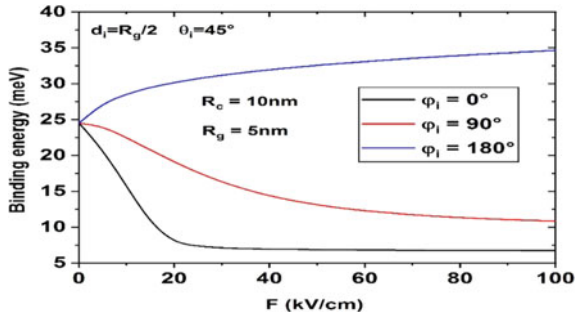
Fig. 4 Binding energy as a function of the electric field, for three impurity positions $\varphi_i = 0°$, $90°$ and $180°$, with a fixed $R_g = 5$ nm, $R_c = 10$ nm, $d_i = \frac{R_g}{2}$ and $\theta_i = 45°$

features: (i) When $\varphi_i = 0°$, the binding energy decreases rapidly as the applied field strength increases and tends to a saturation value with the electric field $20\,\text{kV cm}^{-1}$, due to the displacement of the electron away from the impurity position. (ii) For $\varphi_i = 90°$, the binding energy decreases slowly because the charge distribution is less sensitive to the influence of the electric field. (iii) For $\varphi_i = 180°$, the impurity is in the opposite direction than the applied electric field, it is clear that the binding energy increases with increasing electric field.

In Fig. 5, we studied the variation of the binding energy as a function of the impurity position φ_i for different of the electric field strength ($F = 40$, 60 and $80\,\text{kV cm}^{-1}$). Results are with a fixed value of the radius $R_g = 5$ nm, $R_c = 10$ nm and impurity position $d_i = \frac{R_g}{2}$ and $\theta_i = 45$. For each electric field strengths, we notice a clearly increasing monotonic behavior of the binding energy as the impurity moves away from the electric field direction to their opposite direction, this is due to the fact that the impurity approaches the electronic probability density directed by the direction of the electric field. Indeed, three regions can be discussed: For the impurity position $\varphi_i < 30°$, the binding energy is insensitive to the combined effect of the impurity position and the applied electric field intensity, which becomes constant at the average value of 7 meV. When the azimuthal impurity position at the region $30° < \varphi_i < 150°$, we clearly observe that an increase of the electric field leads to a decrease of the binding energy. Also, for $\varphi_i = 150°$, the effect of the electric field is negligible. Also, when $\varphi_i > 150°$, the increase of the electric field leads to the increase of the binding energy. The effects of electric field and impurity position in this figure are exactly the same as those observed in the previous figures. As a summary to this section of the study, we can conclude that the electron-donor atom binding energy under electric field strength is an essential basic parameter that influences measurable properties of a nanostructures, such as solar cell efficiency and optoelectronic devices.
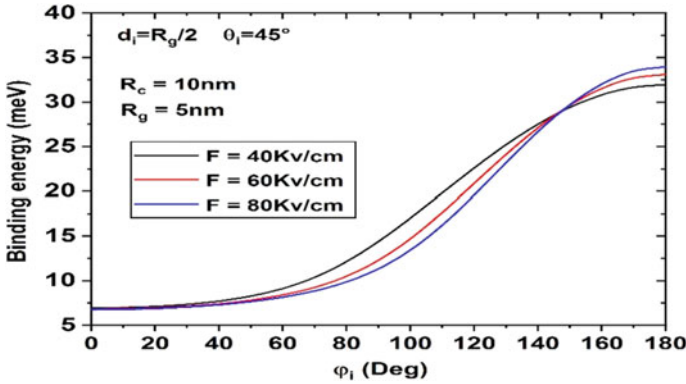
Fig. 5 Binding energy as a function of the impurity position (φ_i) for three electric field strength F = 40, 60 and 80 kV cm^{-1}. Results are with a fixed value of the radius R_g = 5 nm, R_c = 10 nm and impurity position $d_i = \frac{R_g}{2}$ and $\theta_i = 45°$

4 Conclusion

In conclusion, considering an infinite confinement potential, the dependence of the binding energy on the applied electric field, an off-center donor atom and the CQR size has been analyzed in detail within the effective mass approximation. By using the finite difference method, we have found that the binding energy depends on the geometry of the structure, the impurity position and the electric field strength. Our results indicate that the binding energy is insensitive to the position of the impurity and the high electric field strength when the impurity is close to the direction of the applied electric field, and it is decreased with the electric field when the impurity is located at the azimuthal-interval $30° < \varphi_i < 150°$, as well as this electron-donor atom bound energy increases at $\varphi_i > 150°$. Also, it is important to note the critical impurity position ($\varphi_i = 150°$), in which the electric field effect is negligible whatever their intensity.

References

1. Dubertret B (2004) Boîtes quantiques en biologie: progrès récents. Med Sci 20(8–9):737–740
2. Danek M, Jensen KF, Murray CB, Bawendi MG (1994) Preparation of II–VI quantum dot composites by electrospray organometallic chemical vapor deposition. Crystal Growth 145(1–4):714–720
3. Franchi S, Trevisi G, Seravalli L, Frigeri P (2003) Quantum dot nanostructures and molecular beam epitaxy. Prog Cryst Growth Charact Mater 47(2–3):166–195
4. Boussetta R, Mommadi O, Belamkadem L, Chouef S, Hbibi M, El Moussaouy A, El-Miad AK (2022) Deformation and size effects on electronic properties of toroidal quantum dot in the presence of an off-center donor atom. Micro Nanostruct 207209

5. Barseghyan MG, Kirakosyan AA, Duque CA (2009) Donor-impurity related binding energy and photoionization cross-section in quantum dots: electric and magnetic fields and hydrostatic pressure effects. Eur Phys J B 72(4):521–529
6. Heyn C, Duque CA (2020) Donor impurity related optical and electronic properties of cylindrical GaAs/Al$_x$Ga$_{1-x}$As quantum dots under tilted electric and magnetic fields. Sci Rep 10(1):1–18
7. Duque CM, Acosta RE, Morales AL, Mora-Ramos ME, Restrepo RL, Ojeda JH, Duque CA (2016) Optical coefficients in a semiconductor quantum ring: electric field and donor impurity effects. Opt Mater 60:148–158
8. Llorens JM, Trallero-Giner C, Garcia-Cristobal A, Cantarero A (2001) Electronic structure of a quantum ring in a lateral electric field. Phys Rev B 64(3):035309
9. Mommadi O, El Moussaouy A, Chnafi M, El Hadi M, Nougaoui A, Magrez H (2020) Exciton–phonon properties in cylindrical quantum dot with parabolic confinement potential under electric field. Phys E 118:113903
10. Mommadi O, El Moussaouy A, El Hadi M, Chnafi M, Meziani YM, Duque CA (2021) Stark shift and exciton binding energy in parabolic quantum dots: hydrostatic pressure, temperature, and electric field effects. Phil Mag 101(6):753–775
11. Niculescu EC, Bejan D (2015) Nonlinear optical properties of GaAs pyramidal quantum dots: effects of elliptically polarized radiation, impurity, and magnetic applied fields. Phys E 74:51–58
12. Chnafi M, Belamkadem L, Mommadi O, Boussetta R, El Hadi M, El Moussaouy A, Duque CA (2021) Hydrostatic pressure and temperature effects on spectrum of an off-center single dopant in a conical quantum dot with spherical edge. Superlattices Microstruct 159:107052
13. Bala KJ, Peter AJ, Lee CW (2017) Simultaneous effects of pressure and temperature on the optical transition energies in a Ga$_{0.7}$In$_{0.3}$N/GaN quantum ring. Chem Phys 495:42–47
14. Restrepo RL, Barseghyan MG, Mora-Ramos ME, Duque CA (2013) Effects of hydrostatic pressure on the nonlinear optical properties of a donor impurity in a GaAs quantum ring. Phys E 51:48–54

Numerical Analysis of Various Hole Transport Material (HTM) for an Efficient Perovskite Solar Cell

Touria Ouslimane , **Lhoussayne Et-taya** , **and Abdellah Benami**

Abstract Component hybrid perovskites solar cells (PSC) are emerging as a new class of semiconductors that combine high absorption coefficient and excellent transport properties. In this context, we are modeling perovskite solar cells using a number of organic and inorganic hole transport materials (HTMs) using SCAPS-1D software. A comparative theoretical study on the performance of PSCs is reported using methyl ammonium lead iodide (MAPbI$_3$) as an absorber layer and tungsten trioxide WO$_3$ as the electron transport material (ETM). After examining various solar cell configurations, the Glass/FTO/WO$_3$/MAPbI$_3$/SnS/Au design has the optimum photovoltaic performance. This model is made to optimize the thickness of the absorbing layer, and its acceptor dopant density N_A. Simulation results reveal that the higher power conversion efficiency is 32.99% for an absorber thickness of 400 nm, and for dopant density N_A of 1×10^{17} cm^{-3}, with a fill factor (FF) of 73.09%.

Keywords HTM · ETM · PSC · SCAPS-1D · Perovskite

1 Introduction

Perovskite solar cells (PSCs) research is substantially increasing from 2013, because of the fast improvement in their power conversion efficiency (PCE), cheapness, possibility to tune the bandgap, low recombination rate, high open-circuit voltage, and strong and broad optical absorption. Power conversion efficiency (PCE) with

T. Ouslimane (✉) · L. Et-taya · A. Benami
LM3ER-OTEA, Department of Physics, Faculty of Sciences and Techniques, Moulay Ismail University of Meknes, BP 509 Boutalamine, 52000 Errachidia, Morocco
e-mail: t.ouslimane@edu.umi.ac.ma

L. Et-taya
e-mail: l.ettaya@edu.umi.ac.ma

A. Benami
e-mail: a.benami@fste.umi.ac.ma

adequate stability is the most crucial challenge for PSCs for commercial applications at a large scale. Various methodologies have been developed for realizing this goal, among them absorber layer optimization. In contrast to absorber layer, very scant studies have been reported on HTMs, which can be the electrical contacts between absorber layers, and electrodes, and lead to enhancement of charge transport and collection. Thus, the HTM in a photovoltaic device plays a pivotal role in device stability, PCE, influencing the work function of the electrode and the properties of the absorber layer. The types of organic HTL materials are limited; the most widely used is water soluble PEDOT:PSS, while several investigations have demonstrated that due to the hygroscopic, acidic, and protonation nature of PSS in PEDOT:PSS, it influences the device stability and degradation which limits their possible commercial applications [1].

In this paper, we have analyzed several recent studies on the most adequate ETM. Then, WO_3 shows optimal results at the level of PCE and FF comparing it with ZnO. Its properties like tunable bandgap (2.6–3.5 eV) [2] correspond to a wavelength of 475 nm which in turn enables it to absorb light in the visible spectrum. In addition, it has an appreciable transmittance of (80%) in the visible region. Also, WO_3 has good conductivity of the order of $10^{-3}/\Omega$-cm with n-type semiconductor properties, stable structure, non-toxic nature, availability, high surface area, low cost, and good chemical stability which are some of its advantages [3]. A variety of materials such as NiO, CuO_2, CuI, SnS, a-Si, SpiroMeOTAD, P3HT, $CuSbS_2$, CuSbSe, PEDOT:PSS, D-PBTTT are simulated in PSCs as HTMs and analyzed their PV parameters. The perovskite layer is of $CH_3NH_3PbI_3$ (methyl ammonium lead Iodide) having a bandgap between 1.5 and 2.4 eV, showing both organic and inorganic properties which gives increased carrier mobility, long charge diffusion length, low non-radiative auger recombination, and higher absorption coefficient [2, 4].

2 Device Simulation with SCAPS-1D

In this work, a device simulator, SCAPS-1D developed by University of Gent was employed to model PSCs. It is a more popular software in photovoltaic cells simulation, and it is characterized by its ability to model a maximum seven-layer structure. J–V, C–V, C–f characteristics and quantum efficiency of the solar cell can be measured [5–7]. It numerically solves one-dimensional equations that govern the semiconductor material under steady state conditions.

Poisson Eq. (1), which presents the relationship between electric field of a p–n junction (E) and the space charge density (ρ), is given by:

$$\frac{\partial^2 \psi}{\partial^2 x} = -\frac{\partial E}{\partial x} = -\frac{\rho}{\varepsilon_S} = -\frac{q}{\varepsilon_S}\left[p - n + N_D^+(x) - N_A^-(x) \pm N_{\text{def}}(x)\right] \quad (1)$$

where is ψ the electrostatic potential, q is elementary charge, ε_S is the static relative permittivity of the medium, $n(p)$ is the electron (hole) density, $N_D^+ (N_A^-)$ is the density

of ionized donors (acceptors), and N_{def} is the possible defect (Acceptor or donor) density. The electron and hole continuity equations in steady state are given by Eqs. (2) and (3)

$$\frac{\partial j_n}{\partial x} + G - U_n(n, p) = 0 \tag{2}$$

$$-\frac{\partial j_p}{\partial x} + G - U_p(n, p) = 0 \tag{3}$$

j_n, j_p are the electron and hole current densities; $U_{n,p}$ is the net recombination rates; G is the electron–hole generation rate. The electron and hole current density are given by Eqs. (4) and (5) [2]:

$$j_n = qn\mu_n E + qD_n\frac{\partial n}{\partial x} \tag{4}$$

$$j_p = qn\mu_p E - qD_p\frac{\partial p}{\partial x} \tag{5}$$

q is the elementary charge, $\mu_n(p)$ is electron (hole) mobility, and $D_n(p)$ is diffusion coefficient of electrons (holes).

The perovskite ($CH_3NH_3PbI_3$) solar cell simulated is a normal planar architecture; it is illustrated in Fig. 1. The n region is the ETL, and the p region is the HTL. When the cell is subject to light, excitons (bound state of an electron and a hole) are created mainly in the perovskite layer. According to their diffusion length, they can reach the n(p) region. The dissociation of excitons and the migration of electrons and holes is favored by the electrical field between the n and p layers.

In a first step, we simulated different structures using organic and inorganic HTMs, and we extracted that SnS gives good results. Then, we did the simulation with the best HTM chosen previously by varying the ETM between ZnO and WO_3, and we conclude that structure Glass/FTO/WO_3/$MAPbI_3$/SnS/Au gives the best results. Table 1 summarizes the input parameters for the layers of the optimal structure.

Fig. 1 Perovskite solar cell structure simulated under SCAPS
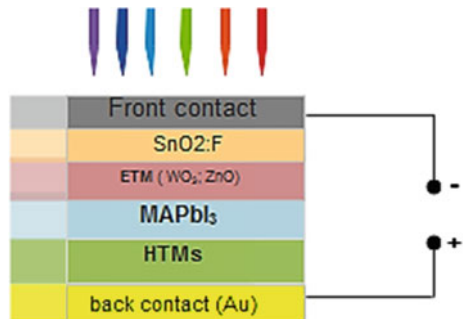
Table 1 Input parameters of different layers used in the simulation [6–8]

Material properties	FTO (SnO:F)	ETL (WO$_3$)	Absorber (MAPbbI$_3$)	HTM (SnS)
d (nm)	500	50	50^a	50
E_g (eV)	3.5	2.98	1.55	1.1
χ (eV)	4	2.1	3.9	3.52
ε (eV)	9	6.5	6.5	12.5
N_c (cm^{-3})	2.2×10^{18}	2.8×10^{19}	2.2×10^{18}	7.5×10^{18}
N_v (cm^{-3})	1.8×10^{19}	1×10^{19}	1.8×10^{19}	10^{19}
V_{the-} (cm/s)	10^7	10^7	10^7	107
V_{thp} (cm/s)	10^7	10^7	10^7	10^7
μ_n (cm^2/V s)	20	1.7×10^{-4}	50	100
μ_h (cm^2/V s)	10	2×10^{-4}	50	4
N_D (cm^{-3})	2×10^{19}	10^{18}	0	0
N_A (cm^{-3})	0	0	10^{13a}	10^{19}
N_t	10^{15}	10^{15}	10^{15}	10^{15}
α (cm^{-1})	File	File	File	File

d Thickness, E_g Bandgap, χ Electron affinity, ε Dielectric permittivity, N_c/N_v Density of states in CB/VB, V_{the-}/V_{thp} Thermal velocity of electron/hole, μ_n/μ_h Electron/Hole mobility, N_D/N_A Donor/Acceptor density and α Absorption coefficient, N_t Total defect density
^{a} Is a variable parameter

The simulation is carried out under AM1.5G solar spectrum with an incident power density of 100 mW/cm^2 at room temperature (300 K) [8]. We have assumed a Gaussian defect energy level of 0.6 eV below the perovskite's conduction band with a characteristic energy of 0.1 eV. To circumvent the simulator's empirical absorption coefficient, we have to input it from external files. Related to the extinction coefficient calculated using

$$\alpha = \frac{4\pi k}{\lambda}, \tag{6}$$

where k and λ are extinction coefficient and wavelength, respectively. While the absorption coefficient of the PVK was extracted from [6].

3 Results and Discussions

In this section, we present the results obtained from the effect of the thickness of the perovskite absorber layer as well as the acceptor dopant density, on the electrical performance of the PSC.

3.1 Effect of Dopant Density of the Absorber Layer

To improve the performance of the solar cell, n-type or p-type dopants are introduced into the absorber layer. In this study, the acceptor dopant density of the absorber layer is varied from 10^{13} to 10^{20} cm^{-3}. The PCE of the solar cells increases gradually with increasing dopant concentration and reaches an optimum value at a concentration of 10^{17} cm^{-3}. However, if the dopant concentration of the absorber layer is below the optimum value, the performance parameters are found to be stable, but not optimal. As a result, in this range of dopant density, the rate of generation and recombination effects of charge carriers are almost in equilibrium, leading to a stable performance of the device. The J_{sc} decreases as the dopant concentration increases, resulting in a decrease in PCE at higher dopant concentrations, as shown in Fig. 2. At lower concentrations, the J_{sc} and PCE values increase due to a decrease in the dark current of the solar cell. However, at higher dopant concentrations, the diffusion length of minority carriers decreases, and heavy doping causes an increase in dark current, resulting in a decrease in JSC and PCE [9]. This result is consistent with the literature results for a controllable design of solid-state PSCs [10].
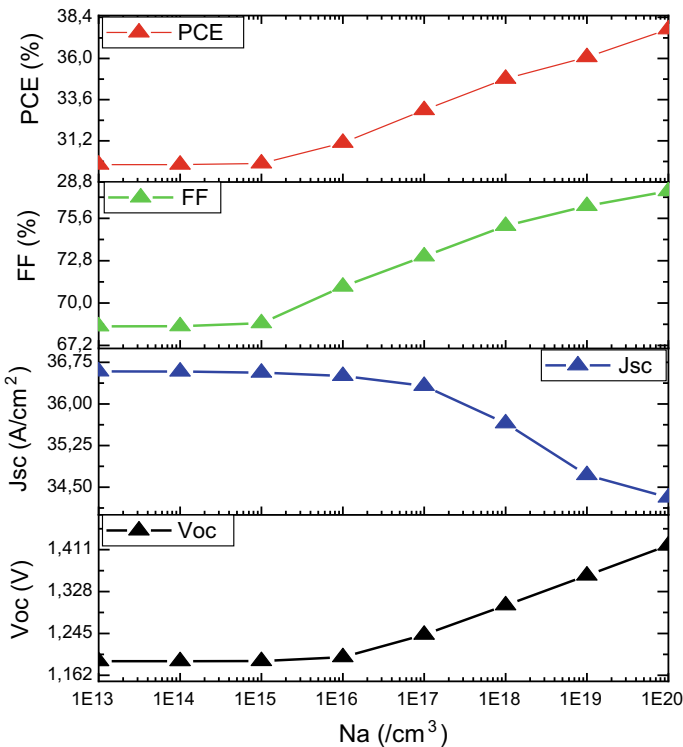
Fig. 2 Variation of J_{sc}, PCE, FF, and V_{oc} with variation of doping concentration

3.2   Effect of Thickness of the Absorber Layer

HTM and ETM thicknesses are optimized and fixed at 50 nm for both. The thickness of the absorber layer is a critical parameter that influences the electrical properties of the PSC. The thickness of the perovskite light harvesting layer is varied between 50 and 1000 nm in order to determine the optimum value required for maximum performance parameters. The variation of V_{OC}, J_{SC}, FF and PCE with the thickness of the perovskite absorber layer is shown in Fig. 3. The PCE of the solar cell increases with thickness because the wider absorber layer absorbs more light, resulting in the generation of a large number of electrons–hole pairs and a corresponding increase in current density. The PCE reaches its maximum at a thickness of 400 nm and then begins to decline gradually. Furthermore, as the absorber layer thickness is gradually increased, the FF decreases with an increase in J_{sc}. However, when the thickness of the absorber layer is increased above an optimal value, the generated electron–hole pairs recombine, resulting in a reduction in the device's PCE.

Fig. 3 The variation of J_{sc}, PCE, FF, and V_{oc} with absorber thickness
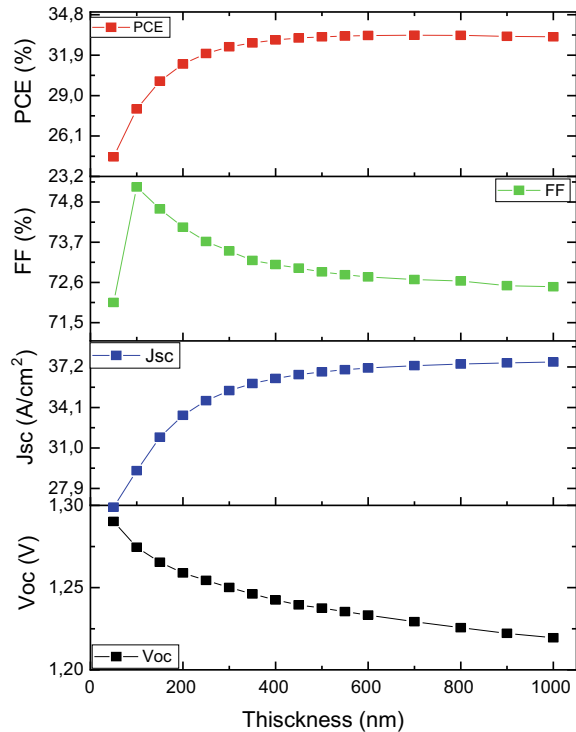
4 Conclusion

Numerical simulation using SCAPS-1D software allows us to get into the details of thin film solar cell structure operation. PSC performance is improved in this work by employing a variety of HTMs. The results show that the photovoltaic performance of organic compounds such as HTM, such as PEDOT:PSS, P$_3$HT, and others, is poor. This demonstrates that inorganic HTMs, such as NiO, CuI, CuO$_2$, and others, can replace the costly and degradable organic HTMs in PSCs. It has been discovered that the best performance is obtained with SnS, a non-toxic inorganic earth abundant material, as HTM and WO$_3$ as ETM. A detailed analysis of the performance variation with absorber layer thickness reveals that a thickness of 400 nm produces the best results. The performance of lead-based PSCs suffers as the absorber becomes thicker. The acceptor dopant concentration in the absorber layer is also changed to see how it affects the performance of the PSC. The chosen configuration Glass/FTO/WO$_3$/MAPbI$_3$/SnS/Au has the best PCE value of 32.99% when the dopant concentration N$_A$ and thickness are 10^{17} cm^{-3} and 400 nm, respectively. It should be advantageous to design a device and comprehend its operation mechanism, which can be used to guide and innovate experiment approaches and fabrication processes.

References

1. Chaudhary N et al (2015) Copper thiocyanate (CuSCN): an efficient solution-processable hole transporting layer in organic solar cells. J Mater Chem C 3(45):11886–11892. https://doi.org/10.1039/C5TC03124A
2. Benami A et al (2022) Comparison of the effects of ZnO and TiO$_2$ on the performance of perovskite solar cells via SCAPS-1D software package. J Nano-Electron Phys 14(1):01033. http://doi.org/10.21272/jnep.14(1).01033
3. Swarnkar A et al (2016) Quantum dot–induced phase stabilization of α-CsPbI$_3$ perovskite for high-efficiency photovoltaics. Science 354(6308):92–95. http://doi.org/10.1126/science.aag2700
4. Rai N et al (2020) Analysis of various ETL materials for an efficient perovskite solar cell by numerical simulation. J Mater Sci Mater Electron 31(19):16269–16280. http://doi.org/10.1007/s10854-020-04175-z
5. Azri F, Meftah A, Sengouga N, Meftah A (2019) Electron and hole transport layers optimization by numerical simulation of a perovskite solar cell. Sol Energy 181:372–378. http://doi.org/10.1016/j.solener.2019.02.017
6. Ouslimane T, Et-taya L, Elmaimouni L, Benami A (2021) Impact of absorber layer thickness, defect density, and operating temperature on the performance of MAPbI$_3$ solar cells based on ZnO electron transporting material. Heliyon 7(3):e06379. http://doi.org/10.1016/j.heliyon.2021.e06379
7. Garain R, Basak A, Singh UP (2021) Study of thickness and temperature dependence on the performance of SnS based solar cell by SCAPS-1D. Mater Today Proc 39:1833–1837. https://doi.org/10.1016/j.matpr.2020.06.185
8. Arvizu MA, Granqvist CG, Niklasson GA (2016) Electrochromism in sputter deposited W1-yMoyO3 thin films. J Phys Conf Ser 682:012005. http://doi.org/10.1088/1742-6596/682/1/012005

9. Jayan KD, Sebastian V (2021) Comparative study on the performance of different lead-based and lead-free perovskite solar cells. Adv Theory Simul 4(5):2100027. http://doi.org/10.1002/adts.202100027

10. Tan K et al (2016) Controllable design of solid-state perovskite solar cells by SCAPS device simulation. Solid-State Electron 126:75–80. https://doi.org/10.1016/j.sse.2016.09.012

Photonic Flat Bands of Asymmetric Star Waveguides Structure

Y. Errouas, Y. Ben-Ali, I. El Kadmiri, and Driss Bria

Abstract Using the interface response theory, we investigate the band structure and transmission rate of 1D photonic star waveguides with asymmetric resonators grafted onto N equidistant sites. The resonator lengths in each site depend on each other following an arithmetic sequence. The results show that when we have one resonator in each site, we obtain large band gaps and narrower pass bands. On the other hand, the difference between the lengths of resonators in each site gives a rise to new pass bands (called flat bands) inside the gaps. These flat bands move to lower frequencies when the length step (difference between two successive resonators) of the resonators increases. The quality factors of the modes situated in these flat bands become very important, when the number of the resonators in each site increases. In addition, the width of the gap bands becomes larger, when the sizes of the asymmetric resonators in each site increase. The thinner created pass bands will be useful for telecommunication applications of types: filtering, dense frequency division multiplexing (DFDM) and drop channel for the extraction of frequencies.

Keywords SWGs · Flat bands · Asymmetric resonators · Gaps · Pass bands

1 Introduction

Electromagnetic waveguide systems are used for a variety of applications, such multi-frequency coherent electromagnetic waves communications, generation of low-phase-noise or agile ultra-broadband microwaves, electromagnetic arbitrary waveform generation, signal processing and filtering [1–6]. Recently, the study of star waveguides (SWGs) systems based on the coaxial cables has been the subject of

Y. Errouas (✉) · Y. Ben-Ali · I. El Kadmiri · D. Bria
Laboratory of Materials, Waves, Energy and Environment, Team of Acoustics, Photonics and Materials, Faculty of Sciences, Mohamed First University, Oujda, Morocco
e-mail: younes.errouas@gmail.com

Y. Ben-Ali
Engineering Sciences Laboratory (LSI), Multidisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, B.P. 1223, Taza Gare, Morocco

several works due to their interest as well as their extraordinary properties [7–10]. These SWGs are composed of the periodicity of the segment and grafted in each site by a finite number of resonators that have the same length. The periodicity of the system creates pass and larger gap bands where the propagation of electromagnetic waves is forbidden [11]. These gaps are due to the periodicity of the segment and the number of symmetric resonators grafted in each site. When the number of symmetric resonators in each site increases, the width of the gaps increases too [12]. In this structure, the theoretical model assumed that the cross sections of the resonators and segments are small compared to their linear dimensions, which can be considered one-dimensional waveguide. In this paper, we study the behavior of electromagnetic waves in 1D SWGs composed by the periodicity of the segment and grafted in each site by a finite number of asymmetric resonators. The resonator lengths in each site depend on each other following an arithmetic sequence (see Fig. 1). Many of the previous works are concentrated on the photonic comb-like structure due to their applications, despite all these works which take the resonators with the same lengths (symmetric resonators) [12]. The advantage of this work is that we have been able to introduce our narrower flat bands, which is impossible to obtain if we have symmetrical resonators. Recently, our team have presented a perfect structure of 1D SWGs with the presence of Fibonacci sequences at the resonators level; this study shows that this structure can present localized bands of higher amplitude in the transmission spectrum. Such structure can be used in the field of electromagnetic telecommunications, the design of new multichannel filters, demultiplexers and other engineering devices [7]. The dispersion relation according to an infinite SWGs system and the transmission rate concerning the propagation of electromagnetic waves in finite SWGs are calculated using the formalism of the green function cited in [7–9, 11, 13, 14].
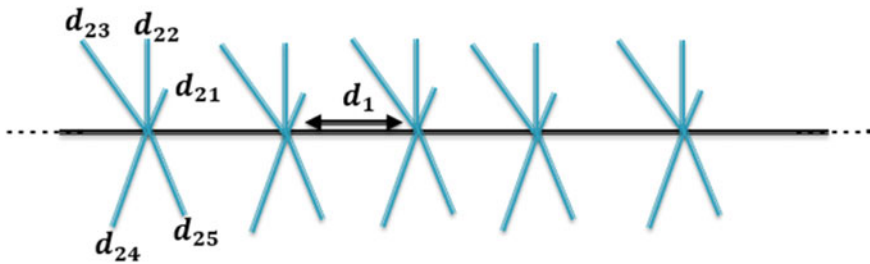
Fig. 1 One-dimensional star waveguides structure with a periodic segment of length d_1 grafted in each site by a finite number of resonator lengths d_{2j} [$j = 1$–5], $D = d_1$ represents the period of the structure

2   Results and Discussions

Firstly, we study the dispersion relation and the transmission rate of the 1D SWGs system composed of the periodicity of segments of lengths d_1 and grafted in each site by the resonators, which have different lengths. The resonators lengths depend on each other by following an arithmetic sequence with d_{2j} [$j = 1$–5]. The materials constituting the segments and the resonators are assumed homogeneous, nonmagnetic ($\mu_1 = \mu_{2j} = 1$), with identical relative permittivity, i.e., $\varepsilon_1 = \varepsilon_{2j} = 4$. We present two principal results; the first one is the effect of the variation of the length of each resonator on the behavior of the gaps corresponding to the infinite SWGs, and in the second case, we are interested in the impact of changing the length of a single resonator on the transmission spectrum. In this study, we define that the reduced frequency $\Omega = \omega\sqrt{\varepsilon_1\mu_1}D/(c)$ with c is the velocity of electromagnetic waves in a vacuum and ω is the pulsation (in s^{-1}), and $D = d_1$ represents the period of the system.

2.1   Effect of the Different Resonator Lengths on the Creation of Thin Pass Bands

In this section, we study in Fig. 2 the variation of the reduced frequency Ω versus the real part of KD for different lengths of the resonators grafted in the structure. The black lines show the pass bands, while the white areas correspond to the gap bands where the electromagnetic waves cannot propagate. In the case (a), where the resonators have the same lengths $d_{2j} = 0.5D$ [$j = 1$–5], we observe the existence of large gap bands, separated by pass bands. This result is identical to the experimental and theoretical results found in the reference [15]. In the case (b), where we take the length of a single resonator $d_{21} = 0.6D$ and the other resonators have the same lengths $d_{2j} = 0.5D$ [$j = 2$–5], we observe the creation of narrow pass bands (flat bands) inside the gap bands. For example, in the first gap, a flat band is created at the reduced frequency $\Omega = 0.86$, and in the second gap, it appears at the reduced frequency $\Omega = 2.8$. In addition, we observe the creation of narrow flat bands in the third pass band at the reduced frequency Ω range varying between $\Omega = 4$ and $\Omega = 4.1$. In the case (c), where $d_{21} = 0.5D$ and the four other resonators have different lengths, we observe the creation of new flat bands inside the gaps. These flat bands are due to the asymmetric resonator lengths that break the symmetry at the resonators level. Consequently, this new structure allows the creating of new pass bands (flat bands) inside the gaps. Therefore, we deduce that the creation of resonators of different lengths allows the creation of narrower flat bands, which is not possible if we take the identical resonator lengths in each site [9–13].
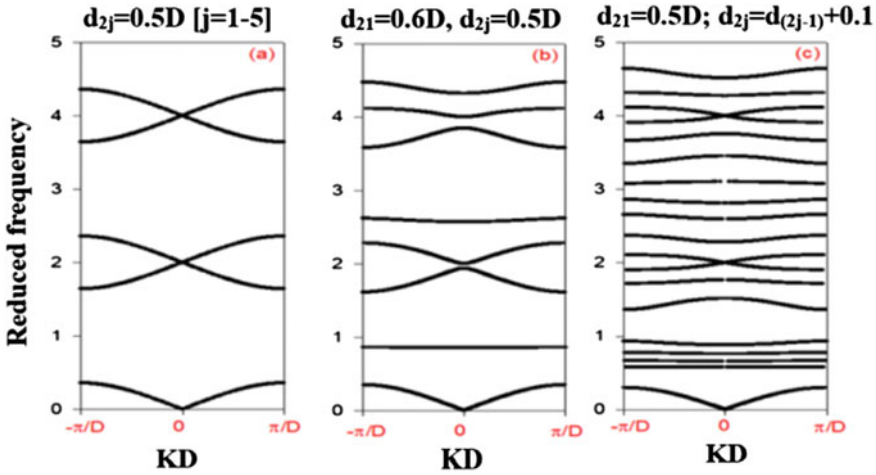
Fig. 2 Variation of the reduced frequency Ω versus the real part of KD for different cases; **a** $d_{2j} = 0.5D$ [$j = 1–5$]; **b** $d_{21} = 0.6D$, $d_{2j} = 0.5D$ [$j = 2–5$]; and **c** $d_{21} = 0.5D$; $d_{2j} = d_{(2j-1)} + 0.1$

2.2 Effect of the Number of the Resonators

In this part, we study the variation of the thinner pass bands width as a function of the number N' of resonators grafted in each site. In each site, we consider N'_1 resonators of length $d_{21} = 0.5D$ and N'_2 resonators of length $d_{22} = 0.6D$ ($N' = N'_1 + N'_2$). Figure 3 shows the variation of the reduced frequency Ω as a function of N'_2 for different values of N'_1. For N'_2 greater than or equal to 1, we notice the creation of flat bands inside the band gaps. These flat bands shift slightly toward higher frequencies when N'_2 increases. At the same time, we can observe that the width of the gaps becomes larger when the number of N'_1 increases.

2.3 Effect of the Resonators Lengths

For a step (difference between two successive resonators step $= d_{2j} - d_{2(j-1)}$) equals to 0, i.e., resonators of the same lengths, we show the existence of pass bands separated by large gaps. The first gap is situated at $0.4 < \Omega < 1.6$. When we increase the step, we remark the creation of new permissible bands inside the gap band. These flat bands shift to lower frequencies, when the step increases. In addition, we observe the creation of new gaps inside the pass bands associated to step different to zero. These new gaps shift to lower frequencies as the step increases and become pass bands for a well-determined step (step equal to 0.5 for $\Omega = 2.07$). Figure 4b is a zoom of Fig. 4a for the step range varying between 0 and 0.08. We notice the apparition of a narrower flat band with a lower reduced width of $\Delta\Omega = 0.008$. Figure 4c
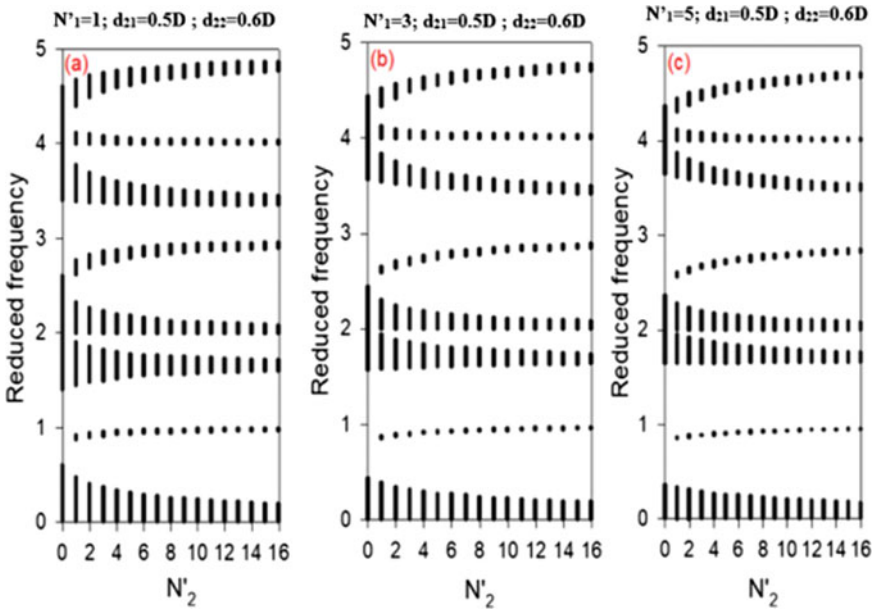
Fig. 3 Variation of the reduced frequency Ω versus N'_2 for different values of N'_1. The others parameters are $d_1 = 1D$; $d_{21} = 0.5D$; and $d_{22} = 0.6D$

represents the variation of the transmission rate through a finite asymmetric SWG system composed of $N = 7$ sites, as a function of the reduced frequency Ω for a step equal to 0.08. This Fig. 4c demonstrates that the transmission rates of the flat bands located in the first and the second gaps are higher with a frequency shift equal to $\Delta\Omega = 0.008$. Figure 4d is the same as Fig. 4b but with range step belonging to the range [0.08–0.16]. This figure shows also that the flat bands shift into a lower frequency by increasing the step. Figure 4e represents the variation of the transmission rate through a finite asymmetric SWG system composed of $N = 7$ sites, as a function of the reduced frequency Ω for a step equal to 0.16. Figure 4e demonstrates that the transmission rates of the flat bands located in the first and the second gaps are very important with the frequency shift equals to 0.014. Due to the presence of these flat bands that are characterized by a higher value of transmission rate and a higher quality factor, this system makes it possible to filter several frequencies belonging to the bandwidth of these flat bands.

2.4 Effect of the Number of the Grafted Resonators

In this part, we focus our attention on the frequency range that varies between $\Omega = 1.6$ and $\Omega = 2.3$. For $N'_2 = 1$ (Fig. 5a), we observe the presence of two absolute gaps
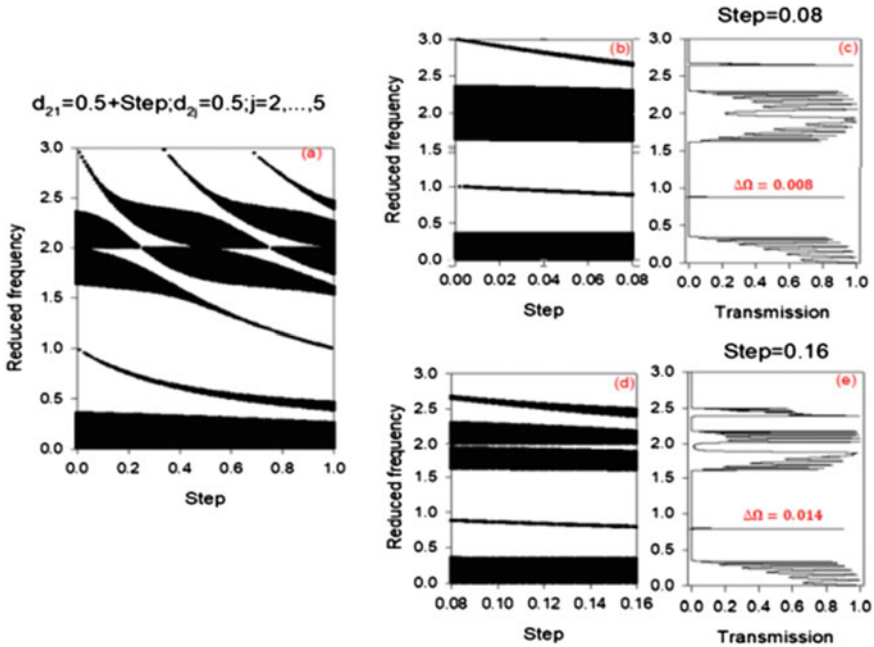
Fig. 4 **a** Variation of the reduced frequency Ω as a function of the length step, **b** and **d** represent the zoom of the case (a), respectively, when the step is varying between 0 and 0.08 (case b) and between 0.08 and 0.16 (case d). **c** and **e** represent the variation of the transmission rate as a function of the reduced frequency Ω for a step equal to 0.08 (case c) and for a step equal to 0.16 (case e)

(transmission equal to zero), one mini-gap (transmission different to zero) around $\Omega = 2$. When we increase the number of resonators N'_2 (Fig. 5b–d), we observe that the mini-gaps observed in the case Fig. 5a ($N'_2 = 1$) become the absolute gap. At the same time, the width of these band gaps becomes larger when N'_2 increases while the pass bands become narrower. This result is identical with the theoretical result found in reference [15], in which we have consider metamaterial symmetric resonators in each site. Figure 5 shows also that the modes located in the second pass band become narrower, and therefore, their quality factors become very important. The frequency shifts between two consecutive modes are very important, which allowed us to deduce that this frequency range associated to the second pass band behaves like a high-performance multichannel system when $N'_2 = 10$. According to these results, we conclude that we can control and manipulate electromagnetic waves through a wider range of frequencies (gaps), when we increase the number of resonators N'_2 and fix the number of other resonators at $N'_1 = 5$.

In this last section, we study in Fig. 6, the effect of the number N'_2 on the transmission rate and the frequency of the flat band created in the first gap observed in Fig. 3a ($N'_1 = 1$). When $N'_2 = 1$, a low transmission rate of this flat band (located between $\Omega = 0.85$ and $\Omega = 0.86$) is observed at this band (transmission cannot exceed 0.03). When we increase N'_2 from 2 to 7, we observe that the transmission rate of
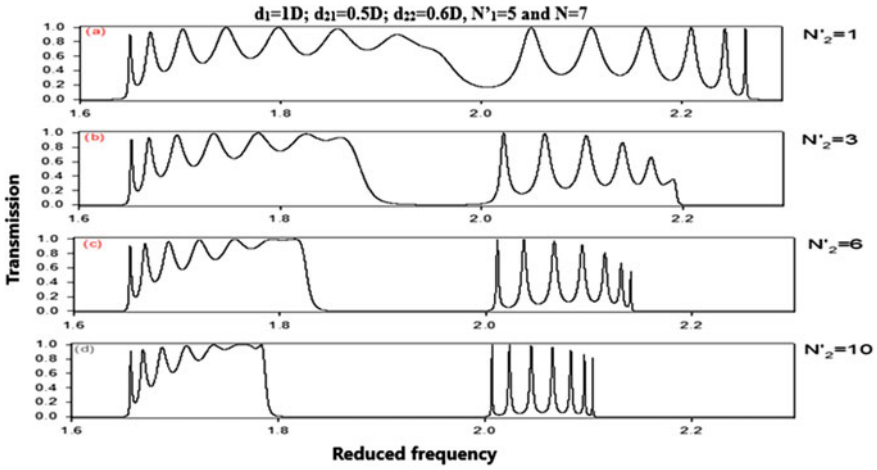
Fig. 5 Variation of the transmission rate versus the reduced frequency Ω belonging to [1.6–2.3] for different values of N'_2. We take $d_1 = 1D$; $d_{21} = 0.5D$; $d_{22} = 0.6D$, $N'_1 = 5$ and $N = 7$

this flat band increases, and at the same time, this flat band moves toward higher frequencies. These narrower flat bands with higher quality factor are important for the enhancement of the unidimensional electromagnetic interaction of materials [10, 11]. In addition, this system may present several applications in guiding, stopping and filtering the electromagnetic waves with higher performance, when the number of one resonator in each site increases.
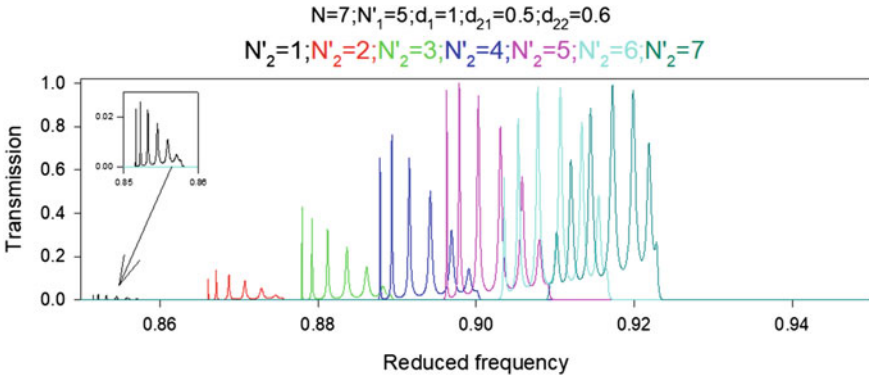
Fig. 6 Variation of the transmission rate versus reduced frequency Ω belonging to [0.85–0.95] for different values of N'_2 ($N'_2 = 1$ black curve; $N'_2 = 2$ red curve, $N'_2 = 3$ green curve, $N'_2 = 4$ blue curve, $N'_2 = 5$ purple curve). We take $d_1 = 1D$; $d_{21} = 0.5D$; $d_{22} = 0.6D$, $N'_1 = 5$ and $N = 7$

3 Conclusions

To sum up, we have investigated the band structure of 1D infinite SWGs and the transmission through finite SWGs. This system is composed of the periodicity of segments and grafted in its extremity by finite number of resonators of different lengths. The lengths of resonators depend on each other following an arithmetic sequence. In the case when the resonators grafted possess the same length (symmetric resonators), we can obtain large gaps and pass bands. These large gaps and pass bands are due to the resonators grafted and the segment periodicity. The width of these gaps becomes very important when we consider the asymmetric resonators in each site, which depend on each other following an arithmetic sequence. In addition, the results show that when we have asymmetric resonators in each site, this system is able to create narrower pass bands (flat bands) in the photonic gaps corresponding to the system that has the same resonators lengths in each site. The number of these flat bands depends on the different parameters of the system, in particular the length of each resonator. Moreover, we have shown that the transmission rate of these flat bands increases when the number of resonators increases. The creation of these narrower flat bands will be useful for telecommunication applications of types: filtering, DFDM and drop channel for the extraction of frequencies. Note that we can find these results by experiments using coaxial cables in the microwave domain.

References

1. Song M, Long C-M, Wu R, Seo D, Leaird D-E, Weiner A-M (2011) Reconfigurable and tunable flat-top microwave photonic filters utilizing optical frequency combs. IEEE Photonique Technol Lett (23):1618–1620
2. Supradeepa V-R, Long C-M, Wu R, Ferdous F, Hamidi E, Leaird D-E, Weiner AM (2012) Comb-based radiofrequency photonic filters with rapid tunability and high selectivity. Nat Photonics 6:186–194
3. Hamidi E, Leaird D-E, Weiner AM (2010) Tunable programmable microwave photonic filters based on an optical frequency comb. IEEE Trans Microw Theory Tech (58):3269–3278 (2010)
4. Chen H-W, Fang A-W, Peters J-D, Wang Z, Bovington J, Liang D, Bowers J-E (2010) Integrated microwave photonic filter on a hybrid silicon platform. IEEE Trans Microw Theory Tech (58):3213–3219
5. Norberg E-J, Guzzon R-S, Parker J-S, Johansson L-A, Coldren L-A (2011) Programmable photonic microwave filters monolithically integrated in InP-InGaAsP. J Lightwave Technol (29):1611–1619
6. Ghoumid K, Mkaoui S, Ouariach A, Cheikh R, Nougaoui A, Gharbi T (2015) Tunable filter based on cavity electro-optic modulation for DWDM applications. Opt Commun 334:332–335
7. Errouas Y, Ben-Ali Y, Ouariach A, Tahri Z, Bria D (2020) Propagation of the electromagnetic waves in one-dimensional asymmetric photonic comb-like structure based on Fibonacci chains of grafted resonators. Mater Today Proc 27:3058–3064
8. Errouas Y, Ben-Ali Y, Tahri Z, Bria D (2020) Propagation of electromagnetic waves in one dimensional symmetric and asymmetric comb-like photonic structure containing defects. Mater Today Proc 31:S16–S23

9. Ben-Ali Y, El Kadmiri I, Tahri Z, Bria D (2020) High quality factor microwave multichannel filter based on multi-defectives resonators inserted in periodic star waveguides structure. Prog Electromagnet Res 104:253–268
10. Ben-Ali Y, Ghadban A, Tahri Z, Ghoumid K, Bria D (2020) Accordable filters by defect modes in single and double negative star waveguides grafted dedicated to electromagnetic communications applications. J Electromagnet Waves Appl 34:539–558
11. Ben-Ali Y, Tahri Z, Ouariach A, Bria D (2019) Double frequency filtering by photonic comb-like. IEEE, pp 1–6
12. Dobrzynski L, Akjouj A, Djafari-Rouhani B, Vasseur JO, Zemmouri J (1998) Giant gaps in photonic band structures. Phys Rev B (57):R9388
13. Ben-Ali Y, Tahri Z, Bria D (2019) Electromagnetic filters based on a single negative photonic comb-like. Prog Electromagnet Res C 92:41–56
14. Errouas Y, Tahri Z, Ben-Ali Y, Bria D (2019) Band structure and photonic band gap in comb-like structure asymmetric with left-handed materials. In: 2019 international conference on intelligent systems and advanced computing sciences (ISACS). IEEE, pp 1–5
15. Yin CP, Wang HZ (2009) Narrow transmission bands of quasi-1D comb-like photonic waveguides containing negative index materials. Phys Lett A 11:1093–1096

Binding Energy of an Off-Center Shallow Donor Impurity in Wedge-Shaped Quantum Dot Under Electric Field Effect

Mohamed Chnafi, Omar Mommadi, Soufiane Chouef, Reda Boussetta, Laaziz Belamkadem, Mohammed Hbibi, Abdelaziz El Moussaouy, Juan Alejandro Vinasco, Carlos Alberto Duque, and Farid Falyouni

Abstract In the present work, we have investigated the influence of the axial electric field on an electron-donor atom confined in a GaAs wedge-shaped quantum dot. To this end, we used the finite difference method to solve the Schrödinger equation. The binding energy and the average value of the electron position were calculated for different quantum dot radius, electric fields, and impurity positions. From the results obtained, it can be seen that the electron-impurity binding increases with increasing electric field intensity in low confinement and almost constant in high confinement, and also, we noticed that the binding energy is very sensitive to the displacement of the donor atom, it is maximum when the impurity is located in the center of the quantum dot (QD). These different effects give a great vision to understand the behavior of the electron-impurity interactions in this type of nanostructures.

Keywords Wedge-shaped quantum dot impurity · Electric field · Binding energy

1 Introduction

Researchers have been paying close attention to low-dimensional semiconductor devices with specific geometrical patterns in recent years. Quantum pseudo-dots, double ring-shaped quantum dots, wedge-shaped quantum dots (WSQD), hexagonal-shaped quantum wires, etc. [1–6]. These structures have the ability to create new

M. Chnafi · O. Mommadi (✉) · S. Chouef · R. Boussetta · L. Belamkadem · M. Hbibi · A. El Moussaouy · F. Falyouni
Laboratory of Materials, Waves, Energy and Environment, OAPM Group, Department of Physics, Faculty of Sciences, Mohamed I University, 60000 Oujda, Morocco
e-mail: omommadi@gmail.com

A. El Moussaouy
Laboratory of Innovation in Science, Technology and Education, Regional Centre for the Professions of Education and Training, 60000 Oujda, Morocco

J. A. Vinasco · C. A. Duque
Grupo de Materia Condensada-UdeA, Facultad de Ciencias Exactas y Naturales, Instituto de Física, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia

phenomena and have applications in laser and optical modulation technologies. The behavior of their electrical and optical characteristics in the presence of an external electric field, in particular, has been extensively investigated, both theoretically and experimentally [7–14]. A good way to study impurity atoms and a method to control their optoelectronic states is the application of an external perturbation (electric field). For example, the electric field weakens the electron-impurity bond as in the hydrogen atom. When the field intensity increases, the absorption peak in the absorption spectrum shifts toward the low frequencies (long wavelengths), the intensity of the peak attenuating. The confined quantum Stark effect is the name given to this issue. Several experimental and theoretical types of research have been conducted on this phenomenon.

Using a parabolic confining potential, the effects of temperature, hydrostatic pressure, and external electric fields on the confined exciton in cylindrical quantum dots are examined by Mommadi et al. [15]. They have obtained that the binding energy and interband emission is very sensitive to the electrical field, and they have found that the existence of a dipole moment is demonstrated by the behavior of the excitonic Stark shift as a function of the applied electric field. By using a variational and perturbative approach, El Khamkhami et al. [16] have found that the polarizability of a donor enclosed in a spherical quantum dot. Considering the perturbation approach and parabolic lateral confinement, Xie [17] has evaluated the binding energy for many states. It is assumed that the electric field is applied along the Ox-axis. He concluded that the confinement effect and electric field have a significant impact on the optical performance of the quantum disk. Wei et al. [18] calculated the binding energy of donor atom confined in coupled ZnO/MgZnO quantum dots by considering the effect of the internal electric field due to the piezoelectric and spontaneous polarizations of the material. The results of the calculations show that the binding energy of the impurity depends on the position of the impurity, the structural parameters of the coupled quantum dots such as the radius and the x-magnesium composition. The strong internal electric field also produces an asymmetric distribution of impurity binding energy toward the center of the coupled QDs, which is observed.

However, new theoretical investigations on the effect of an axial electric field on (WSQD) suggest that the photonic properties of a specific QD can be controlled for the emission of entangled photon pairs. We will investigate the influence of an axial electric field on a donor confined in a WSQD in connection with this work, to better understand the phenomena connected to the application of an axial electric field on WSQD.

The following is a paper's structure. The theoretical model for a shallow donor impurity confined in a WSQD of GaAs type, surrounded by an assumed infinite potential barrier, is presented in Sect. 2. The results of our calculations and discussion are presented in Sect. 3. In Sect. 4, you will find the conclusions.

2 Theory

We consider an electron-donor atom confined in a WSQD of GaAs type, surrounded by an assumed infinite barrier, with the geometrical parameters: R, θ_0, and H, which describe, respectively, the radius, the geometrical angle, and the height of QD, all we keep the height fixed at $H = 5$ nm and the angle of the nano-system 45°, the illustration of the WSQD is presented on Fig. 1.

In the presence of a shallow donor impurity, the basic Hamiltonian of the electron-impurity system in WSQD can be described in the effective mass approximation by

$$H_D = -\frac{\hbar^2}{2m_e^*}\left(\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial z^2}\right) + V_D(\rho, \theta, z) + V_C(\rho, \theta, z) + W, \tag{1}$$

where the first term represents the kinetic energy and $m_e^* = 0.067m_0$ is the electron masse effective for a GaAs semiconductor, m_0 is the free electron masse, the shallow impurity is fixed in ($\theta_i = 0°$ and $z_i = H/2$) and the second terms of Eq. 1 is the coulombic interaction between electron-impurity is given by

$$V_D(\rho, \theta, z) = \frac{-e^2}{4\pi\varepsilon_0\varepsilon_r r_{i-e}}, \tag{2}$$

where e, ε_0, and ε_r are the electron charge, vacuum permittivity, and dielectric constant of GaAs, respectively. r_{i-e} represents the distance between electron and impurity which can be expressed by

$$V_D(\rho, \theta, z) = \frac{-e^2}{4\pi\varepsilon_0\varepsilon_r\sqrt{\left(\rho^2 + r_i^2 - 2\rho r_i\cos(\theta)\right) + (z - z_i)^2}} \tag{3}$$

Fig. 1 Illustration for a confined electron and an off-center donor atom in WSQD defined by the radius R, the height $H = 5$ nm, and $\theta_0 = 45°$
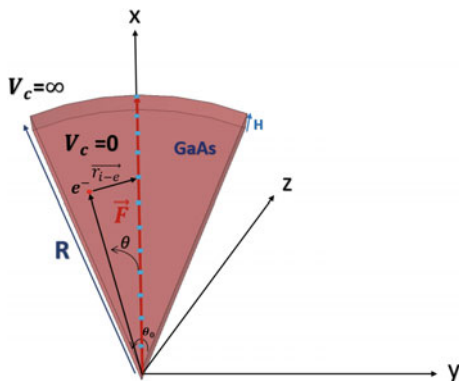
The third term in Eq. 1 represents the infinite confinement potential which can be written as

$$V_c(\rho, \theta, z) = \begin{cases} 0 & \text{if } 0 \leq \rho \leq R \quad 0 \leq \theta \leq \theta_0 \text{ and } 0 \leq z \leq H \\ \infty & \text{otherwise} \end{cases}, \qquad (4)$$

where outside the surface of the WSQD, the probability density of the electron is zero. As a result, the wave function is also zero. The contribution of the external electric field is given by last term of Eq. 1 $W = -eF\vec{\rho} = -eF\rho\cos(\theta)\cos(\varphi_F)$ with $\varphi_F = 0°$. The eigenvalue equation of the system can be expressed as follows:

$$\left(-\frac{\hbar^2}{2m_e^*}\left(\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial z^2}\right) + V_D(\rho, \theta, z)\right.$$
$$\left. + V_C(\rho, \theta, z) + W\right)\psi_D(\rho, \theta, z) = E_D\psi_D(\rho, \theta, z) \qquad (5)$$

where $\psi_D(\rho, \theta, z)$ and E_D are, respectively, the eigenfunctions and eigenenergies of a single donor confined in the WSQD. In absence of the shallow impurity donor, the eigenvalue equation of the system becomes as following:

$$\left(-\frac{\hbar^2}{2m_e^*}\left(\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial z^2}\right) + V_C(\rho, \theta, z) + W\right)\psi_0(\rho, \theta, z)$$
$$= E_0\psi_0(\rho, \theta, z) \qquad (6)$$

where E_0 presents the electron energy and $\psi_0(\rho, \theta, z)$ is the electron eigenfunction in the ground state. By the difference between the electron energies with and with and without impurity, we obtained the ground state binding energy as follows:

$$E_B = E_0 - E_D \qquad (7)$$

Equation 5 represents the three-dimensional second-order differential equation. We used the finite difference method (FDM) to solve this equation, which involves replacing partial derivatives with divided differences or combinations of function point values in a finite number of discrete points or nodes of the mesh. Their advantage is the ease with which they can be written and the low computational cost. We have discretized the Eq. 5 on 3D containing $T_\rho * T_\theta * T_z$ grid points by using the FDM. We divide the $[0, R]$, $[0, \theta_0]$, and $[0, H]$, respectively, into $(T_\rho + 1)$, $(T_\theta + 1)$, and into $(T_z + 1)$ parts. We consider $1 \leq l \leq T_\rho$, $1 \leq m \leq T_\theta$ and $1 \leq n \leq T_z$ are, respectively, the integers identifying the positions of the grid points along the WSQD directions $\vec{e}_\rho$, $\vec{e}_\theta$, and $\vec{e}_z$, as well as the spaces between the grids along the surface are presented by $\Delta\rho$, $\Delta\theta$, and Δz with $\rho = l\Delta\rho$, $\theta = m\Delta\theta$, and $z = n\Delta z$. Using the first- and the second-order central difference approximation, respectively, for $\frac{\partial}{\partial\rho}$; $\frac{\partial^2}{\partial\rho^2}$, $\frac{\partial^2}{\partial\theta^2}$, and $\frac{\partial^2}{\partial z^2}$, we have obtained the discretized equation at a grid point (l, m, n) in the following form:

$$
\begin{aligned}
\frac{-\hbar^2}{2m_e^*}\Bigg[&\left(\frac{1}{2l(\Delta\rho)^2}+\frac{1}{(\Delta\rho)^2}\right)\psi_{l+1,m,n}\\
&-\left(\frac{2}{(\Delta\rho)^2}+\frac{2}{l^2(\Delta\rho)^2(\Delta\theta)^2}+\frac{2}{(\Delta z)^2}\right.\\
&\left.-\frac{2m_e^*}{\hbar^2}\left(V_D^{l,m,n}+W^{l,m,n}+V_c^{l,m,n}\right)\right)\psi_{l,m,n}\Bigg]\\
+\frac{-\hbar^2}{2m_e^*}\Bigg[&\left(\frac{-1}{2l(\Delta\rho)^2}+\frac{1}{(\Delta\rho)^2}\right)\psi_{l-1,m,n}\\
&+\frac{1}{l^2(\Delta\rho)^2(\Delta\theta)^2}\psi_{l,m-1,n}+\frac{1}{l^2(\Delta\rho)^2(\Delta\theta)^2}\psi_{l,m+1,n}\Bigg]\\
+\frac{-\hbar^2}{2m_e^*}\Bigg[&\frac{1}{(\Delta z)^2}\psi_{l,m,n-1}+\frac{1}{(\Delta z)^2}\psi_{l,m,n+1}\Bigg]=E_D\psi_{l,m,n}
\end{aligned}
\tag{8}
$$

where $\Delta\rho = R/(T_\rho + 1)$, $\Delta\theta = \theta_0/(T_\theta + 1)$, and $\Delta z = H/(T_z + 1)$ are the pitch (the distance between two consecutive grid nodes in the $\overrightarrow{e_\rho}$, $\overrightarrow{e_\theta}$, and $\overrightarrow{e_z}$ direction, respectively). $\left(V_D^{l,m,n},\ V_c^{l,m,n},\ \text{and } W^{,m,n}\right)$ are the coulombic interaction between the electron and the donor, confinement potential, and contribution of the external electric field, respectively, at point (l, m, n). The energy eigenvalues of the electron in the presence and the absence of the impurity are obtained by numerically solving Eqs. 5 and 6.

3 Numerical Results and Discussion

The problem we studied depends on several parameters: the intensity of the electric field, the position of the donor atom, and the dimensions of the quantum dot. The angle and thickness are fixed: $\theta_0 = 45°$ and $H = 5\,\text{nm}$.

In Fig. 2, we have illustrated the variation of the binding energy as a function of the WSQD radius for different intensities of the electric field (0, 20 and 40 kV/cm). It can be noted that for small values of the radius ($R \leq 9\,\text{nm}$), the effect of quantum confinement is predominant in relation to the effect of the electric field, on the other hand, for large values of the QD radius ($R \geq 9\,\text{nm}$), the electric field tends to enlarge the electron-impurity distance and the WSQD behaves like a massive semiconductor, in this case, the carriers of the charges are free to move in the three directions of WSQD.

In order to fully understand the above effects, Fig. 3 shows that the numerical results of the impurity binding energy as a function of the intensity of the electric field, for three values of the radius ($R = 10, 20,$ and 30 nm) of the quantum dot and for a given values of the angle and the height. We have noticed that for the low confining regimes (large WSQD size) the electron-impurity bound followed an increasing function with the intensity of the electric field, this effect is due to the
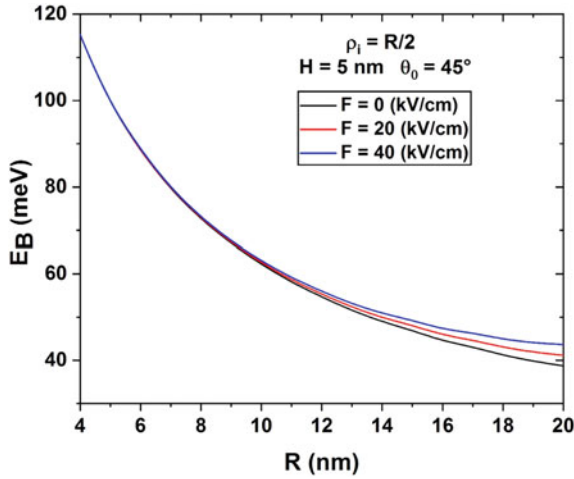
Fig. 2 Variation of the impurity ground state binding energy as a function of radius of GaAs quantum dot. The impurity is fixed in the symmetry axis at $\rho_i = \frac{R}{2}$ with $H = 5$ nm and $\theta_0 = 45°$

electric field intensity which tends to diminish the impurity electron distance. For the large confinement regime (large WSQD size), the electron-impurity bound is presupposed insensitive to the electric field intensity due to the predominance of quantum confinement in relation to the applied electric field.
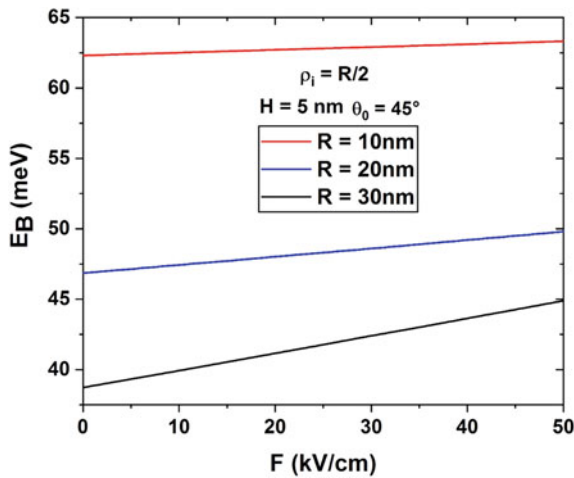
Fig. 3 Variation of the impurity ground state binding energy as a function of the intensity of the electric field in GaAs quantum dot, the impurity is fixed in the symmetry axis at $\rho_i = \frac{R}{2}$ with $H = 5$ nm and $\theta_0 = 45°$

In Fig. 4, we have illustrated the variation of the impurity binding energy as a function of the normalized position of the impurity along the axis of symmetry, considering three values of the electric field intensity. Our results show that the binding energy is maximal in the region of the center of gravity. This behavior will be close to the impurity that is localized near to the gravity center of WSQD edge. As the impurity moves toward the WSQD extremity ($\frac{\rho_i}{R} \to 0$ or $\frac{\rho_i}{R} \to 1$), the average electron-impurity distance increases, leading to a decrease in the Coulombic interaction and corresponding binding energy. We then analyze the oscillating behavior shown by these curves, the electric field induces double comportment on the binding energy, and they strengthen the coulombic interaction in small sizes which produces an increase in the binding energy, and entrain an increase in the average electron-impurity distance which implies a decrease in the electron-impurity binding. It is noted that this figure is consistent with the results obtained in Figs. 2 and 3. As well as this comportment is in agreement with the results obtained by Mohajer et al. [19] in the case of hemispherical quantum dot and Chnafi et al. [4] in the case of a conical quantum dot.

In order to understand the effect of the electric field on the average electron cloud, we have examined in Fig. 5 the average electron distance from the origin of our structure as a function of the quantum dot radius for three values of the electric field intensity ($F = 0, 20$, and 40 kV/cm), the position of the impurity is fixed at the center of the structure's gravity $\left(\rho_i = \frac{R}{2}\right)$. We found that the average value of the electron distance increases linearly with WSQD radius. Also, the electric field effect is important in the low confinement regime which is in argument with the results presented above and in good agreement with the results found by Mohajer [19].
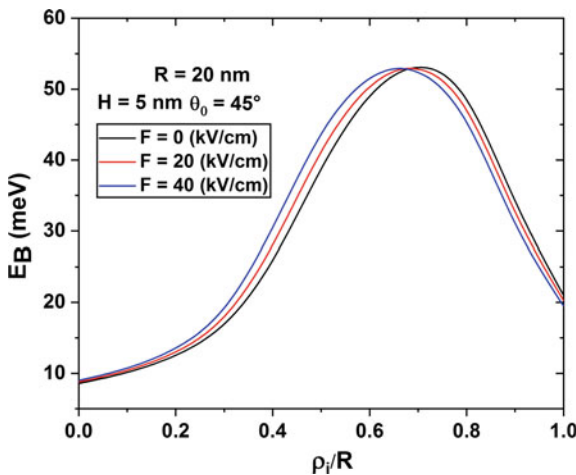
Fig. 4 Variation of the impurity ground state binding energy as function of normalized impurity position along the symmetry axis in GaAs quantum dot, the results are considering for three values of the intensity of electric field with $H = 5$ nm and $\theta_0 = 45°$
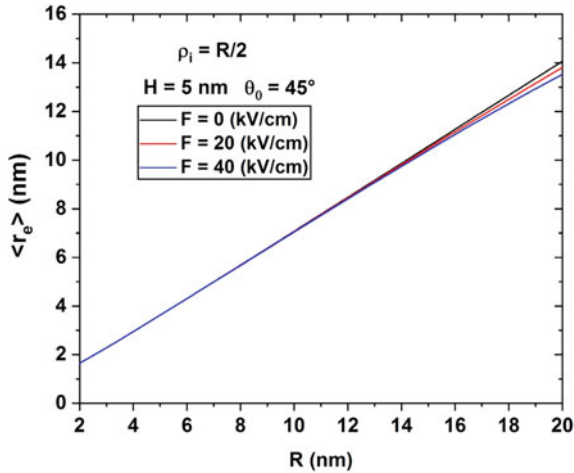
Fig. 5 Variation of the average electron as a function of radius of GaAs quantum dot, the impurity is fixed in the symmetry axis at $\rho_i = \frac{R}{2}$ with $H = 5$ nm and $\theta_0 = 45°$

4 Conclusion

In this work, we have exposed the numerical method solution (finite difference method) of the Schrödinger equation to an electron-donor atom confined in a WSQD, by using the effective mass approximation and an infinite square confinement potential. We have given the expression of the discriminated Hamiltonian while discussing each term and the boundary conditions. Thus, we have studied the effect of an axial electric field on electron-impurity binding energy in a specific structure (WSQD). We have obtained that the impurity binding energy is very sensitive to the combined effects of WSQD size and electric field. The impurity binding energy varies linearly with the electric field, and it follows a forte increasing function in the low confinement regimes. Which enhance the non-dissociation of the quasi-particles and improve their radiative lifetime. Therefore, it can be concluded that the effect of the electric field leads to a shift of the energy peak toward the blue. Thus, we have deduced that the binding energy strongly depends on the position of the donor atom. These influences allow us to adjust the optoelectronic properties of nano-systems.

References

1. Hartmann H, Schuck R (1980) Spin-orbit coupling for the motion of a particle in a ring-shaped potential. Int J Quant Chem 18:125
2. Khordad R, Bahramiyan H, Pramana (2017) Effects of electron–phonon interaction and impurity on optical properties of hexagonal-shaped quantum wires. J Phys 88:50

3. Holovatsky VA, Chubrei MV, Duque CA (2022) Core-shell type-II spherical quantum dot under externally applied electric field. Thin Solid Films 747:139142
4. Chnafi M, Belamkadem L, Mommadi O, Boussetta R, El Hadi M, El Moussaouy A et al (2021) Hydrostatic pressure and temperature effects on spectrum of an off-center single dopant in a conical quantum dot with spherical edge. Superlattices Microst 159:107052
5. Chnafi M, El Moussaouy A, Mommadi O, Belamkadem L (2020) Energy and stability of negatively charged trion in cylindrical quantum dot under temperature effect. Phys B 594:412333
6. Iqraoun E, Sali A, Rezzouk A, Feddi E, Dujardin F, Mora-Ramos ME, Duque CA (2017) Donor impurity-related photoionization cross section in GaAs cone-like quantum dots under applied electric field. Philos Mag 97:1445–1463
7. Tulupenko V, Akimov V, Demediuk R, Tiutiunnyk A, Duque CA, Sushchenko D, Laaroze D (2020) Hydrogenic impurity states in a delta-layer within quantum wells in a transversal electric field. In: IEEE 40th international conference on electronics and nanotechnology
8. Ungan F, Martínez-Orozco JC, Restrepo RL, Mora-Ramos ME, Duque CA (2019) Effect of applied electric field on the nonlinear optical properties of modulation-doped GaAs/Al$_x$Ga$_{1-x}$As double quantum well. Superlattices Microst 126:89–97
9. Al EB, Kasapoglu E, Sakiroglu S, Duque CA, Sokmen I (2018) Binding energy of donor impurity states and optical absorption in the Tietz-Hua quantum well under an applied electric field. J Mol Struct 1157:288–291
10. Strozecka A, Jaskolski W, Zielinski M, Bryant GW (2004) Stark effect in semiconductor nanocrystals: tight-binding approach. Vacuum 74:259
11. Harotyunyan VA, Aramyan KS, Petrosyan HSh, Demirjian GH (2004) Stark effect in a wedge-shaped quantum box. Phys E 24:173
12. Belamkadem L, Mommadi O, Vinasco JA, Laroze D, El Moussaouy A, Chnafi M, Duque CA (2021) Electronic properties and hydrogenic impurity binding energy of a new variant quantum dot. Phys E 129:114642
13. Feng YP, Tan HS, Spector HN (1995) Quantum well excitons in an electric field: two versus three-dimensional behavior. Superlattices Microstr 17:267
14. Mommadi O, El Moussaouy A, Chnafi M, El Hadi M, Nougaoui A, Magrez H (2020) Exciton–phonon properties in cylindrical quantum dot with parabolic confinement potential under electric field. Phys E 118:113903
15. Mommadi O, El Moussaouy A, El Hadi M, Chnafi M, Meziani YM, Duque CA (2021) Stark shift and exciton binding energy in parabolic quantum dots: hydrostatic pressure, temperature, and electric field effects. Philos Mag 101:753–775
16. El Khamkhami J, Feddi E, Assaid E, Dujardin F, Stébé B, Diouri J (2001) Low magnetic field effect on the polarisability of excitons in spherical quantum dots. Phys Scr 64:504
17. Xie W (2009) Nonlinear optical rectification of a hydrogenic impurity in a disc-like quantum dot. Phys B 404:4142–4145
18. Wei S, Chang Q, Zeng Z (2011) Hydrogenic impurity states in wurtzite symmetric ZnO/MgZnO coupled quantum dots. Curr Appl Phys 11:16–20
19. Mohajer SN, Ibral A, El Khamkhami J, Assaid EM (2018) Quantum confined Stark effects of single dopant in polarized hemispherical quantum dot: two-dimensional finite difference approach and Ritz-Hasse variation method. Phys B 537:40–50

Electronics and Biomedical Applications

Design and Optimization of a WPT System for Powering Biomedical Implants

Brahim Ouacha, Hamid Bouyghf, Mohammed Nahid, and Said Abenna

Abstract With inductive coupling, wireless power transfer (WPT) is possible. Transmission of electricity without the need of cables has become one of the most advanced approaches for powering biomedical implants in recent decades. The key elements (indicators) of a wireless power transfer system are PTE and PDL. The geometrical dimensions of the coils, the distance between transmitter (TX) and receiver (RX), and the operating frequency are all reliant on these keys in the WPT system. This work focuses on the creation, design, and optimization of coils square spirals in a wireless energy transfer system employing a resonant inductive link. Among the optimization approaches employed are metaheuristic algorithms.

Keywords Inductive coupling (IP) · Wireless power transfer (WPT) · Metaheuristic algorithms · Power transfer efficiency (PTE) · Power delivered to load (PDL)

1 Introduction

Due to the requirement for invasive surgery to replace the battery, implantable devices cannot employ a large-capacity battery as the principal energy source in long-term investigations. In recent years, WPT has grown in popularity, and it is now routinely used in near-field inductive coupling systems. Induction is now used to power the bulk of bio-implanted devices [1–3]. Power transmission to biomedical implants is often accomplished by inductive coupling [2]. This method eliminates the need for transcutaneous connections, which can cause infection, as well as implantable batteries, which must be replaced after a set amount of time [4, 5]. Power transfer efficiency (PTE) and power transferred to load (PDL) are two significant parameters (indicators) in wireless power transfer systems. The size of the transceiver (geometric

B. Ouacha (✉) · H. Bouyghf · M. Nahid · S. Abenna
LSIB Laboratory, Faculty of Sciences and Techniques, Hassan II University of Casablanca, Mohammedia, Morocco
e-mail: brahim.ouacha-etu@etu.univh2c.ma

characteristics), the separation distance, and the operating frequency all influence these two variables. The problem resurfaces as an optimization problem, requiring the determination of ideal values for parameters that are dependent on PTE and PDL in order to maximize these two parameters. This is our contribution in this paper. In [6], According to Kiani et al., a WPT system consists of two transmitting and receiving coils as well as a repeater. In other words, a three-coil system. Despite increasing the separation distance to 120 mm and using a resonance frequency of $f = 13.56$ MHz, the transfer efficiency dropped to 55% and the PDL power dropped to only 83 mW.

In [7], the author showed and contrasted a multicoil wireless power transmission system to systems with two, three, and four coils. He claims that the four-coil approach yields superior PTE outcomes, such as 66.7% for a 200 mm distance. On the other hand, the usage of repeated coils exacerbates the disadvantages, such as infections and implant size distribution. However, as shown in [8], using a multicoil inductively coupled array can help achieve better PTE and PDL. The authors were able to raise PTE and PDL to 76% and 115 mW, respectively.

The rest of the article is organized as follows: Sect. 1 depicts the WPT's introduction as well as the inductive coupling transfer technique, as well as some similar works that influenced the development of the suggested technique. Section 2 discusses the theoretical underpinning of inductive coupling. Section 3 discusses metaheuristic algorithms, while Sect. 4 compares the DEA and GA results, and Sect. 5 wraps up the paper.

2 Theoretical Background

2.1 Energy Transfer System

Thanks to inductive coupling, electrical energy can be distributed without the use of any material support. Implantable devices, for example. A simple instance of a two-coil inductive link is shown in Fig. 1. The parasitic resistance of coil windings is a defining characteristic. The primary and secondary coils are designated by the letters L1 and L2, respectively (R1, R2).

2.2 Expressions of Objective Functions PTE and PDL

The primary goal of implanted device developers is to improve power transfer efficiency (PTE) and delivered to load (PDL). It is enough to maximize the quality and coupling variables in order to maximize this coefficient. Also, both members of the system should have the same resonance frequency. As a result, the following equation may be used to calculate the PTE:

Fig. 1 Equivalent schematic of a two-coil inductive connection
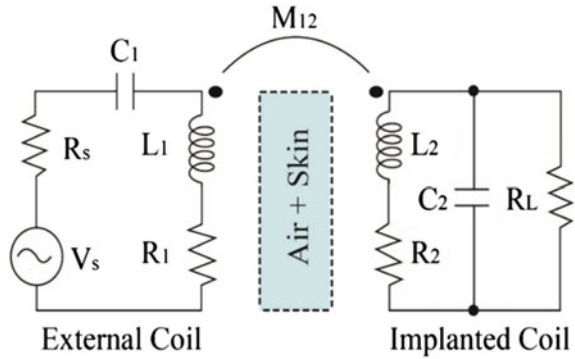
$$\text{PTE} = \frac{k_{12}^2 Q_1 Q_{2L}}{1 + k_{12}^2 Q_1 Q_{2L}} * \frac{Q_{2L}}{Q_L} \tag{1}$$

where $Q_{2L} = \frac{Q_L Q_2}{Q_L + Q_2}$, $Q_L = \frac{R_L}{w L_2}$, $k_{12} = \frac{M_{12}}{\sqrt{L_1 L_2}}$.

As shown in the equation below, the PDL is a quantity connected to the PTE.

$$\text{PDL} = \frac{V_s^2}{R_s} * \text{PTE} \tag{2}$$

where

V_s: source voltage.

R_s: resistance to the source.

3 Metaheuristic Algorithms

Many metaheuristic algorithms have been successfully used to solve difficult situations [10, 11]. Differential evolution algorithm (DEA) and genetic algorithm (GA) are heuristic search methodologies that can be used to solve a variety of optimization problems. They are appealing for many optimization issues in practice because of their versatility [12].

DEA is a part of a stochastic and self-organizing metaheuristic search method. Because DE employs real integers as solution strings, no encoding or decoding is necessary. The population of a DE algorithm is randomly initialized within the initial parameter bounds. The three basic processes employed in optimization are selection, crossover, and mutation [13, 14]. Classical optimization approaches are not the same as genetic algorithms. GA simulates the biological evolution of chromosomes. There
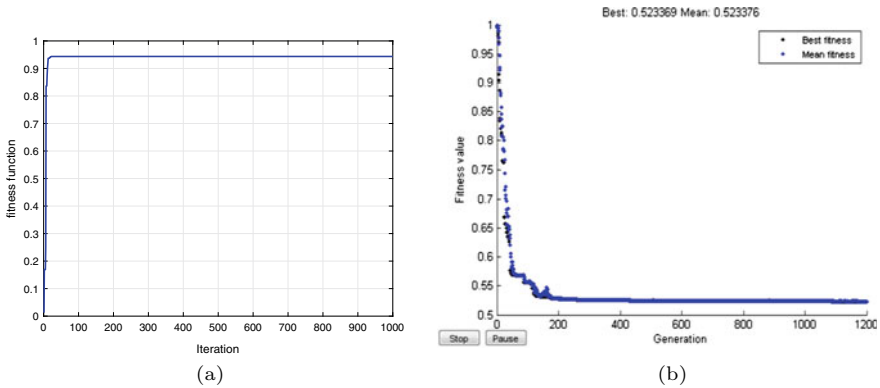
Fig. 2 Objective function (PTE) versus number of iterations by DEA (**a**) and inverse of efficiency (PTE) versus generation by applying GA (**b**)
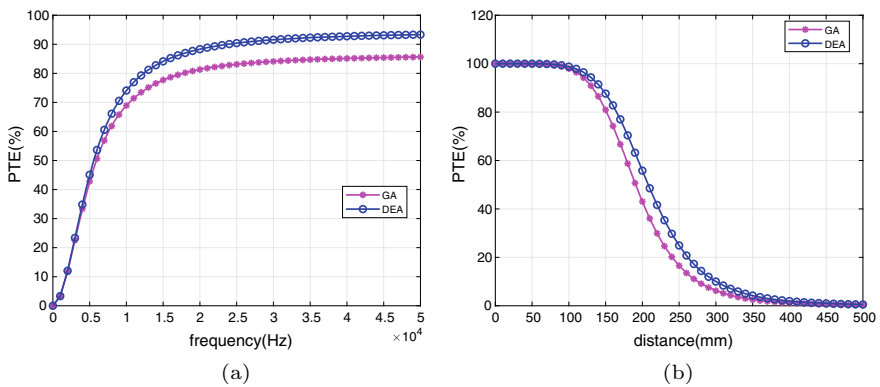
Fig. 3 Objective function (PTE) versus frequency (**a**) and separation distance (**b**)

are also different genetic operators in genetic algorithms like DEA [15] (selection, crossover, and mutation). Figure 2 depicts the method's convergence toward the optimal value as iterations progress.

4 Simulation Results

MATLAB software was used to simulate these two technique proposals. Figure 3 shows the results of the two methods.

The power transfer efficiency changes as a function of operating frequency and separation distance between the two connected coils, as shown in Fig. 3a and b. The graph's objective is to show Fig. 4 depicts the effect of transmission distance on the quantity of energy delivered to the load (implant).

Fig. 4 Power delivered to load (PDL) versus separation distance
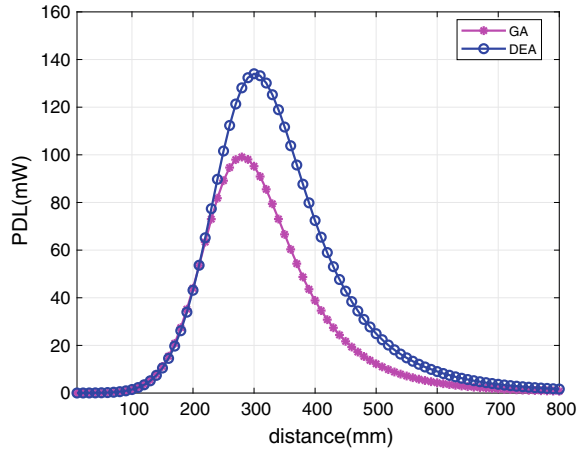
4.1 Results and Discussion

When compared to the genetic algorithm, the differential evolutionary algorithm (DEA) shows a significant improvement in power transfer efficiency and power delivered to load over long separation distances. The ability of the differential evolutionary technique to avoid striking local maximums when hunting for an ideal number enabled this development (maximum PTE).

5 Conclusion

These two metaheuristics approaches successfully optimize where to discover the ideal values of the geometric parameters of the two connected coils, separation distance, and operating frequency in order to get the highest PTE and PDL values. In the future, we will concentrate on improving the PTE and PDL at a longer distance while also reducing the implant's size.

References

1. Mehri S, Ammari AC, Ben Hadj Slama J, Rmili H (2016) Geometry optimization approaches of inductively coupled printed spiral coils for remote powering of implantable biomedical sensors. J Sens 2016
2. Song S, Konijnenburg M, van Wegberg R, Xu J, Ha H, Sijbers W, van Helleputte N (2019) A 769 µW battery-powered single-chip SoC with BLE for multi-modal vital sign monitoring health patches. IEEE Trans Biomed Circ Syst 13(6):1506–1517

3. Lakhdari A. Développement d'un systéme de transfert d'énergie sans fil: application au domaine biomédical. Doctoral dissertation, Université Mohamed Boudiaf des Sciences et de la Technologie-Mohamed Boudiaf d'Oran

4. Duan Z, Guo YX (2011) Rectangular coils modeling for inductive links in implantable biomedical devices. In: 2011 IEEE international symposium on antennas and propagation (APSURSI). IEEE, pp 388–391

5. Mirbozorgi SA (2015) High-performance wireless power and data transfer interface for implantable medical devices

6. Kiani M, Jow UM, Ghovanloo M (2011) Design and optimization of a 3-coil inductive link for efficient wireless power transmission. IEEE Trans Biomed Circ Syst 5(6):579–591

7. Kiani M, Ghovanloo M (2012) A figure-of-merit for designing high-performance inductive power transmission links. IEEE Trans Ind Electron 60(11):5292–5305

8. Mirbozorgi SA, Bahrami H, Sawan M, Gosselin B (2014) A smart multicoil inductively coupled array for wireless power transmission. IEEE Trans Ind Electron 61(11):6061–6070

9. Mehri S, Slama JBH, Ammari AC, Rmili H (2014) Genetic algorithm based geometry optimization of inductively coupled printed spiral coils for remote powering of electronic implantable devices. In: 2014 global summit on computer & information technology (GSCIT). IEEE, pp 1–6

10. Bouyghf H, Benhala B, Raihani A (2019) Analysis of the impact of metal thickness and geometric parameters on the quality factor-Q in integrated spiral inductors by means of artificial bee colony technique. Int J Electr Comput Eng 9(4):2918

11. Heidarian M, Burgess SJ (2020) A design technique for optimizing resonant coils and the energy transfer of inductive links. IEEE Trans Microwave Theory Tech 69(1):399–408

12. Kramer O (2017) Genetic algorithms. In: Genetic algorithm essentials. Springer, Cham, pp 11–19

13. Varadarajan M, Swarup KS (2008) Differential evolution approach for optimal reactive power dispatch. Appl Soft Comput 8(4):1549–1561

14. Moreno L, Garrido S, Muñoz ML (2006) Evolutionary filter for robust mobile robot global localization. Robot Auton Syst 54(7):590–600

15. Abdel-Basset M, Abdel-Fatah L, Sangaiah AK (2018) Metaheuristic algorithms: a comprehensive review. Computational intelligence for multimedia big data on the cloud with engineering applications, pp 185–231

Prediction of Port A Cath Complications Using Machine Learning Techniques

Hanane El Oualy [ORCID]**, Hajji Bekkay, Adel Mellit, Mouhsine Omari, Kamal Ahsayan, and Hamid Madani**

Abstract The use of the Port A Cath in chemotherapy is increasing. It improves the management of cancer patients by giving them a permanent and secure administration route, and although the placement of an implantable port is a common practice, it is sometimes dangerous because of the significance of certain complications that can range from a simple hematoma to septic shock threatening the patient's vital prognosis. It is necessary therefore to have a reliable, accurate, and feasible system to detect these complications in time for appropriate management. Machine learning (ML) techniques have been implemented on diverse medical datasets to facilitate the analysis of large and complex data. Recently, many scientists have used various machine learning techniques to help the medical industry and professionals to diagnose diseases, but rarely to predict complications related to Port A Cath. The main objective of this work is to examine and compare the accuracy of different data mining classifications, applying a set of ML algorithms, for the prediction of Port A Cath complications. The dataset is built by conducting a retrospective study at the Hassan II Oncology Center in Oujda (Morocco). The models are developed using the Python language. The investigated ML techniques are support vector machine (SVM), decision tree (DT), random forest (RF), and logistic regression (LR). The experimental results indicate that LR method performs better than other investigated ML techniques.

Keywords Port A Cath · Complications · Machine learning algorithms

H. El Oualy (✉) · H. Bekkay · H. Madani
Laboratory of Energy, Embedded System and Information Processing, National School of Applied Sciences, Mohammed First University, B.P. 665, 60000 Oujda, Morocco
e-mail: Hanane.eloualy@ump.ac.ma

A. Mellit
Laboratory of Renewable Energy, University of Jijel, Jijel, Algeria

M. Omari · K. Ahsayan · H. Madani
Laboratory of Medical Oncology, Faculty of Medicine and Pharmacy, Mohammed First University, B.P. 665, 60000 Oujda, Morocco

1 Introduction

The Port A Cath, also known as a totally implanted central venous access (TIVAPS), is a small reservoir connected to a venous catheter and placed in the subcutaneous tissue. Its use began in the early 1980s in oncology patients, and today, these devices are an integral part of daily clinical routine [1]. The complications of Port A Cath related to their placement and long-term use are still a matter of concern. These complications, such as infection, catheter fracture thrombosis, and extravasation, may require device replacement, resulting in additional stress for the patient and delays in treatment. Cancer patients have an inherently high risk of venous thrombosis, both from the disease itself and from certain cancer treatments. This risk is further increased by the placement of a catheter or TIVAPS [2–4].

In this paper, we use machine learning algorithms to predict the presence or absence of Port A Cath complications. The dataset used to build the predictive models is from a retrospective study conducted at the Hassan II Oncology Center Fig. 1. The data is loaded in CSV format and cleaned for use. After data crushing and attribute selection, machine learning (ML) algorithms including support vector machine (SVM), decision tree (DT), random forest (RF), and logistic regression (LR) are used, and then, the accuracy is compared to select the best model. All the analysis and visualizations are performed in Python language.

This paper is organized as follows: Data collection and methodology are described in Sect. 2. Experimental results and discussion are given in Sect. 3. Concluding remarks are reported in the last Sect. 4.

Fig. 1 Implantation of the Port A Cath

Table 1 Extract of the dataset

Age	Gender	Cancer types	WHO	Pathway	PU	Difficulty	Early complications	Late complications
30	Woman	Ovarian	0	Right	6 days	6	Absence of complications	Absence of complications
56	Man	Gastric	1	Left	20 days	3	Absence of complications	Absence of complications
71	Man	Colon	3	Left	7 days	3	Thrombosis	Absence of complications
85	Woman	Breast	4	Left	15 days	2	Absence of complications	Infection

2 Data Collection and Methodology

2.1 Dataset

The dataset used for this study contains 794 patients with a total of 9 attributes (age, gender, tumor type, WHO, pathway (left or right), time between placement and use of the chamber (PU), difficulty of placement, and early and late complications). Table 1 presents an extract of the dataset. However, several studies in cancer patients have examined potential risk factors for TIVAPS-related complications, including patient age and sex [5]. During the study conducted at the Hassan II Oncology Center, WHO (World Health Organization) indicators were found to be another risk factor. So, the attributes are age, sex, and WHO indicators (from 0 to 4). The main class has two values, "false" and "true", which correspond to the absence or presence of a complications, respectively.

2.2 Methodology

In this article, four types of ML techniques have been employed to predict complications of implantable catheter chamber. The proposed method compares the classification performance of SVMs, DTs, RF, and LR. The proposed processes for building the predictive models are shown in Fig. 2.

- Exploring the dataset: The dataset is explored in the Python environment after the nominal attributes have been transformed into binary attributes.
- Data preprocessing: The data available in the real world is more often incomplete, with some missing entries. Therefore, data preprocessing is one of the data mining techniques to solve this problem. In this work, there is no missing data since the database is extracted from a retrospective study.
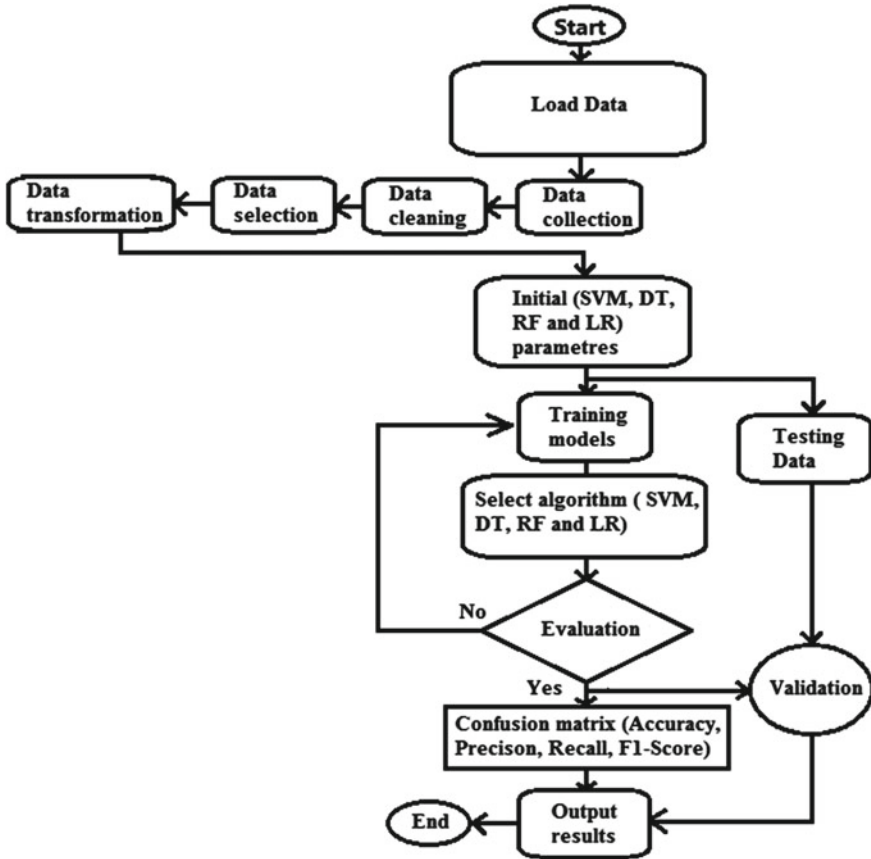
Fig. 2 Flowchart of developing the model to predict the Port A Cath complication

- Fitting and testing: After the selection of features, four classification algorithms, which include SVM, DT, RF, and LR, have been employed with the selected features. Performance comparison was done based on the confusion matrix, including the recall, precision, accuracy, and $F1$-score. The size was set to 80% of the data for classifier training to obtain the best RF and DT parameters. Thus, to find the performance of our classifiers.

2.3 Programming Language

There are many languages for implementing ML algorithms, such as WEKA, TANAGRA, MATLAB, ORANGE, Python, and others. Python is one of the most popular programming languages for data science and the open-source community for many reasons:
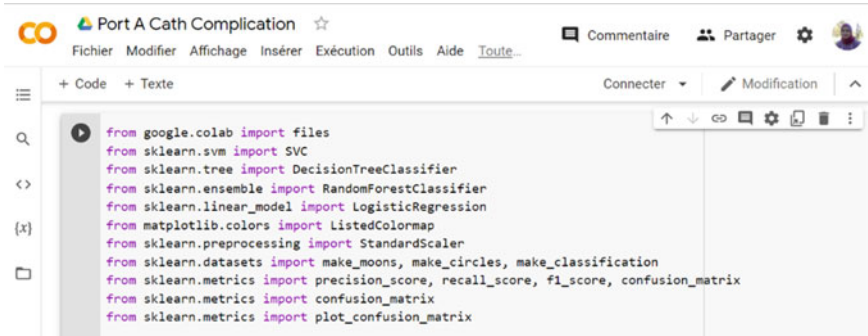
Fig. 3 Google Colab tool used for developing machine learning methods

It has a variety of libraries and frameworks such as NumPy, which is used for scientific computations, SciPy for more advanced computations, and scikit [6] for learning data mining and analysis.

Simplicity: Developing applications with Python is fast and simple compared to many other programming languages. In addition, it allows the developer to test algorithms without implementing them.

Massive online support: Python is an open-source programming language and has strong support from many resources and quality documentation around the world.

All these are reasons why Python is the most suitable for this application. The developed codes were implemented and executed online by using Google Colab [7]. Figure 3 shows Google Colab tool used for developing machine learning methods.

2.4 Machine Learning Algorithms

2.4.1 Support Vector Machine (SVM)

Support vector machines are a set of supervised learning methods used for classification, regression, and outlier detection. To classify, it finds a hyperplane in the feature space that differentiates the classes. The SVM model presents training data points as points in the feature space, mapped so that points in different classes are separated by as large a margin as possible. The testing data points are then mapped into this same space and classified according to which side of the margin they are on (Fig. 4).

The general form of Support Vector Classification estimating function is:

$$y_n = w^T x_n + b \quad \text{where } y_n \in \{1, -1\} \quad \text{and} \quad n \in \{1, \ldots, N\}$$

Here, x_n denotes a data point, w is a weight vector, b is the bias, and y_n is the model prediction. In this case, each prediction can be equal to 1 or -1 for points that lie on either side of the hyperplane and 0 for points that lie on the hyperplane itself. Given

Fig. 4 Support vector
machine
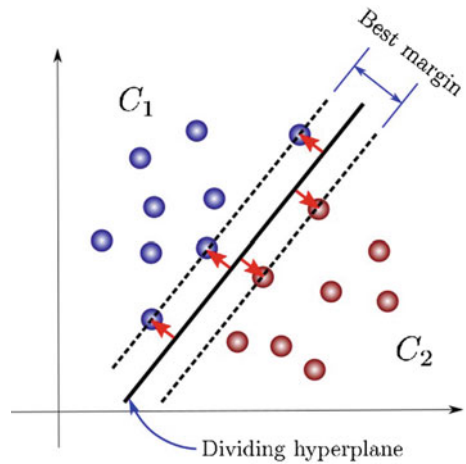
a dataset that contains several x's with their corresponding y's, what we want to do
is find values for w and b that give us the best possible margin.

2.4.2 Decision Tree (DT)

The decision tree is a supervised learning algorithm, resembling the tree analogy
in real life, used for classification and regression analysis [6]. It is a tree graph that
starts with one node and branches out based on its possible outcomes.

The dataset is then divided into two or more similar sets based on the most
significant predictors. The decision tree algorithm starts by calculating the entropy
of each attribute. Next, the dataset is divided using the variables or predictors with
maximum information gain or minimum entropy. These two operations are then
performed recursively with the rest of the attributes [8] (Fig. 5).

$$\text{Entropy}\,(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

$$\text{Gain}(S,\, A) = \text{Entropy}(S) - \sum_{\vartheta \,\epsilon\, \text{Values}(A)} \frac{|S_\vartheta|}{|S|} \text{Entropy}(S_\vartheta)$$

2.4.3 Random Forest (RF)

The random forest is another supervised machine learning algorithm. This technique
is useful for both regression and classification problems, but it is usually more accu-
rate for classification problems. Random forest classifier is a series of decision tree
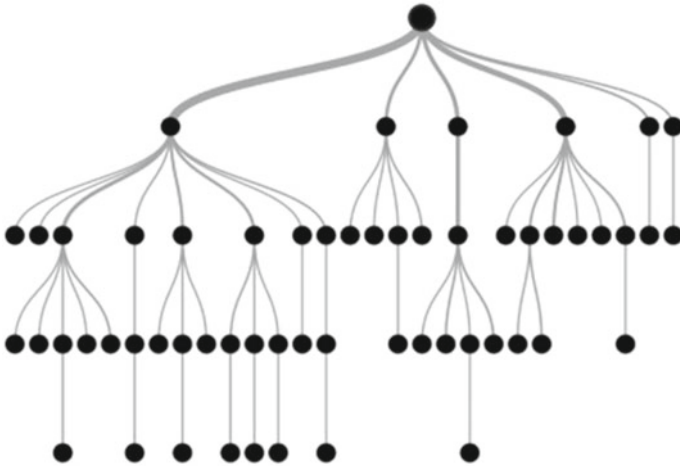
Fig. 5 Decision tree

classifiers in which each tree is built with a number of random vectors and is capable of voting for the most privileged class for prediction. The inclusion of randomness in the model protects it from being overfitted and gives a better score for the classification analysis (Fig. 6).
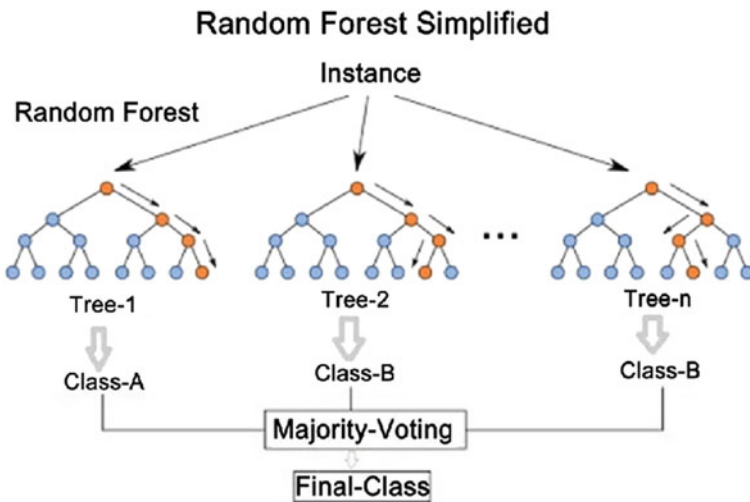
Fig. 6 Random forest

2.4.4 Logistic Regression (LR)

Logistic regression is an algorithm for classifying the probability of occurrence of an event, i.e., whether or not the event will occur. It is used to represent a binary or categorical result with only 2 classes. It is like linear regression, except that the result of the variable is a categoric variable instead of a continuous variable. It utilizes the logit link function, in which data values are adjusted, to predict. The mathematical interpretation will define the logit function as the natural logarithm of the probability that Y is equal to one of the categories [9]. If p is a probability, then the logit function for p is defined as follows:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$$

2.5 Performance Metrics

The confusion matrix can be generated for a classifier on a binary dataset, and it can be used to describe the performance of the classifier. It is performed to obtain the recall, precision, accuracy, and $F1$-score. These metrics are (case of two classes):

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{3}$$

$$F1\_\text{score} = 2\frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{4}$$

where TP is a number of true positive, TN is a number of true negative, FP is a number of fault positive, and FN is a number of fault negative.

To explain how a confusion matrix works, let us consider a confusion matrix with 2×2 dimension (two classes: true and false) as shown in Table 2.

Table 2 Confusion matrix: 2 × 2 dimensions

	Observed true	Observed false
Predicted true	TP	FP
Predicted false	FN	TN

3 Results and Discussion

Experiments and results from the different ML algorithms (SVM, DT, RF, and LR) are being trained to predict the Port A Cath complication. We have encoded the prediction attribute to classes 0 and 1 to represent the absence or presence of complication. The experiments are constructed under Google Colab environment.

The dataset is divided into two groups, one for training and one for testing. The proportions of training and testing data are 80% and 20% respectively. To evaluate the performance of each binary classifier, the accuracy and confusion matrix are calculated for each algorithm. Table 3 shows the calculated error metrics, i.e., accuracy, precision, recall, and $F1$-score.

For DT and RF, we experimented with depth = 3, 4, 5, and 10 and found that the best performance for both techniques is for depth = 3. The accuracy of DT and RF with different numbers of depth is shown in Table 3.

From Table 4, a good accuracy is obtained, since all values are between 88.05.5% and 91.82%. It is also important to mention that these results are obtained without optimization of hyperparameters except for DT and RF we optimized the number of depths.

If we verify the confusion matrix (CM) shown in Fig. 7a (e.g., an SVM classifier), out of 159 samples, 140 are correctly classified (140/159 = 88.05%, which is the accuracy). Only 19 are incorrectly classified. The classification error rate is 19/159 (11.95%) (Fig. 7).

Let us study the case of the confusion matrix DT and RF (see Fig. 7b, c), out of 159 samples, 143 are correctly classified, so the accuracy is 143/159 = 89.93%, and the misclassification rate is 16/159 = 10.07%. Regarding the case of LR (see Fig. 7d), out of 159 samples, 146 are correctly classified (146/159 = 91.82%, which is the accuracy). Only 13 are incorrectly classified. The classification error rate is 13/159

Table 3 Accuracy of DT and RF based on different values of depth

Depth	3	4	5	10
Decision tree (%)	89.93	87.42	87.42	87.43
Random forest (%)	89.93	88.05	86.79	87.42

Table 4 Error metrics: accuracy, precision, recall, and $F1$-score based on ML algorithms (SVM, DT, RF, and LR)

Classifiers	Accuracy (%)	Precision (%)	Recall (%)	$F1$-score (%)
Support vector machine	88.05	96	57.14	71.64
Decision tree	89.93	96.42	64.28	77.13
Random forest	89.93	96.42	64.28	77.13
Logistic regression	91.82	83.72	85.71	84.7
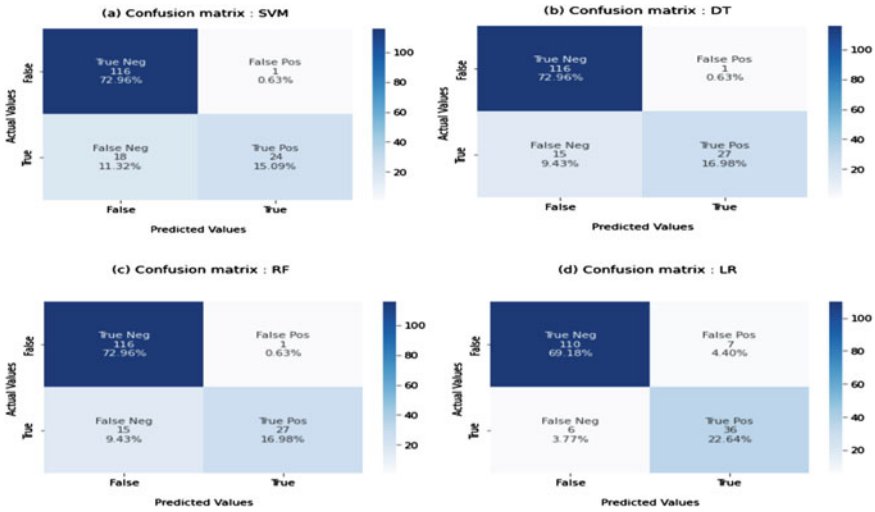
Fig. 7 Confusion matrices: **a** SVM, **b** DT, **c** RF, and **d** LR

(8.18%). Experimental results reveal that LR has achieved the highest percentage of accuracy.

4 Conclusion

In this paper, we present prediction models employing machine learning methods, namely support vector machine, decision tree, random forest, and logistic regression to predict the Port A Cath complication. From the experimental results, it can be seen that LR classifier gives the highest accuracy. In addition, LR has the highest recall and $F1$-score. Therefore, we can conclude that the LR classifier is suitable for predicting the existence of complications.

Future improvements to this project are the use of a pipeline structure for data preprocessing and optimization of hyperparameters of our classifiers, which could contribute to improve the results.

References

1. Niederhuber JE, Ensminger W, Gyves JW, Liepman M, Doan K, Cozzi E (1982) Totally implanted venous and arterial access system to replace external catheters in cancer treatment. Surgery 92:706–712
2. Vescia S, Baumgärtner AK, Jacobs VR et al (2008) Management of venous port systems in oncology: a review of current evidence. Ann Oncol 19:9–15
3. Yildizeli B, Lacin T, Batirel HF, Yüksel M (2004) Complications and management of long term central venous access catheters and ports. J Vasc Access 5:174–178
4. Hartkamp A, van Boxtel AJ, Zonnenberg BA, Witteveen PO (2000) Totally implantable venous access devices: evaluation of complications and a prospective comparative study of two different port systems. Neth J Med 57:215–223
5. Hsieh C-C, Weng H-H, Huang W-S et al (2009) Analysis of risk factors for central venous port failure in cancer patients. World J Gastroenterol 15:4709–4714
6. Scikit-Lean. Available: https://scikit-learn.org/
7. Google Colab. https://colab.research.google.com/
8. Mitchell TM (1997) Decision tree learning. Mach Learn 52–80
9. Sperandei S (2014) Lessons in biostatistics Understanding logistic regression analysis. Biochem Med 24(1):12–18

Fast and Accurate Color Image Classification Based on Quaternion Tchebichef Moments and Quaternion Convolutional Neural Network

Abdelmajid El Alami, Abderrahim Mesbah, Nadia Berrahou, Aissam Berrahou, Mohammed Ouazzani Jamil, and Hassan Qjidaa

Abstract This paper introduces a new architecture named QTMCNN for color image classification based on quaternion discrete Tchebichef moments (QTM) and quaternion convolutional neural network (QCNN) to improve the classification accuracy and to reduce the time of learning process. Color image is represented as a single quaternion matrix where each color pixel is represented as a pure quaternion. From this representation, quaternion Tchebichef moments are used to generate a matrix of low-dimensional significant features and fed to QCNN as input layer instead of color image. The proposed architecture reduces tremendously the number of parameters and consequently decreases the computational complexity while improving the classification rates. Experiments are conducted on Coil-100 and ETH-80 datasets to demonstrate the performance of the proposed architecture. The obtained results outperform other approaches in terms of classification accuracy and GPU elapsed time.

1 Introduction

Quaternion moments have been used in many applications in computer vision and color image classification. In [1], Wang et al. represented color image based on quaternion radial harmonic Fourier moments (QRHFMs) and derived moment invariants for image retrieval. Chen et al. [2] extended conventional Zernike moment used

A. El Alami (✉) · N. Berrahou · H. Qjidaa
Sidi Mohamed Ben Abdellah University, Fez, Morocco
e-mail: abdelmajid.elalami@usmba.ac.ma

A. Mesbah · A. Berrahou
Mohammed V University, Rabat, Morocco

M. O. Jamil
Private University of Fez, Fez, Morocco
e-mail: ouazzani@upf.ac.ma

in grayscale image to color image based on quaternion numbers. They constructed moment invariants to recognize color image under geometric transformations. The main drawback of continuous moments is the discretization error that increasing-by-increasing moment order. To avoid this inconvenience, various methods extracted moment invariants from discrete orthogonal moments. In [3], a set of moment invariants derived from quaternion discrete orthogonal Tchebichef moments introduced by Zhu et al. for color object recognition and image classification. However, invariant moments require a high computation time and are limited in the classification process. To resolve this problem, different methods based on deep learning and discrete moments without deriving invariant moments are realized for image classification and pattern recognition [4–9]. El Alami et al. [10] used quaternion discrete moments and neural networks for color face recognition.

Recently, many methods based on quaternion convolutional neural networks (QCNN) have been developed. Zhu et al. [11] developed a QCNN model to extract pertinent features from images for denoising and classification tasks. Yin et al. [12] designed a QCNN model for color image classification and color image forensics. They demonstrated the performance of QCNN against CNN with same structure. El Alami et al. [13] developed a model based on quaternion representation of color image as input vector in quaternion neural network for face recognition.

In this paper, we propose a new architecture based on QCNN and quaternion discrete Tchebichef moments named QTMCNN. In fact, the QTM are integrated as input layer to QCNN. QTM have the ability to extract pertinent information from color image in first orders. The quaternion moments are calculated from the quaternion representation of color image. This makes it possible to reduce enormously the parameters in QCNN and consequently to accelerate learning process.

The rest of this paper is organized as follows: Sect. 2 presents an overview on quaternion discrete Tchebichef moments. The proposed architecture is explained in Sect. 3. The simulation results based on our QTMCNN are provided in Sect. 4. We conclude the present work in Sect. 5.

2   Quaternion Discrete Tchebichef Moments

2.1   Tchebichef Polynomials

The Tchebichef polynomials [14] are defined by using hypergeometric function as:

$$t_n(x, N) =_3 F_2 \left(\begin{matrix} -n, -x, 1+n \\ 1, 1-N \end{matrix} \middle| 1 \right), \tag{1}$$

Tchebichef polynomials satisfy the orthogonal property:

$$\sum_{x=0}^{N-1} t_m\,(x;\,N)\,t_n\,(x;\,N)\,\omega_t\,(x) = \rho_t\,(n,\,N)\,\delta_{mn},\tag{2}$$

with $\omega_t\,(x) = 1$ and $\rho_t = (2n)!\begin{pmatrix} N+n \\ 2n+1 \end{pmatrix}$.

The normalized Tchebichef polynomials are given by:

$$\tilde{t}_n\,(x;\,N) = t_n\,(x;\,N)\sqrt{\frac{\omega_t\,(x)}{\rho_t\,(n,\,N)}}.\tag{3}$$

The discrete polynomials of Tchebichef satisfy the following three-term recurrence relation:

$$\tilde{t}_n\,(x;\,N) = \frac{(2n-1)\,\tilde{t}_1\,(x;\,N)\,\tilde{t}_{n-1}\,(x;\,N) - (n-1)\left(1 - \frac{(n-1)^2}{N^2}\right)\tilde{t}_{n-2}\,(x;\,N)}{n},\tag{4}$$

with $\tilde{t}_0\,(x;\,N) = 1$ and $\tilde{t}_1\,(x;\,N) = \frac{(2x+1-N)}{N}$.

2.2 2D Quaternion Discrete Tchebichef Moments

Quaternion numbers are defined by Hamilton [15] as a generalization of complex numbers. A quaternion number q is given by:

$$q = q_0 + q_1 i + q_2 j + q_3 k,\tag{5}$$

where $q_0, q_1, q_2, q_3 \in \mathbb{R}$ while i, j, and k are imaginary units that satisfy the following rules:

$$i^2 = j^2 = k^2 = ijk = -1, ij = -ji = k, jk = -kj = i, ki = -ik = j.\tag{6}$$

The conjugate and norm of q are:

$$q^* = q_0 - q_1 i - q_2 j - q_3 k, \, |q| = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}.\tag{7}$$

Tchebichef's quaternion discrete orthogonal moments (QTM) of a $(m+n)$th order of a two-dimensional color image with intensity function $f(x, y)$ are defined over the domain $[0, M-1] \times [0, N-1]$ as:

$$QTM_{mn} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (f_R(x, y)i + f_G(x, y)j + f_B(x, y)k)\tilde{t}_m(x)\tilde{t}_n(y)\mu, \quad (8)$$

where $f_R(x, y)$, $f_G(x, y)$, and $f_B(x, y)$ correspond to three channels of color image $f(x, y)$ and $\mu = \frac{1}{\sqrt{3}}(i + j + k)$ is the unit pure quaternion that corresponds to the gray line in RGB color space.

Based on the orthogonality property of Tchebichef polynomials, color image $f(x, y)$ is reconstructed perfectly, when all moments are taken, by employing the following inverse transform:

$$f(x, y) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} QTM_{mn}\tilde{t}_m(x)\tilde{t}_n(y)\mu, \quad (9)$$

Table 1 shows some reconstructions of original color image with size of 128×128 selected from Coil-100 dataset [17] using quaternion discrete Tchebichef moments up to orders (11, 11), (23, 23), (63, 63), and (127, 127), respectively. We can clearly see a greater resemblance between the original and reconstructed images in lower orders. Indeed, the normalized image reconstruction error (NIRE) decreases with increasing moment orders. This demonstrates the ability of quaternion Tchebichef moments to extract sufficient features from image that are useful for classification process.
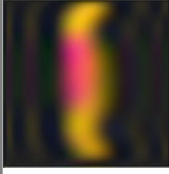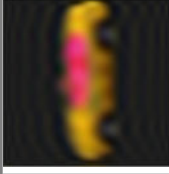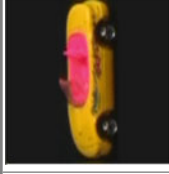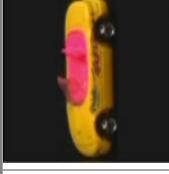
3 2D Quaternion Techebichef Moments Convolutional Neural Network

The proposed architecture QTMCNN for color image classification contains quaternion Tchebichef moments input layer, three blocks of quaternion convolution. We downsample the features in maxpooling layer according to the split method from the associated real-valued layers. Then, the extracted features are provided to quaternion dense layers. The last softmax layer produces the number of labels which corresponds to the used classes. To optimize our architecture, we adopted categorical cross-entropy loss function, normalized Adam optimizer with the learning rate of 0.0001 and batch size of 80. The structure of the proposed QTMCNN is given in Fig. 1. The activation functions in quaternion domain are used as a split activation function $\overline{\psi}$ [16] defined by:

$$\overline{\psi}(q) = \psi(q_0) + \psi(q_1)i + \psi(q_2)j + \psi(q_3)k, \quad (10)$$

with $\psi(\cdot)$ as a standard activation function.

Table 1 Reconstruction results of color image using quaternion Tchebichef moments

Original image (128 × 128)	Reconstructed color images up to orders			
	(11, 11)	(23, 23)	(63, 63)	(127, 127)
	NIRE = 0.069684	NIRE = 0.036742	NIRE = 0.005791	NIRE = $3.2e^{-19}$
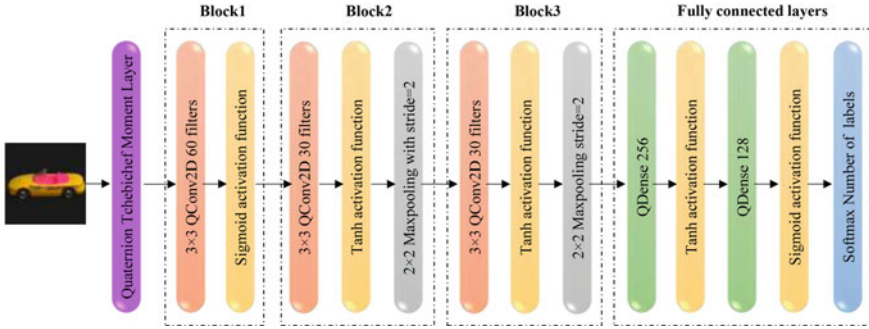
Fig. 1 Proposed architecture (QTMCNN)

4 Experiments

In this section, based on the proposed architecture illustrated in Fig. 1, the simulation experiments are performed by using two splits on Coil-100 and ETH-80 datasets. For the first split noted S1, we randomly select 20% of each dataset for training and keep the rest for testing; for the second split noted S2, we randomly select 50% for training and 50% for testing. All experiments are realized in Google Collaboratory framework with Tesla K80 GPU and 12.72 GB of RAM.

4.1 Results on Coil-100 and ETH-80 Datasets

Coil-100 Dataset. Columbia University Image Library (COIL-100) [17] dataset consists of 7200 RGB images with size of 128×128 pixels divided into 100 objects. Each object is captured under 72 different viewing directions and reflectance characteristics.

Table 2 gives the classification accuracy results. We obtained of 98.83% and 99.72% at order 16 in S1 and S2, respectively. The comparison to other methods is shown in Table 3; we can see the superiority of our QTMCNN against other approaches.

ETH-80 Dataset. The ETH-80 dataset [18] consists of 3280 color images divided into eight classes; each class contains ten object sets. Each object set is composed of 41 images taken under different view angles with the size of 256×256 pixels. The classification accuracy results are displayed in Table 4.

The best results are achieved at order 24 by 90.45% and at order 20 by 96.01% in S1 and S2, respectively. Table 5 shows the comparison to other methods; it can be observed that our method outperforms other algorithms except DRMWV [22] which exceed ours by 2.24%.

Table 2 Classification accuracy (%) on Coil-100 dataset

Order	4	8	12	16	20	24	28	32
S1	98.33	98.07	98.38	**98.83**	98.41	98.02	98.29	97.67
S2	99.28	99.56	99.64	**99.72**	99.69	99.64	99.56	99.64

Table 3 Comparison accuracy (%) on Coil-100 dataset

Methods	SalBayes [19]	SIFT [19]	HMAX [19]	ALP-TMR [20]	Our
S1	97.20	87.19	77.02	96.92	98.83

Table 4 Classification accuracy (%) on ETH-80 dataset

Order	4	8	12	16	20	24	28	32
S1	89.05	88.56	89.24	89.85	90.08	**90.45**	90.04	89.43
S2	94.29	94.76	95.00	95.30	**96.01**	95.36	95.36	95.06

Table 5 Comparison accuracy (%) on ETH-80 dataset

Methods	ALP-TMR [20]	IMTF [21]	LRRC [23]	DRMWV [22]	Our
S1	88.92	89.84	–	–	90.45
S2	–	–	94.75	98.25	96.01

4.2 Ablation and Complexity

In this subsection, we compare the computational complexity between our QTM-CNN architecture and the same QCNN structure using quaternion representation (QR) of image as an input layer instead of quaternion Tchebichef moments (QTM). Three simulations are performed on split S1 from Coil-100 dataset. We obtained 96.71% classification rate with QCNN, while the best classification result 98.83% was achieved at order (16, 16) with QTMCNN. In addition, the number of parameters is significantly decreased by using QTM compared to QR. The storage memory usage is optimized by 98.38% approximately. Table 6 shows the GPU elapsed time after 200 epochs by both methods. We observe that our architecture is 35 times faster than QCNN model.

Table 6 GPU elapsed time in second on Coil-100 dataset

Methods	First simulation	Second simulation	Third simulation	Average
QCNN	3784.97	3825.08	3816.04	3808.69
QTMCNN	108.25	108.04	107.65	107.98

5 Conclusion

This paper introduced a new architecture QTMCNN for color image classification based on quaternion Tchebichef moments and quaternion convolutional neural network. The simulation results on Coil-100 and ETH-80 datasets demonstrate the performance of the proposed architecture and give high recognition accuracy against other existing methods. Furthermore, quaternion discrete Tchebichef moments have the ability to represent perfect color image and to accelerate learning process by decreasing the parameters. As future work, we will use quaternion moments with QCNN to investigate and explore large datasets.

References

1. Xiang-yang W, Wei-yi L, Hong-ying Y, Pan-pan N, Yong-wei L (2015) Invariant quaternion radial harmonic Fourier moments for color image retrieval. Opt Laser Technol 66:78–88. https://doi.org/10.1016/j.optlastec.2014.07.020
2. Chen BJ, Shu HZ, Zhang H, Chen G, Toumoulin C, Dillenseger JL, Luo LM (2012) Quaternion Zernike moments and their invariants for color image analysis and object recognition. Signal Process 92(2):308–318. https://doi.org/10.1016/j.sigpro.2011.07.018
3. Zhu H, Li Q, Liu Q (2014) Quaternion discrete Tchebichef moments and their applications. Int J Signal Process Image Process Pattern Recognit 7(6):149–162. https://doi.org/10.14257/ijsip.2014.7.6.13
4. Lakhili Z, El Alami A, Mesbah A, Berrahou A, Qjidaa H (2019) Deformable 3D shape classification using 3D Racah moments and deep neural networks. Procedia Comput Sci 148:12–20. https://doi.org/10.1016/j.procs.2019.01.002
5. Mesbah A, Berrahou A, El Alami A, Berrahou N, Berbia H, Qjidaa H (2019) 3D object classification using 3D racah moments convolutional neural networks. In: Proceedings of the 2nd international conference on networking, information systems & security, pp 1–6. https://doi.org/10.1145/3320326.3320397
6. Lakhili Z, El Alami A, Mesbah A, Berrahou A, Qjidaa H (2019) 3D shape classification using 3D discrete moments and deep neural networks. In: Proceedings of the 2nd international conference on networking, information systems & security, pp 1–6. https://doi.org/10.1145/3320326.3320398
7. Mesbah A, Berrahou A, Hammouchi H, Berbia H, Qjidaa H, Daoudi M (2019) Lip reading with Hahn convolutional neural networks. Image Vis Comput 88:76–83. https://doi.org/10.1016/j.imavis.2019.04.010
8. Lakhili Z, El Alami A, Mesbah A, Berrahou A, Qjidaa H (2020) Robust classification of 3D objects using discrete orthogonal moments and deep neural networks. Multimed Tools Appl 79(27):18883–18907. https://doi.org/10.1007/s11042-020-08654-7

9. Lakhili Z, El Alami A, Qjidaa H: Enhancing the performance of grayscale image classification by 2D Charlier moments neural networks. In: International conference on electronic engineering and renewable energy. Springer, Singapore, pp 151–159. https://doi.org/10.1007/978-981-15-6259-4_14

10. El Alami A, Berrahou N, Lakhili Z, Mesbah A, Berrahou A, Qjidaa H (2022) Efficient color face recognition based on quaternion discrete orthogonal moments neural networks. Multimed Tools Appl 81(6):7685–7710. https://doi.org/10.1007/s11042-021-11669-3

11. Zhu X, Xu Y, Xu H, Chen C (2018) Quaternion convolutional neural networks. In: Proceedings of the European conference on computer vision (ECCV), pp 631–647. https://arxiv.org/abs/1903.00658

12. Yin Q, Wang J, Luo X, Zhai J, Jha SK, Shi YQ (2019) Quaternion convolutional neural network for color image classification and forensics. IEEE Access 7:20293–20301. https://doi.org/10.1109/ACCESS.2019.2897000

13. El Alami A, Lakhili Z, Mesbah A, Berrahou A, Qjidaa H (2019) Color face recognition by using quaternion and deep neural networks. In: 2019 international conference on wireless technologies, embedded and intelligent systems (WITS). IEEE, pp 1–5. https://doi.org/10.1109/WITS.2019.8723788

14. Mukundan R, Ong SH, Lee PA (2001) Image analysis by Tchebichef moments. IEEE Trans image Process 10(9):1357–1364. https://doi.org/10.1109/83.941859

15. Hamilton WR (1866) Elements of quaternions. Green & Company, Longmans

16. Xu D, Zhang L, Zhang H (2017) Learning algorithms in quaternion neural networks using ghr calculus. Neural Netw World 27(3):271–282. https://doi.org/10.14311/nnw.2017.27.014

17. Nene S, Nayar S, Murase H (1996) Columbia object image library (COIL-100). Technical report, Columbia University

18. Leibe B, Schiele B: Analyzing appearance and contour based methods for object categorization. In: 2003 IEEE Computer Society conference on computer vision and pattern recognition, 2003. Proceedings, 2003, pp II-409. https://doi.org/10.1109/CVPR.2003.1211497

19. Elazary L, Itti L (2010) A Bayesian model for efficient visual search and recognition. Vis Res 50(14):1338–1352

20. Zhang H, Zhang Z, Zhao M, Ye Q, Zhang M, Wang M (2020) Robust triple-matrix-recovery-based auto-weighted label propagation for classification. IEEE Trans Neural Netw Learn Syst 31(11):4538–4552

21. Yang C, Yu Q (2021) Invariant multiscale triangle feature for shape recognition. Appl Math Comput 403:126096. https://doi.org/10.1016/j.amc.2021.126096

22. Hayat M, Bennamoun M, An S (2014) Deep reconstruction models for image set classification. IEEE Trans Pattern Anal Mach Intell 37(4):713–727. https://doi.org/10.1109/TPAMI.2014.2353635

23. Shah SA, Nadeem U, Bennamoun M, Sohel F, Togneri R (2017) Efficient image set classification using linear regression based image reconstruction. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 99–108

Slew-Rate Enhancement of a Full-On Chip CMOS LDO Based on a Capacitorless Push–Pull Current Booster Circuit

Kamal Zared, Hatim Ameziane, Aicha Alami Hassani, Hicham Akhamal, Mohammed Jamil Ouazzani, and Hassan Qjidaa

Abstract In this paper, a full-on chip low drop-Out voltage regulator (LDO) with a simple Slew-Rate Enhancement Circuit (SREC) has been proposed and simulated in TSMC 0.18 μm CMOS process. The proposed SREC technique has been designed mainly to improve the transient response at full load by using only active components to minimize the space area of the chip as low as possible. The proposed LDO presents a quiescent current (I_Q) of 14 μA at input voltage (V_{int}) of 1.8 V, providing a 60 mA load current (I_{load}) with 1.6 V output voltage (V_{out}) and drop-out voltage (V_{DO}) of 200 mV. The simulation results show the system stability under different load conditions. The load regulation has been improved with a nominal output voltage recovery time less than 900 ns and the output over/undershoot values are 156 mV and 190 mV, respectively.

Keywords LDO · Full-on-chip · Capacitorless · SREC · SoC · CMOS

K. Zared (✉) · A. Alami Hassani
Faculty of Science and Technology, Sidi Mohamed Ben Abdellah University, Fez, Morocco
e-mail: kamal.zared@usmba.ac.ma

A. Alami Hassani
e-mail: aicha.alamihassani@usmba.ac.ma

H. Ameziane
National School of Applied Sciences, Khouribga, Morocco
e-mail: h.ameziane@usms.ma

Sultan Moulay Slimane University, Beni-Mellal, Morocco

H. Akhamal · H. Qjidaa
Faculty of Sciences and Technology, Sidi Mohamed Ben Abdellah University, Fez, Morocco
e-mail: hichamfsdm@hotmail.com

M. Jamil Ouazzani
Private University of Fez, Fez, Morocco
e-mail: ouazzani@upf.ac.ma

1   Introduction

High performance of the latest generation of LDO voltage regulators such as sensitivity to noise, low consumption, and stability of the provided power supply makes them more used in power manager integrated circuit (PMIC) block for battery powered embedded systems like medical devices, biomedical, domotics, electronic wearables, Internet of Thing (IoT), etc. The LDO's function in the PMIC is to provide a stable and regulated power supply independently of the system line and load conditions.

The major challenges of LDO conventional structure (LDOCS) are stability, transient response rapidity and low consumption without a large off-chip capacitor (OCC). The LDOCS consists of an error amplifier (EA), feedback, and output Power Transistor (PT). Many methods have been used to increase LDOCS performance and decrease its OCC. Moreover, the integration of some pF capacitor on the System on Chip (SoC) of PMIC presents more difficulties. In literature, many approaches are focused on reducing the OCC using Miller compensation and slew-rate enhancement techniques. In Ref. [1], a buffer stage is inserted at the gate of the PT to realize low quiescent current LDO for low consumption is presented. In [2], a Q-reduction circuit is implemented to minimize both the OCC for high integration density and the quiescent current for low power consumption. To eliminate the undesirable effect of removing or reducing the OCC a technique that generates a zero internally for stability is presented and discussed in [3]. In [4, 5], a current amplifier is used as a fast transient path between LDO output and EA output to improve the system transient response and stability with a small value of OCC. In Ref. [6], the Miller compensation technique is used to guarantee the stability using active capacitor and optimized frequency shaping. In [7], an active capacitor in CMOS technology is used with a Miller Compensation technique to stabilize the system and reduce the LDO chip area.

In [8], a push–pull stage is used to enable output driving capability and enhance the slew-rate at the PT gate by Common Mode FeedBack resistors. An evolved version of Flipped Voltage Follower has been proposed in [9, 10] to ensure full-on chip LDO with enhanced slew rate. In [11], a capacitorless LDO using multipath nested Miller compensation with embedded feedforward path for fully integrated system is studied. An active compensation network based on a bootstrap flipped-voltage follower is presented and discussed in [12, 13]. However, most of the previous studies do not consider the difficulty of integrating a few picofarad capacitors on the chip.

In this work, a simple capacitorless circuit SREC consisting only of CMOS transistors to detect changes at the LDO output instead a large capacitor which used in all references cited above to sense the output voltage variation. Firstly, a large capacitor value implies a large chip area. Secondly, the time for charging and discharging capacitor and detect variation is not neglected is implemented to enhance the slew-rate of a full-on chip LDO. The proposed SREC reduces the chip area of the full-on

chip LDO and decreases its quiescent current at normal operation. Due to this performance, the proposed full-on chip LDO is suitable for any SoC power management and embedded systems.

The remainder of this paper is organized as follows. Section 2 presents the bloc schematic of proposed LDO architecture. In Sect. 3, the proposed circuit SREC of slew-rate enhancement will be presented and discussed. The stability analysis of the proposed LDO is given in Sect. 4. Simulation results of the proposed LDO with SREC circuit implementation are presented in Sect. 5. Section 6 summarizes the results of this work and draws conclusions.

2 Implementation of the Proposed LDO

The proposed LDO structure is based on the LDOCS with three-stage EA which is used to increase the regulation accuracy. The output capacitor C_{pp} (generally up to 100 pF) is used to module the line and load distributions.

The proposed technique is implemented in three-stage gain LDO as shown in Fig. 1. The parameters g_{m1}, g_{m2}, g_{m3} and g_{mp} are the transconductances of the EA, second gain stage, source follower and PT transistor, respectively. The source follower is used to split the no dominant pole at the gate of PT. R_{O1}, R_{O2}, $R_{O3,}$ and $R_L//r_p$ are the output resistances of the EA, second stage, source follower and PT resistance in parallel with the load resistance, respectively. C_{O1}, C_{O2}, C_{O3}, and C_{pp} are the output equivalent capacitances of the EA, second stage, source follower, and load, respectively. C_{m1} and C_{G-PT} represent the internal compensation capacitor and the gate power transistor, respectively , $G_{m-\text{push}}$, $G_{m-\text{pull}}$ are the transconductances of the push and pull current amplifier of the SREC circuit. Finally, R_{push} and R_{pull} present the equivalent resistances of the output push and pull current amplifiers, respectively.

3 The Proposed SREC

A capacitorless slew-rate enhancement circuit, as shown in Fig. 2, is implemented to convert the output voltage variation to the necessary current pushed or pulled (not both) to (or from) the PT gate capacitor during load transient. In the proposed SREC, no coupling capacitor is used to detect the load variation with no complexity in bias voltage of slew-rate enhancement circuit as in many works in the art state. As shown in Fig. 2, the proposed SREC circuit consists of three common source amplifiers [$(M_{22} - M_{23})$, $(M_{24} - M_{25})$, $(M_{28} - M_{29})$], a common gate amplifier (M_{26} – M_{27}) and a push–pull transistors (MPUSH, MPULL). $K1$ and $K2$ are the gains of the current mirrors in SREC circuit.
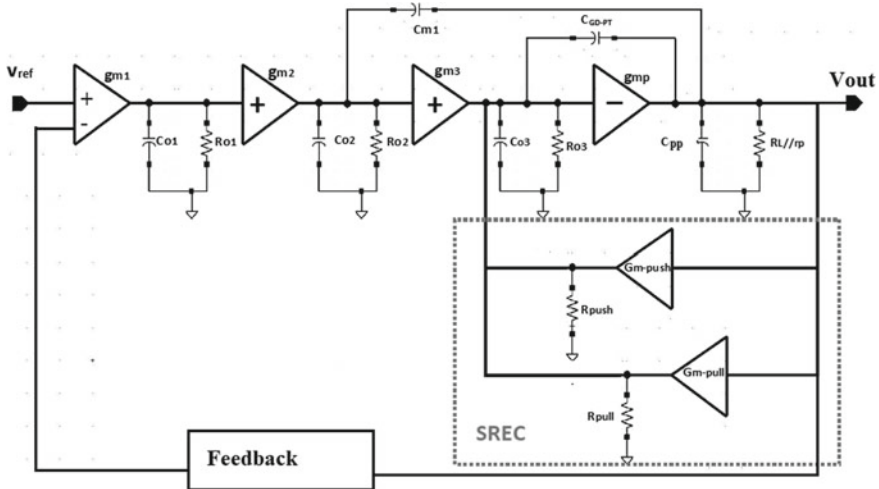
Fig. 1 Topology of the LDO with the proposed SREC
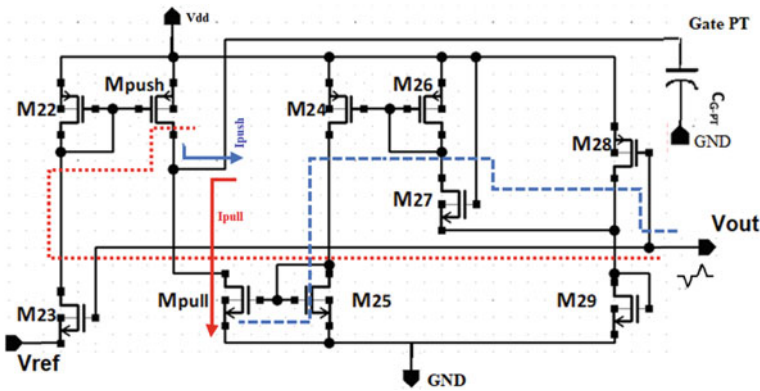
Fig. 2 Schematic of the proposed SREC

The operating of the proposed SREC is presented as follows: the variation voltage at the LDO output will be converted to current, the latest provides the requirement current through the PUSH-PATH (PULL-PATH) to charge (discharge) PT gate capacitor $C_{G\text{-}PT}$ as given in Eq. (1).

At steady state, the SREC bloc isolates the system output node from the PT gate because the MOS transistors M_{23} and M_{28} are controlled by output voltage variation, and the SREC block does not affect the function of the principal loop regulation.

The common source amplifiers $M_{23} - M_{24}$ and $M_{28} - M_{29}$ are used to detect and amplify the LDO output voltage variation when the load current changes from low to high values and vice versa. The charging and discharging current of the PT gate

capacitor can be written as:

$$i_{G-PT} = f(V_{out}) = \begin{cases} i_{push}, & \Delta V_{out} > 0 \\ -i_{pull}, & \Delta V_{out} < 0 \end{cases} \tag{1}$$

With

$$i_{pull} = \frac{g_{m-pull}g_{m24}R_{o23}g_{m27}g_{m28}R_{o26}R_{o28}R_{o24}}{(1 + SR_{o26}C_{G24})(1 + SR_{o28}C_{D28})(1 + SR_{o24}C_{G-pull})}V_{out} \tag{2}$$

$$i_{push} = \frac{g_{m-push}g_{m23}R_{o23}}{1 + SR_{o23}C_{D23}}V_{out} \tag{3}$$

where i_{G-PT}, i_{push}, and i_{pull} are the charging/discharging current of PT gate capacitor, the output current of the PUSH-PATH and PULL-PATH, respectively. g_{m-push} and g_{m-pull} are the transconductances of M_{pull} and M_{push} transistors, respectively. $g_{mi}(i = 23, 24, 26, 27)$ and $R_{Oj}(j = 23, 24, 26, 27, 28)$ are the transconductances and the equivalent resistances of transistors M_i and M_j, respectively. C_{D23} and C_{D28} are the parasitic capacitances at transistor's drains M_{23} and M_{28}. C_{G24} and C_{G-pull} are the parasitic capacitances at the transistor's gates M_{24}, M_{pull}.

4 Frequency Domain Stability Analysis

To analyze the behaviour of the proposed LDO in frequency domain, two assumptions are taken into consideration:

Assumption 1: $C_{pp} \gg C_{G-PT} \gg C_{m1} \gg C_{o1} > C_{o2} > C_{o3}$.

Assumption 2: $R_L \gg R_{o1} \gg R_{o2} \gg R_{o3} \gg r_p$.

Based on the above assumptions, the open loop transfer function of the proposed LDO expressed by:

$$H_{(s)} \approx G_{DC}\frac{(1 + a_1s)(1 + a_2s + a_3s^2 + a_4s^3)}{(1 + b_1s)(1 + b_2s + b_3s^2 + b_4s^3)} \tag{4}$$

where G_{DC} is the low frequency gain and is given by:

$$G_{DC} = -g_{m1}g_{m2}g_{m3}g_{mp}R_{o1}R_{o2}R_Gr_p \tag{5}$$

The nominator and denominator parameters of Eq. (4) are:

$$a_1 = R_{o1}C_m, \quad a_2 = \frac{C_m}{g_{m2}g_{m3}g_{mp}R_{O2}R_G},$$
$$a_3 = \frac{C_mC_G}{g_{m2}g_{m3}g_{mp}R_{O2}}, \quad a_4 = \frac{C_mC_GC_{02}}{g_{m2}g_{m3}g_{mp}} \tag{6}$$

$$b_1 = g_{m2}g_{m3}g_{mp}R_{o1}R_{o2}R_G r_p C_m,$$

$$b_2 = \frac{1}{g_{m2}g_{m3}g_{mp}r_p}\left[C_{o2}\left(g_{mpp}g_{mp} + \frac{1}{R_{01}R_G} + \frac{1}{R_G r_p}\right)\right.$$
$$\left. + C_G\left(\frac{1}{R_{o2}r_p} + \frac{C_{o2}}{R_{O1}r_p C_m} + \frac{1}{R_G r_p} + \frac{C_m}{C_G}\frac{1}{R_{o1}r_p}\right)\right],$$

$$b_3 = \frac{C_{o2}}{g_{m2}g_{m3}g_{mp}r_p}\left(\frac{R_{o1}}{R_{o2}R_G} + \frac{C_G}{R_{o1}C_m}\right) + \frac{C_G}{g_{mp}g_{m3}g_{m2}r_p R_{o2}}, \quad b_4 = \frac{C_m C_G C_{02}}{g_{m2}g_{m3}g_{mp}}$$

$$(7)$$

From Eqs. (4), (6) and (7), the poles and zeros can be expressed as:

$$w_{pd} = \frac{1}{g_{m20}g_{mp}R_o R_2 R_3 C_{m1}}, \quad w_{pnd1} \approx \frac{1}{R_{o1}C_{o1}}, \quad w_{pnd2} \approx \frac{-a_2}{a_3} = \frac{g_{m20} + g_{mp}}{C_{m2}}$$

$$w_{pnd3} \approx \frac{-a_1}{a_2} = \frac{g_{m20}}{C_{m2}}, \quad \text{and} \quad w_{zRHP} = \frac{g_{mp}}{C_{m2}}, \quad w_{zLHP} = \frac{g_{m20}}{C_{m1}}$$

$$(8)$$

The zeroes and poles locations are given in Eq. (8). The stability analysis will be discussed in simulation results.

5 Simulation Results

As a part of the validation step of the proposed LDO performance, several simulations have been performed using 0.18 μm CMOS process. Figure 3 shows the simulation result of the open loop frequency response, the system is stable in different load conditions with a phase margin of 76°. The simulation result of the load transient response is shown in Fig. 4. When the load current changes from 100 μA to 60 mA with 500 ns rise and fall times and with the SREC technique, the maximum values of under/overshoot are 156 mV and 148 mV with a recovery times of 270 ns and 810 ns respectively. Without SREC circuit, the under/overshoot are 440 mV and 190 mV respectively with a recovery-times of 280 ns and 1.24 us, respectively. It is obvious that the load regulation characteristic is significantly improved.

To evaluate the transient performance, the transient figure-of-merit (FOM) [9] is used and it is given in Table 1.

$$F.OM = T_R \frac{I_Q}{I_{max}} \tag{9}$$
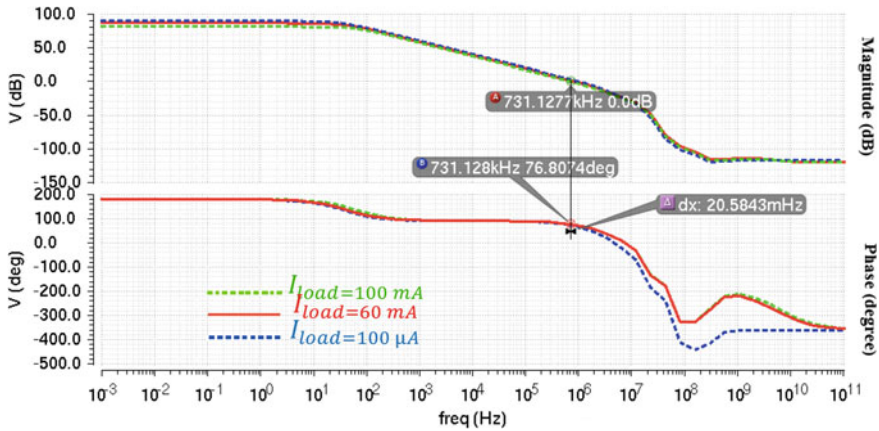
Fig. 3 Open loop frequency response plots of proposed LDO
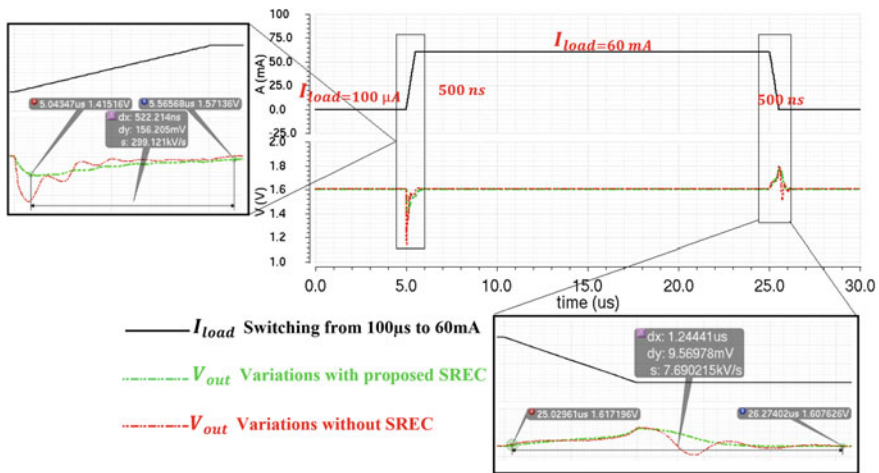
Fig. 4 Load transient response for I_{Load} up from 100 μA to 60 mA

6 Conclusions

A full-on chip regulator LDO that uses a capacitorless SREC to enhance performance has been presented and simulated in 0.18 um CMOS technology. The simulation results of the proposed LDO with capacitorless SREC technique prove the reduction of over/undershoot amplitudes and with active area less than in comparing with other work using coupling capacitor to detect output variation and then act. The proposed LDO is suitable to be integrated on power management of embedded system.

Table 1 The performances and comparison results with other works

	[5]	[12]. Sim	[13] sim and exp	This work. Sim
Technology (μm)	0.18	0.18	0.18	0.18
I_{load} (mA)	0–100	0–100	0–50	0.1–50
V_{in} (V)	1.8–3.8	>1.1	1.8	1.2–1.8
V_{out} (V)	2.25	1	1.6	1.6
V_{DO} (mV)	200	100	200	200
C_{out} (pF)	100	NA	100	Free
I_Q (mA)	0.046	NA	0.014	0.014
Slew-rate enhancement capacitor (pF)	2.5	12	0.3	0
Settling time (μs)	0.067	1.3	2.4	0.522
FOM (ns)	0.030	NA	0.672	0.146

References

1. Rincon-Mora GA, Allen PE (1998) A low-voltage, low quiescent current, low drop-out regulator. IEEE J Solid-State Circuits 33(1):36–44
2. Lau SK, Mok PKT, Leung KN (2007) A low-dropout regulator for SoC with *Q*-reduction. IEEE J Solid-State Circuits 42(3):658–664
3. Saberkari A, Alarcón E, Shokouhi SB (2013) Fast transient current-steering CMOS LDO regulator based on current feedback amplifier. Integr VLSI J 46(2):165–171
4. Milliken RJ, Silva-Martínez J, Sánchez-Sinencio E (2007) Full on-chip CMOS low-dropout voltage regulator. IEEE Trans Circuits Syst Regul Pap 54(9):1879
5. Wen K, Liu S, Li Y (2020) A voltage-adjustable output-capacitorless LDO regulator with split-length current mirror compensation and overshoot/undershoot reduction. Analog Integr Circuits Sig Process 105(3):459–470
6. Huang W-J, Liu S-I (2008) Capacitor-free low dropout regulators using nested Miller compensation with active resistor and 1-bit programmable capacitor array. IET Circuits Dev Syst 2(3):306
7. Ameziane H, Zared K, Akhamal H, Qjidaa H (2019) Full on-chip low dropout voltage regulator with an enhanced transient response for low power systems. Int J Electr Comput Eng 9(6):4637–4648
8. Lim C-C, Lai N-S, Tan G-H, Ramiah H (2015) A low-power fast transient output capacitor-free adaptively biased LDO based on slew rate enhancement for SoC applications. Microelectronics J 46:740–749
9. Zeng Y, Li Y, Zhang X, Zhou Tan H (2017) A push-pulled FVF based output-capacitorless LDO with adaptive power transistors. Microelectron J 64:69–77
10. Li C, Chan PK (2017) FVF LDO regulator with dual dynamic-load composite gain stage. Analog Integr Circuits Sig Process 92(1):131–140
11. Bu S, Guo J, Leung KN (2018) A 200-ps-response-time output-capacitorless low-dropout regulator with unity-gain bandwidth >100 MHz in 130-nm CMOS. IEEE Trans Power Electron 33(4):3232–3246
12. Abdi F, Bastan Y, Amiri P (2019) Dynamic current-boosting based FVF for output-capacitor-less LDO regulator. Analog Integr Circuits Sig Process 101(3):555–564
13. Zamora-Mejia G et al (2020) A 0.18 μm CMOS capacitor-less low-drop out voltage regulator compensated via the bootstrap flipped-voltage follower. Microelectron J 101:104809

Telemedicine in the Era of Covid-19: Teleconsultation Architecture Platform

Ayat Yassine, El Moussati Ali, and Mir Ismail

Abstract Teleconsultation is a type of medical practice similar to face-to-face consultations, and it allows a health professional to give a consultation remotely through information and communication technologies. In the context of the management of the coronavirus epidemic, the use of teleconsultation practices can facilitate healthcare access and limit the risk of avoidable propagation in medical cabinets. This paper presents the monitoring of international teleconsultation referrals in the era of Covid-19 to facilitate and prevent the suspension of access to care, the most common architecture for teleconsultation, communication technologies and protocols, vital body signals, video transmission, and the conduct of teleconsultation. The aim is to develop a teleconsultation platform to diagnose the patient in real time, transmit data from the remote location to the doctor, and provide a teleconsultation.

A. Yassine (✉) · E. M. Ali · M. Ismail
Energy, Embedded System, and Data Processing Laboratory, National School of Applied Sciences Oujda (ENSAO), Mohammed First University (UMP), 60000 Oujda, Morocco
e-mail: yassine.ayat@ump.ac.ma

E. M. Ali
e-mail: a.elmoussati@ump.ac.ma

M. Ismail
e-mail: ismail.mir@ump.ac.ma

1   Introduction

Telemedicine is now encouraged at the political level. It can provide answers to current health problems such as the decrease in the population of general practitioners and specialists, medical desertification in rural areas, and the growing aging of the population. The current pandemic has led to enormous pressure on hospitals and doctors' surgeries to restructure and often cancel elective treatments and procedures. This unprecedented health crisis has rapidly changed the way doctors provide care to their patients [1].

Telemedicine provides essential support for developing new organizations capable of responding to the challenges facing health care today. Telemedicine must be based on medical programs that respond to the priorities and needs of the population and health professionals in the area [2]. Today, many applications and services of telemedicine have been deployed in the field.

Telemedicine is defined by the World Health Organization (WHO) as "the provision of health care services, where distance was a factor, by health care professionals using information and communication technologies" [3, 4]. Suppose telemedicine is to be effective in managing health emergencies and crises. In that case, we need to ensure that it is appropriately integrated into our health services and becomes common practice for healthcare providers and patients [5]. The health crisis was the starting point for the rapid development of telemedicine in Geneva.

The deployment of HUG@home at the University Hospitals of Geneva (HUG) and of docteur@home, ease use of these remote consultation tools is essential for sustainable telemedicine in practice [6]. Telemedicine, including communication skills with caregivers by phone, video, or SMS instantly, is beyond geographical and temporal constraints. It thus provides better access to care, including a wider range of specialists, reduces health costs, and saves time for patients and doctors. So, the need for training in communication and virtual physical examinations during video consultations appears to be necessary to integrate telemedicine into daily practice while ensuring quality care.

Taking into account the feedback from the ALS reference center of the Lille University Hospital Centre [7], the objective is to offer Amyotrophic Lateral Sclerosis patients' teleconsultations with the ALS center's referral doctor. This new consultation modality aims to avoid traveling for a physical consultation and to maintain quality care by the expert center. Especially during the period of confinement in March 2020 have favored the setting up of teleconsultation facilities at the ALS center in Lille, with more than 192 TCs being carried out. The more advanced patients of the disease favor and request follow-up by TC.

The risk of being infected by the virus due to the number of patients in the waiting room for a consultation has prompted patients to change their mode of access to care. The recourse to teleconsultation engenders a decrease in the activity of doctors during confinement, according to the data of the insurance of 2019, 2020, 2021, and 2022. The exceptional crisis and the recourse to teleconsultation require monitoring and quantification of the deployment of TCs.

In order to anticipate the possible results on public health and examine the effect of the integration of new tools in medicine, particularly teleconsultation as can be seen in Fig. 1 that there is a drop in the use of physical consultation in 2020, especially in the first confinement, which causes a breakdown in access to care for the population, teleconsultation is a solution to avoid a breakdown in access to care, as indicated in the statistics on the use of teleconsultation in Fig. 2, which shows that in the first period of confinement, teleconsultation is the only thing that can improve continuity of care.

Moroccan governmental commission considered a law on the practice of telemedicine (No. 2-18-378) in May 2018, and a Moroccan telemedicine company was born. An organization whose mission is to set up the technological infrastructure to develop telemedicine activities to benefit people in rural and remote areas. In Morocco, [8] where telemedicine is still experimental in the province of Sefrou, the experience of teleconsultation through the website tabib24.com launched by the Ministry of Health during the containment phase only allows for simple medical advice via voice or video chat with a volunteer doctor, but does not allow for the full potential of telemedicine technology to be used.

The rest of this document is organized as follows. Section 2 presents a comparative study targeting countries to represent a diversity of situations in the deployment
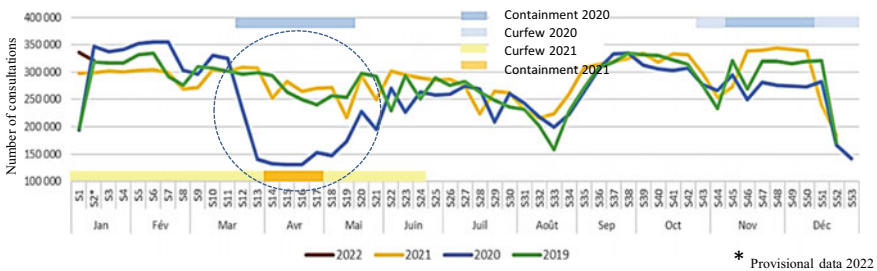
Fig. 1 Weekly evolution of the use of physical consultation in the PROVENCE-ALPES-CÔTE D'AZUR region [9]
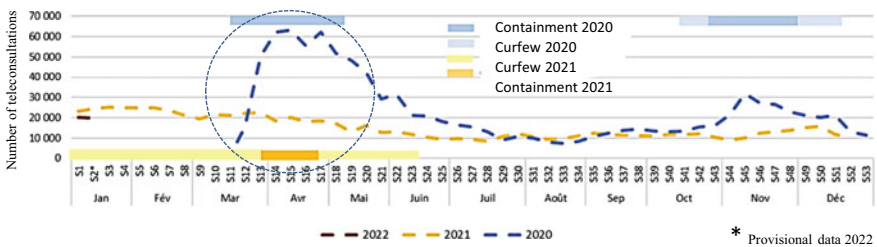
Fig. 2 Weekly evolution of the use of teleconsultation in the PROVENCE-ALPES-CÔTE D'AZUR region [9]

of telemedicine. Section 3 describes an architecture for the implementation of tele-consultation and the technologies (communication protocol, WebRTC…) and the vital human signals. Section 4 presents the sequences of a teleconsultation. Finally, Sect. 5 concludes the paper.

2 Telemedicine Deployment Internationally

Table 1 summarizes the deployment of telemedicine at the international level; it is clear that Morocco needs a complementary practice of telemedicine to compensate the lack of health professionals in order to improve access to care for the population (elderly, medical desert, etc.), which results in the development of the Moroccan health system through the adoption and integration of information technology and communication [10].

3 Teleconsultation IoT Architecture

The IoT has opened up new deployments within the healthcare field that benefits both the patient side by reducing medical expenses, avoiding the risk of having a health crisis and the quality and satisfaction of care [9]. IoT facilitates the task of monitoring patients' health more efficiently. The doctor can monitor patients' adherence to treatment plans or any need for immediate medical attention. In this approach, connected object technology is implemented to communicate between the sensors implemented on the human body and the cloud to visualize them at the platform.

Since teleconsultation relies on the accurate assessment of several vital signals in real time such as blood pressure, heart rate, temperature, blood glucose, implementation of sensor networks to acquire, process, and exchange data between patient and doctor is required. Figure 3 shows the main units of architecture between a doctor and his patient and the exchange of data from sensors in contact with the patient to extract vital signals and the passage of these signals from a gateway to the platform for evaluation.

The patient places each sensor in its appropriate location to capture the desired information. Once the data is processed, the IoT gateway sends it to the cloud for further processing and analysis. Short-range wireless field connectors can hardly communicate in larger networks. Therefore, the gateway uses long-distance wireless field connectors such as cellular data, satellite, LTE. Table 2 summarizes the information about different wireless protocols and their application.

The major signals involve the measurement of magnitudes and are given in Fig. 4.

Table 1 Telemedicine deployment in countries

	Country				
	Germany	Australia	Canada	Morocco	France
Population (in millions)	82.8	23.3	35.2	35.5	67
Density of doctors (per 1000 inhabitants)	4.1	3570	2720	0.71	3.37
Density of hospital beds (per 10,000 inhabitants)	80	38	32	9.9	61
Life expectancy at birth (in years)	80.6	83	82	74.8	82
Internet users (% of population)	90	82.3	86.8	63	77
There is a specific regulatory framework for telemedicine?	Yes	No	No	Yes[1]	Yes
Healthinsurance companies for the coverage of telemedicine acts	Private CMA	Public AMO	Public AMO	Public AMO[2]	Public AMO
Specialized national actor in charge of the development of telemedicine on the territory	No	Yes	Yes	Yes[3]	No
The nature and level of telemedicine services available in the territory	Private offer limited but growing development	Private offer limited sometimes subsidized	Private offer unevenly developed	Very weak[4]	Hybrid offer (public and private abundant)

---

[1] Telemedicine became part of Moroccan law in 2015. A 2018 decree provides a framework for the practice of telemedicine and creates the Moroccan Society of Telemedicine.

[2] Telemedicine procedures, as provided for by law, are reimbursable by the National Health Insurance Agency (ANAM).

[3] The Moroccan Society of Telemedicine (SMT) is an associative structure mandated by the Ministry of Health to strategically pilot the deployment of telemedicine.

[4] There is no public offer; the development of telemedicine in Morocco does not allow conclusions to be drawn about future offer.
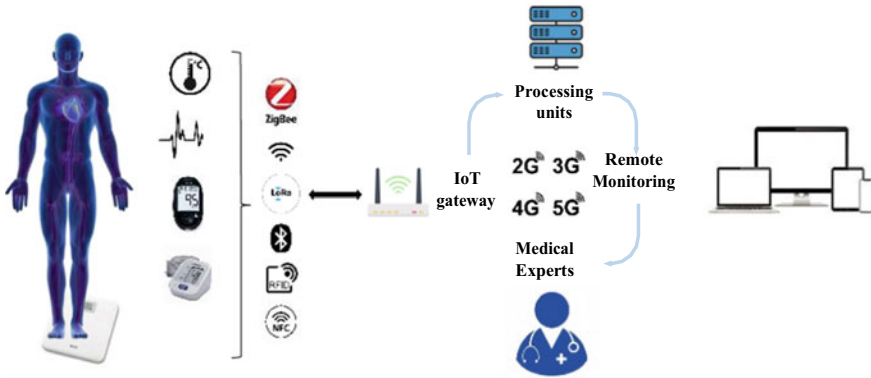
Fig. 3 Patient evaluation architecture from the IoT to the physician

3.1 Video Transmission Solution

The teleconsultation act must be the subject of interpersonal video and audio exchanges between the patient and the health professional. A solution must carry out these exchanges as ergonomic and easy to use for all users. The solution enabling interpersonal exchange may be based on proprietary technology (e.g., flash, silver light, etc.) or open source (e.g., WebRTC). All related exchanges of personal health data must comply with health data confidentiality guidelines and require the information system's data hosting server to be HDS certified.

Web Real-Time Communications (WebRTC) is a technology that allows applications and websites to collect and eventually broadcast audio and/or video media and exchange arbitrary data between browsers without going through an intermediary. The set of standards that comprise WebRTC allows for data sharing and peer-to-peer teleconferencing without requiring the user to install plug-ins or any other third-party software. WebRTC serves several purposes and overlaps considerably with the Media Capture and Streams Application Programming Interface (API). Together, they provide powerful multimedia capabilities to the web, including support for audio and video conferencing, file exchange, identity management, and interfacing with existing telephone systems. Peer-to-peer connections can be established without special drivers or plug-ins and can often be established without any intermediate server.

Table 3 presents the advantages and specifications of WebRTC, which is recommended for peer-to-peer communication and is, therefore, the most widely used communication platform and teleconsultation platform.

Table 2 Wireless technologies and their application

Standards	Range (m)	Data rate	Frequency	Applications	Purpose
ZigBee	10–20 m	250 kbit/s	868/915 MHz; 2.4 GHz	Bowel activity monitoring device [11] Bandage-size non- ECG heartrate monitor Digital stethoscope [12]	Gastrointestinal motility detection after abdominal surgery Heart sound monitoring
Blue-tooth	10–100 m	1 mbit/s	2.4–2.5 GHZ	Digital stethoscope Wearable Fitbit tracker	Heart sounds detec tion and monitoring fitness health track ing during physical activities [13]
RFID	Up to 3 m	Varies with frequencies	LF/HF/UHF/micro wave	Medical equipment tracking system The wearable RFID tag antenna	Quickly locating and tracking medical Equipment under critical time constraints [14]
2 g/3 g	1– 8 km	Varies on network type	Varies on network type	Wireless wearable	Monitor health status remotely Real-time personal skin-caring
Satellite and 3 g/4 g	160–3600 km	Varies on model	Varies on model	Real-time wireless health monitor	Track and approach patients upon emergencies [15]

4 Conduct of Teleconsultation

The teleconsultation process is the same as a traditional physical consultation (face-to-face at the doctor's office). Firstly, a request for an appointment is sent spontaneously or by recommendation of the attending doctor. For this, the doctor sends an e-mail address containing the invitation to join the session at the scheduled appointment time on a secure website or application. A computer or tablet equipped with a webcam and connected to the internet is required. Patient approval prior to the teleconsultation is mandatory regardless of the situation or pathology. At the end

Electrical

(ECG), (EMG), (EEG), eyes (EOG), superficial galvanic response. These very weak signals (millivolt scale for ECG, microvolt scale for EEG) are commonly collected using pre-gelled electrodes (Ag-AgCl) which form an electrode-electrolyte couple with the skin to allow ionic exchange (Fem1), but more and more "dry" electrodes are used, which are more hypoallergenic, exploiting the capacitive effect of the electrode-skin couple;

Thermal

surface temperature, pseudo-internal temperature, subcutaneous tissue hydration rate. These are thermistors for passive measurements, sometimes associated with regulated heat sources in the case of active sensors active sensors;

Mechanical

Pression postures, weights, forces, center of pressure. These are mainly pressure or acceleration sensors, integrated on silicon (MEMS), using capacitive, piezoresistive and piezoelectric effects capacitive, piezoresistive and piezoelectric effects;

Optics

imaging, colors, photoplethysmography. They are either video imagers (CCD cameras) or assemblies of photodiodes and phototransistors (measurement of blood opacity for pulse oximetry);

Fig. 4 Principal physiological parameters

Table 3 WebRTC advantages and specifications

Web RTC advantage	Specification
Free open source	• It provides browsers with end-to-end direct communication and allows developers to facilitate this connection easily
Speed enhancement	• It no longer needs to be routed through a server; it reduces latency and bandwidth consumption • Direct communication improves the speed of data transfer and file sharing
No third-party app required	• Requiring no additional software, plug-ins, or continuous server involvement • Easily be embedded in any website and connect peers across the internet
Provide a safe connection across many browsers	• Encryption is mandatory for all of the WebRTC components. Since it is not a plug-in, it runs inside the browser's sandbox without creating a new process so that no malware can get into the user's system

of the teleconsultation session, the doctor may issue a prescription (for medication or additional tests), which will be sent to the patients in paper form, by post, or in electronic form, in particular via a secure messaging system, under conditions that guarantee the confidentiality and security of the exchanges. Figure 5 summarizes the steps of teleconsultation.
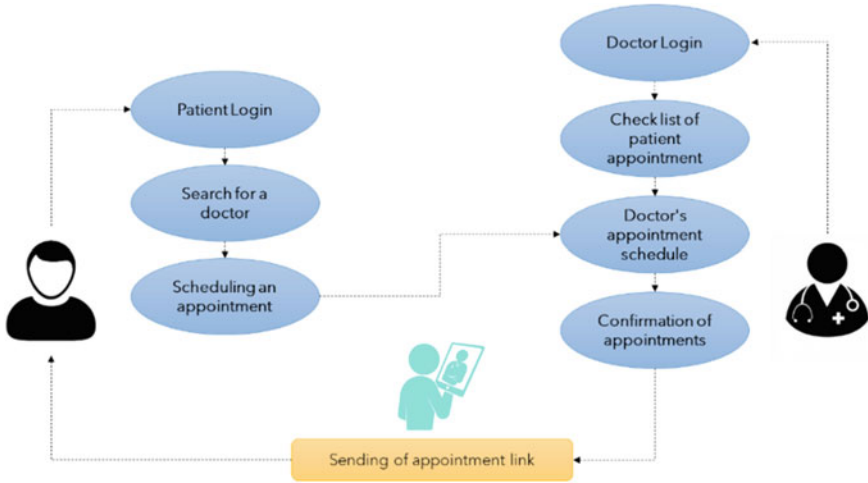
Fig. 5 Consultation processes

5 Conclusion

The integration of new information technologies in the healthcare sector, and more particularly in public sector hospitals, has been a major priority in developing countries, and the implementation of these technologies has been validated as a tool for improving the quality of care, reducing errors, facilitating the tasks of health professionals and the relationship with patients. It also offers tools to improve the continuity of care at a reduced cost and automate the execution of clinical processes to allow health professionals to devote more time to their patients. Telemedicine covers several acts, namely teleconsultation, tele-expertise, remote medical monitoring, remote medical assistance, and remote medical response. This article focuses on the study of the recourse toward the act of teleconsultation in the era of Covid-19, the deployment of telemedicine at the international level and the implementation of an architecture of the sensor networks and the technology of communication to make the acquisition of the vital signals for monitoring the state of the patient's health at the doctor as well as to analyze the technology most responded in the world of the peer-to-peer communication (WebRTC) and the scenario for the teleconsultation process.

As a perspective to this works, the development of a teleconsultation platform at the level of Morocco is planned, contributing to the acceleration of the national deployment of telemedicine. In addition, the project will prepare a prototype of a sensor pack connected to this platform to provide information to the doctor in real time in order to guarantee the quality of the care.

References

1. Kichloo A, Albosta M, Dettloff K, Wani F, El-Amir Z, Singh J, Chugh S et al (2020) Telemedicine, the current COVID-19 pandemic and the future: a narrative review and perspectives moving forward in the USA. Family Med Commun Health 8(3)
2. Haleem A, Javaid M, Singh RP, Suman R (2021) Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2:100117
3. World Health Organization (2010) Telemedicine: opportunities and developments in member states: report on the second global survey on eHealth. Glob Obs eHealth Ser 2
4. Mazouri-Karker S, Karsegard J (2020) Cross glances on the videoconsultation. Rev Med Suisse 16(706):1709–1713
5. Geissbuhler A, Mazourikarker S (2020) Telemedicine at the heart of the COVID crisis: an unexpected boom! Rev Med Suisse 1694–1694
6. Vendeuvre LP, Lecygne C, Jeannot JG, Spahni S, Mazouri-Karker S (2020) Telemedicine in the COVID-19 era: a revolution? Experience of the Geneva University Hospitals. Rev Med Suisse 6(706):1695–1698
7. Danel V et al (2021) ALS and teleconsultation: Hauts-de-France experience. Neurol Rev 177:S154–S155
8. Rouidii M, Elouadi A, Hamdoune A (2020) The acceptance of telemedicine among Moroccan doctors: a qualitative study using the UTAUT model. Moroccan J Quant Qual Res 2(3):41–53
9. Davin-Casalena B, Jardin M, Guerrera H et al (2021) The impact of the COVID-19 epidemic on primary care in the Provence-Alpes-Côte d'Azur region: feedback on the implementation of a real-time monitoring system based on data regional health insurance. J Epidemiol Publ Health 69(3):105–115
10. Al-Turjman F, Nawaz MH, Ulusar UD (2020) Intelligence in the internet of medical things era: a systematic review of current and future trends. Comput Commun 150:644–660
11. Ulusar UD, Turk E, Oztas AS, Savli AE, Ogunc G, Canpolat M (2019) IoT and edge computing as a tool for bowel activity monitoring. In: Al-Turjman F (ed) Edge computing: from hype to reality. Springer International Publishing, Cham, pp 133–144
12. Patil DD, Shastri RK (2012) Design of wireless electronic stethoscope based on zigbee. arXiv preprint arXiv:1202.1680
13. Fitbit Official Site for Activity Trackers and More (n.d.) https://www.fitbit.com/eu/home. Accessed 24 July 2021
14. Lopez-Soriano S, Parron J (2015) Wearable RFID tag antenna for healthcare applications. In: IEEE-APS topical conference on antennas and propagation in wireless communications (APWC 2013). IEEE, Torino, Italy, pp 287–290
15. Aziz K, Tarapiah S, Ismail SH, Atalla S (2016) Smart real-time healthcare monitoring and tracking system using GSM/GPS technologies. In: 3rd MEC international conference on big data and smart city (ICBDSC), pp 1–7

SiNW-ISFET Sensor Modeling Using the *k*-Nearest Neighbor Machine Learning

Nabil Ayadi, Hajji Bekkay, Ahmet Lale, Jerome Launay, and Pierre Temple-Boyer

Abstract The silicon nanowire (SiNW)-based ISFET sensor is experiencing tremendous development due to its multiple attractive advantages (small size, low cost, robust, and portable real time). The ISFET sensor is widely used for pH detection but also for patient health monitoring, cancer detection, water and soil content analysis, and other applications. Currently, few predictive models are used to estimate the performance of the ISFET sensor and determine the most influential parameters. In this study, a new model based on the machine learning (ML) technique named *k*-nearest neighbor (*K*NN) is developed to estimate the performance of silicon nanowire ISFET. The *K*NN model was tested for different values of training data and compared with experimental results. A good agreement was observed with the experimental measurements. Moreover, the test results show that the accuracy percentage of the *K*NN model with 90% training data could reach up to 99.96% and could meet the practical demand.

Keywords SiNW · Sensor · Machine learning (ML) · *K*-nearest neighbor (*K*NN)

1 Introduction

ISFET sensors have undergone unprecedented development over the past 30 years. This development has been identified at all levels: new design, new detection materials, and new measurement circuits compatible with microelectronic technology [1]. ISFET sensors are widely used in multiple applications: pH detection, patient monitoring during surgery, water quality monitoring, food industry, etc. [2, 3].

However, some problems persist, such as the sensitivity which is limited to 59 mv/pH (Nerst sensitivity), the effect of temperature which affects the measurements

N. Ayadi (✉) · H. Bekkay
Laboratory of Renewable Energy, Embedded System and Information Processing, National School of Applied Sciences, Mohammed First University, Oujda, Morocco
e-mail: nabil.ayadi.ma@gmail.com

A. Lale · J. Launay · P. Temple-Boyer
LAAS-CNRS, Université de Toulouse, CNRS, UPS, Toulouse, France

357

and the drift of the measurements over time. This led researchers to innovate a new type of SiNW-ISFET sensor, which has interesting advantages: high sensitivity, fast response time, low cost [4].

To support this technological development, the researchers have simultaneously developed high-performance design tools and methods for modeling ISFET sensors. Several models have been proposed to evaluate the performance of the SiNW-ISFET sensor, such as PSPICE, EKV, and BSIM combined with the site-binding model or using the TCAD tool—COMSOL Multiphysics [5]. However, at present, few predictive models are used to estimate the performance of SiNW-ISFET and the most influential parameters.

Some experts believe that artificial intelligence (AI) is the most suitable solution to effectively predict the performance of the SiNW-ISFET sensor [6]. Broadly, there are three types of machine learning algorithms: supervised learning, unsupervised learning, and reinforcement learning.

In supervised learning, algorithms make predictions based on a set of labeled examples that you provide. In unsupervised learning, the data points aren't labeled; the algorithm labels them for you by organizing the data or describing its structure. Reinforcement learning uses algorithms that learn from outcomes and decides which action to take next. After each action, the algorithm receives feedback that helps it to determine whether the choice it made was correct, neutral, or incorrect.

In this work, the k-nearest neighbor (KNN) technique, a supervised machine learning (ML) technique, is used for the performance prediction of the SiNW-ISFET sensor. This technique is used for both regression and classification. Its working principle is to predict the correct class for the test data by calculating the distance between the test data and all the learning points. Then, the number k of points closest to the test data is selected.

This paper is structured in four sections: Section 1 presents an introduction to ISFET sensors, Sect. 2 develops the KNN machine learning model applied to the SiNW-ISFET sensor, Sect. 3 is devoted to the discussion of the main results obtained, and the conclusion of this work is presented in Sect. 4.

2   SiNW-ISFET Model in Machine Learning Technique

2.1   Model in k-Nearest Neighbor (KNN)

KNN is a machine learning technique applied to classification and regression. The principle of KNN regression is to choose the number of k-nearest neighbors to use in the prediction. The nearest neighbors can be defined as the points with the shortest distance and at an unknown point on its circumference [7]. The data from the simulation were used to predict ($k = 2$), and the power exponent (p) was fixed at 2. The technique has been applied in the Python language. Several ways to extract the neighbor distance include the Minkowski distance function as given in Eq. (1) [8].

$$d(\hat{x}, x_i) = \left(\sum_{j=1}^{n} |\hat{x}_j - x_{ji}|^p \right)^{\frac{1}{p}} \tag{1}$$

where $\hat{x} = (\hat{x}_1; \hat{x}_2; \ldots; \hat{x}_n)$ predicted sample, $x_{ji} = (x_{1,i}; x_{2,i}; \ldots; x_{n,i})$ reference sample, p power exponent, and i the reference sample's increasing rank, in order similarity.

The *k*-nearest neighbor (*K*NN) technique is determined as the distance between the neighbors by applying the *K*NN algorithm. Then, the output result of this method is represented by Eq. (2) [8, 9].

$$\hat{y}(\hat{x}) = \sum_{i=1}^{k} w_i \, y(x_i) \tag{2}$$

where y response variable and $\hat{y}$ response variable prediction.

$$w_i = d(\hat{x}, x_i)^{-p} / \sum_{i=1}^{k} d(\hat{x}, x_i)^{-p} \tag{3}$$

where w_i weighting factor for x_i and k number of neighbors used for interpolation [10].

In this study, the objective output parameter is the source–drain current (I_{ds}) which is estimated as a function of the gate voltage (V_{gs}), the nanowire length (l_{nw}), the number of wires (N), the gate length (L), and the (pH). Those parameters are then inputs data to the *K*NN model, which is given by Eq. (4).

$$(x_1; x_2; x_3; x_4; x_5) = (V_{gs}; l_{nw}; L; N; pH) \tag{4}$$

where $\hat{x} = (\hat{x}_1; \hat{x}_2; \ldots; \hat{x}_n)$ are the predicted samples, as shown in the formula (5):

$$(\hat{x}_1; \hat{x}_2; \hat{x}_3; \hat{x}_4; \hat{x}_5) = \left(\widehat{V_{gs}}; \widehat{l_{nw}}; \hat{L}; \hat{N}; \widehat{pH} \right) \tag{5}$$

According to the *K*NN model computation steps, in the first step, we compute the Minkowski distance function using Eq. (1), given the power exponent ($p = 2$) and $i = 1, \ldots, k$ for our case $k = 2$.

$$d(\hat{x}, x_1) = \sqrt{\left(\widehat{V_{gs}} - V_{gs1} \right)^2 + \cdots + \left(\widehat{pH} - pH_1 \right)^2} \tag{6}$$

$$d(\hat{x}, x_2) = \sqrt{\left(\widehat{V_{gs}} - V_{gs2} \right)^2 + \cdots + \left(\widehat{pH} - pH_2 \right)^2} \tag{7}$$

In the second step, we compute the weights associated with the two neighbors, using Eqs. (3), (6), and (7):

$$w_1 = \sqrt{d(\hat{x}, x_1)}/\sqrt{d(\hat{x}, x_1)} + \sqrt{d(\hat{x}, x_2)} \tag{8}$$

$$w_2 = \sqrt{d(\hat{x}, x_2)}/\sqrt{d(\hat{x}, x_1)} + \sqrt{d(\hat{x}, x_2)} \tag{9}$$

Finally, the output corresponding to the predicted sample has been evaluated by Eq. (2).

$$I_{ds}(k\text{NN}) = w_1 I_{ds1} + w_2 I_{ds2} \tag{10}$$

Figure 1 shows the flowchart of the k-nearest neighbors (KNN) program with the computational phases. First, the input data are presented [gate voltage (V_{gs}), nanowire length (lnw), number of wires (N), gate length (L), and the (pH)] and the outputs are defined (Ids_experimental). Then, the choice of power exponent p and the number k of neighbors are used for interpolation. Next, training and testing of the data are performed. Finally, the predicted source–drain current (I_{ds}) is estimated.

Fig. 1 Program flowchart of the k-nearest neighbors (KNN)
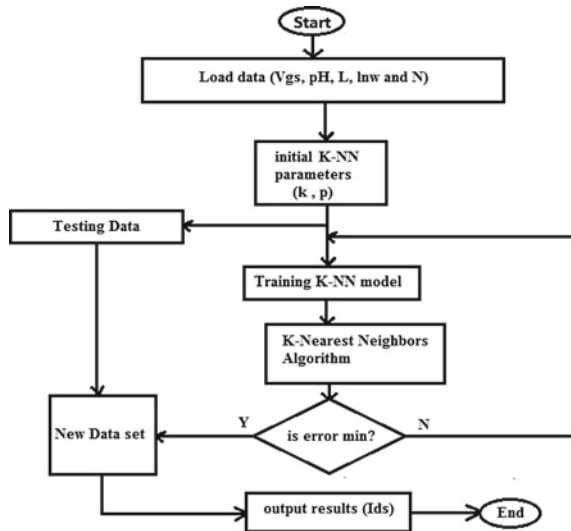
2.2   Evaluation and Validation of the Quality Performance

In this study, two performance indicators were applied to assess the quality of the developed model with respect to the experimental results, which are the coefficient of determination (R^2) and the mean square error (RMSE). Both indicators are represented by Eqs. (11 and 12).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{11}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{12}$$

where n means the number of data items in the entire test process; y_i and $\hat{y}_i$ are the correct value and the predicted value of the target [11, 12].

3   Results and Discussion

In this part, we present the results of the machine learning technique, namely *k*-nearest neighbor (*K*NN), for the performance prediction of the silicon nanowire ISFET sensor. About 300 data were used for training the *k*-nearest neighbor (*K*NN) model. This model is developed using Python language. It is the most popular programming language for data science and open access communities.

To test and validate the efficiency of the SiNW-ISFET sensor model, experimental data are provided by the Laboratory for Analysis and Architecture of Systems (LAAS-CNRS, Toulouse, France).

3.1   The Effectiveness of the KNN Model

Figure 2 shows the $I_{ds} - V_{gs}$ plot of SiNW-ISFET for different pH values 4, 7, and 10. The length of the nanowire used is 20 μm, and the drain–source voltage is fixed at 1 V ($V_{ds} = 1$ V). Measurement data are displayed as circles and simulation results as symbols for the *k*-nearest neighbor (*K*NN) model. It is found that a good agreement is obtained between the SiNW-ISFET measurement data and the simulation results of the *k*-nearest neighbor (*K*NN) model with negligible errors. These results prove that the *k*-nearest neighbor (*K*NN) model is accurate and efficient.

Figure 3 demonstrates the variation of the gate voltage V_{gs} of the SiNW-ISFET sensor as a function of pH. The measured pH sensitivity of the SiNW-ISFET sensor is 54 mV/pH, while that estimated by the *k*-nearest neighbor (*K*NN) model is

Fig. 2 Verification between the results of the technique for machine learning and measurement of SiNW-ISFET of a single nanowire 20 μm long, with a gate length of 3.73 μm and different pH
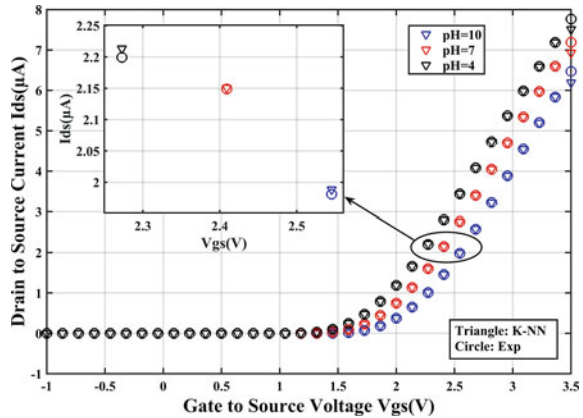
Fig. 3 Comparison the sensitivity of the SiNW-ISFET of the KNN technique with the experimental data ($I_{ds} = 1\,\mu A$ and $V_{ds} = 1$ V)
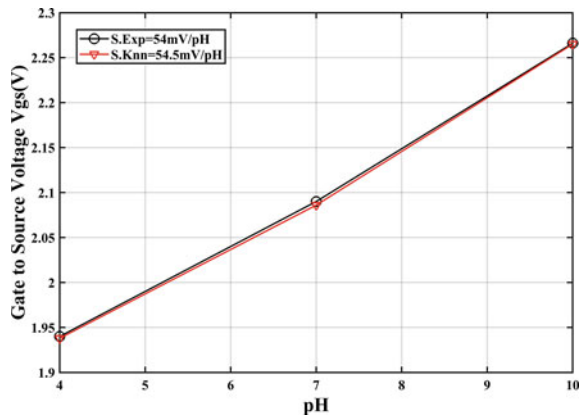
54.4 mV/pH. We note that the k-nearest neighbor (KNN) model gives a very precise estimation of the sensitivity of the SiNW-ISFET sensor.

Figure 4 shows that the KNN model fits well to the measurement data carried out for pH = 4, 7, and 10. In addition, after training with 100 data, the model was able to estimate with great precision the tracing of the $I_{ds} - V_{gs}$ curve for pH value = 8.5. The estimated errors were 0.9995 for R^2 and 5.4×10^{-8} for RMSE.

3.2 Study the Uncertainty Indices for the Training Data Set

The performance of the KNN model is evaluated based on the R^2 and RMSE parameters. Figures 5 and 6 illustrate the evolution of these parameters (R^2 and RMSE) as a function of the percentage of the KNN model training data: 70, 75, 80, 85, and 90%. From Fig. 5, it can be seen that the accuracy of the KNN model tends to increase with

Fig. 4 Prediction (pH = 8.5) of the results of the technique (*K*NN) for machine learning of SiNW-ISFET of a single nanowire 20 μm long, with a gate length of 3.73 μm and different pH
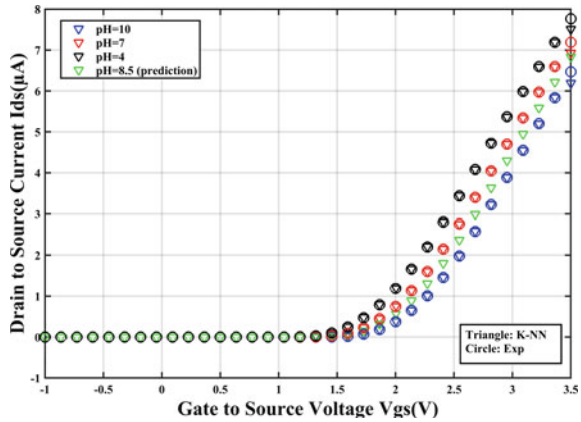
the increasing percentage of training data. This precision can reach up to 99.965% (R^2 is equal to 0.99965) with a percentage of 90% of the training data.

Figure 6 shows that the root mean square error (RMSE) decreases as the *K*NN model training data of the SiNW-ISFET sensor increases. The RMSE displays a value of 5.4×10^{-8} for 70% of the training data, while it can reach 4.6×10^{-8} for 90% of the training data.
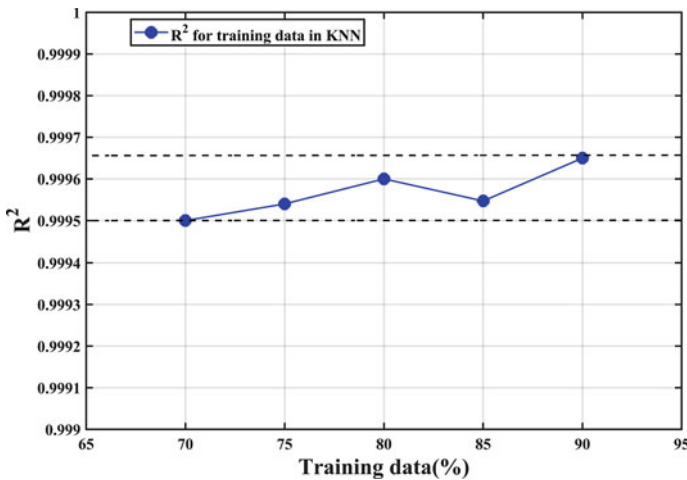
Fig. 5 R^2 according to training data

Fig. 6 RMSE according to training data
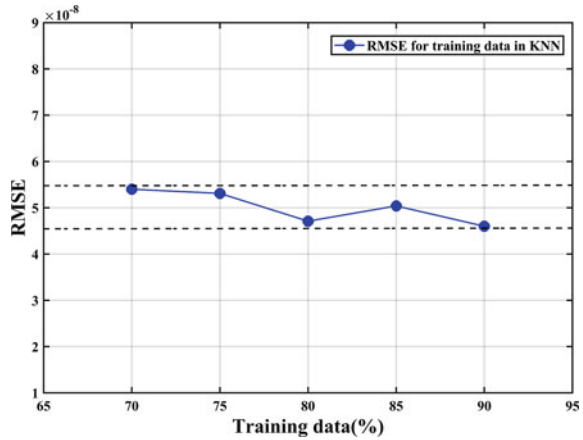
4 Conclusion

In this paper, a new model based on a machine learning technique named *k*-nearest neighbor (*K*NN) was developed to estimate the performance of silicon nanowire ISFET sensor. This model has been implemented in the Python environment.

The developed model was compared to the experimental data, and a good agreement was observed. The performance of the *K*NN model was evaluated based on the R^2 and RMSE parameters. The results obtained indicate that the RMSE displays a minimum value of 5.4×10^{-8}, while R^2 shows a value of 0.9995, close to 1.

Additionally, this *K*NN model was tested against the percentage of training data. It turned out that the model becomes more accurate whenever the percentage of training data exceeds the 70%.

As a result, the *K*NN model is more efficient in predicting the performance of the silicon nanowire ISFET sensor and could meet the practical demand.

References

1. Bergveld P (1970) Development of an ion-sensitive solid-state device for neurophysiological measurements. IEEE Trans Biomed Eng BME-17:70–71
2. Umenyuk I et al (2006) Development of pNH4-ISFETS microsensors for water analysis. Microelectron J 37:475–479
3. Choi S et al (2020) Experimental extraction of stern-layer capacitance in biosensor detection using silicon nanowire field-effect transistors. Curr Appl Phys 20:828–833
4. Ayadi N et al (2021) Simulation and performance study of silicon nanowire (Si-NW) field-effect transistor (FET) pH microsensor. In: Lecture notes in electrical engineering, vol 681. Springer, Singapore. https://doi.org/10.1007/978-981-15-6259-4_42
5. Ayadi N et al (2022) Study of parameters influencing on the performance of SiNW ISFET sensor. In: Lecture notes in electrical engineering, vol 745. Springer, Singapore

6. Liu Q, Hu X, Ye M, Cheng X, Li F (2015) Gas recognition under sensor drift by using deep learning. Int J Intell Syst 30(8):907–922
7. Hamed Y, Mustaffa ZB, Idris NRB (2015) An application of *K*-nearest neighbor interpolation on calibrating corrosion measurements collected by two non-destructive techniques. In: 2015 IEEE 3rd international conference on smart instrumentation, measurement and applications (ICSIMA). IEEE, pp 1–5
8. Mahdiani MR et al (2020) Modeling viscosity of crude oil using *k*-nearest neighbor algorithm. Adv Geo Energy Res 4(4):435–447
9. Hamed Y et al (2020) Two steps hybrid calibration algorithm of support vector regression and *K*-nearest neighbors. Alex Eng J. https://doi.org/10.1016/j.aej.2020.01.033
10. Piikki K et al (2013) Sensor data fusion for topsoil clay mapping. Geoderma 199:106–116
11. Shataee S et al (2012) Forest attribute imputation using machine-learning methods and ASTER data: comparison of *k*-NN, SVR and forest regression algorithms. Int J Rem Sens
12. Pal R (2017) Validation methodologies. In: Predictive modeling of drug sensitivity 2017, pp 83–107 (Chapter 4). https://doi.org/10.1016/B978-0-12-805274-7.00004-X

Portable Device for Real-Time Monitoring of Blood Samples

Ayat Yassine, El Moussati Ali, Choukri Mohammed, Mir Ismail, and Benslimane Anas

Abstract Medical samples (tissues, cells) and accompanying documents require bar-coding to be able to identify the donor of the tissue or cells collected, the location, and the date and time of collection. The time at which the sample is taken this time sets the starting point for the storage time. This paper proposes an application of the Internet of Things in the medical field to resolve the problem of no-conformity of blood samples. A hand-held scanner has been developed to communicate with the developed server to determine and save in real-time the start of the sample conservation and identify the person from the blood vial. A device based on a barcode reader, from the barcode labeled in a vial. We have the possibility to know the person concerned if the sample was already made or not. Once the sample will be made, the sampler has the responsibility to click on the button to validate on the system that a sample has been made at a time will be defined automatically and the information will be returned to our server. Therefore, the proposed system aims at monitoring the date and time of sampling which determines the time and delay of the transport of samples to the laboratory of each hospital department to keep the sample in the best conservation conditions.

Keywords Internet of Things (IoT) · Wireless sensor network · Digital health-systems · 3D conception

---

A. Yassine (✉) · E. M. Ali · M. Ismail · B. Anas
Energy, Embedded System and Data Processing Laboratory, National School of Applied Sciences Oujda (ENSAO), Mohammed First University (UMP), 60000 Oujda, Morocco
e-mail: yassine.ayat@ump.ac.ma

E. M. Ali
e-mail: a.elmoussati@ump.ac.ma

M. Ismail
e-mail: ismail.mir@ump.ac.ma

B. Anas
e-mail: a.benslimane@ump.ac.ma

C. Mohammed
Central Laboratory, Mohammed VI University Hospital Centre, Oujda, Morocco

1 Introduction

IoT is recognized as one of the most crucial future technology fields and attracts attention from a wide range of industries. The Internet of Things, also known as the Industrial Internet, is a new technology model envisaged as a global network of machines and devices capable of interacting with each other [1]. These forms of connection allow new masses of data to be gathered on the network and thus new forms of knowledge. Considered as the third evolution of the Internet is called Web 3.0 (sometimes perceived as the generalization of the Web of objects but also as that of the semantic Web), which follows the era of the social Web. The Internet of Things is universal and designated connected objects with various uses in e-health, home automation or quantified self [2]. According to ISO 15189 2007 [3], the pre-analytical phase represents a series of steps beginning chronologically with the clinician's prescription of the tests [4], including the preparation of the patient for sample collection, specimen collection, the timing of sample collection specimen, the authenticity of specimen identification the patient's contact details and a properly completed test request form, the choice of the right tube or container for the sample, the transport of the sample to the laboratory ending at the start of the analytical procedure.

This paper proposes an application of the Internet of Things in the medical field to resolve the problem of no-conformity of blood samples, including:

- Identify the person from the barcode labeled in a blood vial.
- Monitoring the date and time of sampling determines the time and delay of transporting samples to the laboratory of each hospital department to keep the sample in the best conservation conditions.
- Increase the reliability of analytical results in the biochemistry laboratory from the following three criteria:

  - **Conformity** Control of the non-conformity of the pre-analytical phase to ensure reliable analytical results.
  - **Traceability** One of the laboratory's quality assurance program elements begins with receiving and identifying biological samples. In ISO 15189, a misidentification of a sample leads to a repeat sample [5].
  - **Exchange and save data** Retrieve from the database any information related to the patient Name, type of analysis, service, etc., from the code barre scanning.

The rest of this paper is organized as follows. Section 2 describes the type of no-conformity. It also explains different causes of the no-conformity of the pre-analytical phase. Section 3 presents the proposed system, design, and implementation. Section 4 concludes the paper.

2 Description of Type of No-Conformity (NC)

Pre-analytical (PA) errors could invalidate the proper conduct of the analyzes and the reliability of the results. It is at the origin of 85% of the errors and faults that affect the results of analyses [6].

The pre-analytical phase is divided into two stages: The first is outside the biomedical laboratory (external pre-analytical phase), and the second is inside the laboratory (internal pre-analytical phase).

2.1 External Part of the Pre-analytical Phase

The external part is done by the prescriber and the sampler, whose roles end when the samples have arrived at the laboratory in a state that meets the biologist's condition as expected by the biologist [7].

In this phase, the NC concerns the failure to respect the sampling time (Figs. 1 and 2), particularly for certain biochemical parameters that have a circadian cycle, such as Cortisol. The realization of the glycemic cycle, the postprandial glycemia, or the HGPO is also affected by this type of no-conformity (NC).

Fig. 1 NC (no-conformity) of the external part of the pre-analytical step
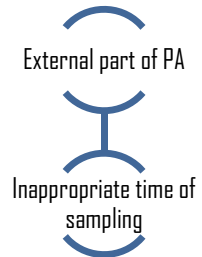
Fig. 2 NC (no-conformity) of the internal part of the pre-analytical step
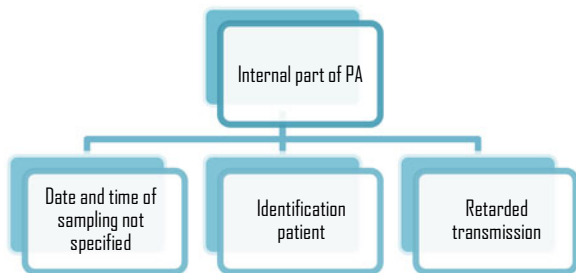
2.2 Internal Part of the Pre-analytical Phase

The NCs noted are related to the sampling procedure conditions.

2.3 Existing Systems and Their Limitations

CHU (university hospital center) adopts the patient management system within the different services in Oujda city. In this system, we can register any patient for all services and departments, two important dates of registration and prescription.

Limitations:

- Cannot identify the date and time of sampling, which determines the time and delay of the delivery of samples to the laboratory;
- Lack of traceability of samples;
- There is no system in existence to manage the pre-analytical phase.

3 Proposed System

A wireless hand-held scanner has been developed to communicate with the server to determine and save in real time the start of the sample conservation and identify the person from the blood vial (Fig. 3).

Barcode scanner is used to scan barcodes, present on the medical biological examination. ESP32 is a series of microcontrollers, a feature-rich MCU, with integrated Wi-Fi, Bluetooth and LoRa connectivity for various applications. We use it to treat the information and ensure communication with the server.
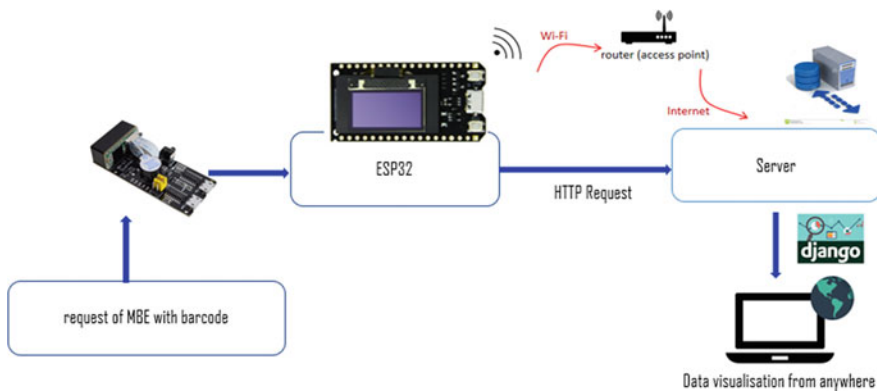
Fig. 3 Architectural diagram for IOT hand-held device

The application has been developed with the Django framework for real-time data visualization. Design is a creative process in which a system organization is established to satisfy the system's functional and non-functional requirements. Great systems are always decomposed into sub-systems that provide a set of related services. The result of the design process is a description of the software architecture [8].

3.1 Architectural Design

MBE Medical biology examinations contain a prescription of the tests by the clinician [9], information about the patient, and a barcode.

Scanner barcode Barcode Scanner is used to scan barcodes present in blood vials or MBE. The scanned barcode information number is eventually stored in the database, and the barcode scanner is connected and controlled by the ESP32. MH-ET LIVE Scanner v3.0 is a barcode and QR code reader module; it adopts an image processing chip for barcode recognition; it can read quickly and accurately beneath the high reading ability in complex environments.

ESP32 with display It is a series of microcontrollers to control the barcode scanner, treat the information, and ensure communication with the server using Wi-Fi, the display for visualization data. For our prototype, we used the ESP32-based TTGO LoRa32 SX1276 OLED board is an ESP32 development board with an integrated LoRa chip, Wi-Fi module, and a 0.96 in. SSD1306 OLED display [10]. The main objective of using a microcontroller with a communicator module (LoRa and Wi-Fi) instead of a hand-held scanner commercialized, because the microcontroller always allows a flexibility of improvement at the level of the programming and an extension of the inputs/output that is in the order to use other sensor to improve the detection of conformity of the samples and to send this information with module communicator send this information over a long or short distance.

Application The application has been developed by Django using Python, REST API, PostgreSQL, and C++ language (Arduino).

Django. The Django framework has remained true to its origins as a model–view–controller (MVC) server-side framework designed to operate with relational databases.

Nevertheless, Django has stayed up to date with most web development tendencies to operate alongside technologies like non-relational databases (NoSQL), real-time Internet communication, and modern JavaScript practices.

All this to the point, the Django framework is now the web development framework of choice for a wide array of organizations, including the photo-sharing sites Instagram and Pinterest; Public Broadcasting System (PBS); United States, National Geographic [11].

Fig. 4 Data flow of a
RESTful API
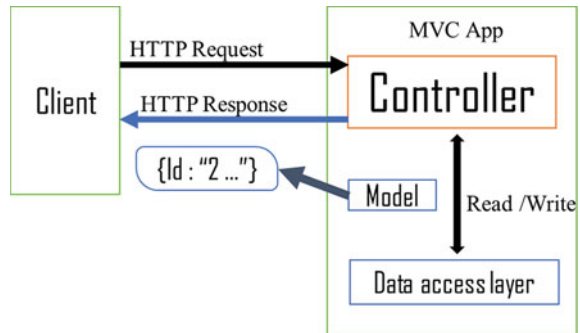
Django REST Framework. By default, Django is not ideal for creating RESTful APIs, as it does not come with a direct way to create APIs. Instead, a parallel, collaboratively funded project called the Django REST Framework. The Django REST framework is a flexible and robust toolkit for building web APIs. Although it is free to use, it is recommended that we fund the project when creating commercial projects [12]. Using Django's REST framework, the MVC logic is moved to the front end. This framework transforms Django's role in the stack into that of a server; it does not display any content but is only responsible for processing client requests to provide and modify data. The RESTful API is shown in Fig. 4 [13].

PostgreSQL is an open-source object-relational database system that uses and extends the SQL language combined with several features that allow the storage and scaling of the most complex data workloads [14].

PostgreSQL began in 1986 as part of the POSTGRES project at the University of California at Berkley and has been under active development ever since.

Fusion 360 is a platform for computer-aided design (CAD), computer-aided manu-facturing (CAM), and computer-aided engineering (CAE). The software allows industrial and mechanical design (surface and mechanical design (surface and volume) and runs on multiple platforms (PC, Mac, and mobile devices). Wu et al. [15] mention that Fusion 360 allows the creation of smooth and accurate surfaces with the surfaces with T-Splines feature as well as with sketch curves and extrusions. We used this software to create our device box Fig. 6.

3.2   Implementation

After the scan, ESP32 communicates with a server to retrieve data from the scanner code name, status, etc. If the status is sampled, display the date of sampling; other-wise, click on the button to register that the sample has been taken and the date will be automatically registered (Fig. 5).

Fig. 5 Portable hand-held device works
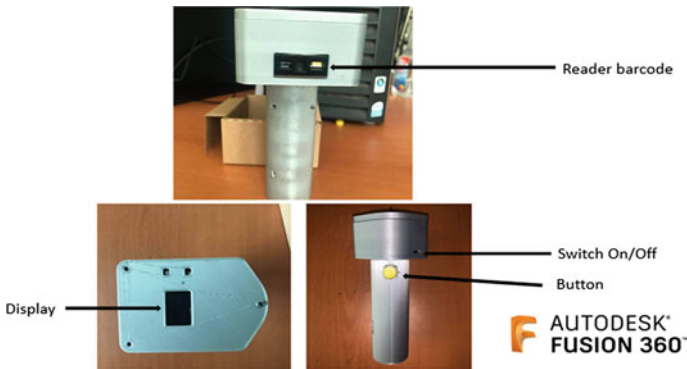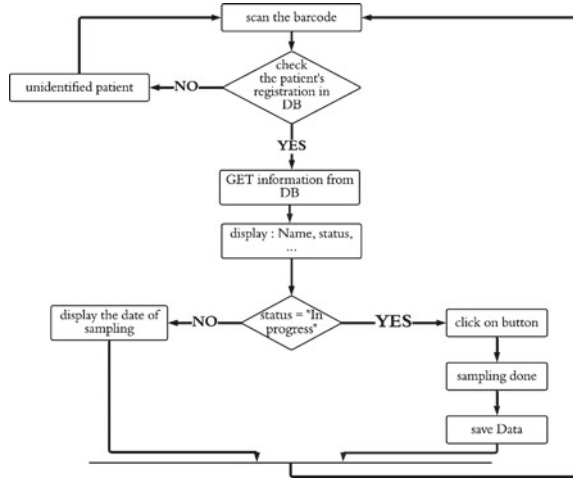
Fig. 6 Portable hand-held device

Here is our prototype (Fig. 6) containing barcode scanner, display, the switch for on/off, and the button has three functions:

- Wake up the microcontroller.
- Scan the label of barcode.
- Register that the sampling is done.

Figure 7a shows our dashboard which displays on the left side the navigator bar (patient registration, graphs as well as the patient data table and the profile of the connected person). In center of the interface, the blue card represents the number of patients that have been sampled. The yellow card shows the number of patients waiting to be sampled. The green card shows the total number of registered patients. Below, the pie chart shows the percentage of samples taken and waiting to be taken (red in progress; blue taken). The line graph shows the number of daily samples taken.

Fig. 7 Application interface: **a** dashboard **b** data of patients

The tables interface Fig. 7b shows the registered patients with several columns last name, first name, status date of prescription and collection. As you can see in the table, the application allows you to reference the date and time of the sample collection. For example in Fig. 7b, in line two so the sample has not been collected yet, the date will be displayed as soon as the sample is collected. This starts the sample storage period; once the sample is collected, the sampler has the responsibility to validate on the hand-held with the yellow button (Fig. 6) to record the date and time.

4 Conclusion

The reliability of laboratory results depends not only on correct analytical technique but on the correct preparation of the pre-analytical phase. Even when run on a good

analyzer, a bad sample will give the wrong result. The concept of IoT is used for various creative. In this paper, an application has been developed that has interfaces to control no-conformity and traceability of the samples take and a wireless device to validate and check the status of the samples, identify the patients and check the date of the samples with a barcode scanner. The information from the scanning device is sent to the application.

The proposed system can be deployed:

- In the organizations that require verification of stock.
- The system can also be used for any type of asset scanning.
- The system can be used anywhere.

References

1. Khan MA, Salah K (2018) IoT security: review, blockchain solutions, and open challenges. Future Gener Comput Syst 82:395–411
2. Mougeot M, Stoltz G (2015) Mathilde Mougeot: connected statistics
3. ISO Document 15189 (2007) Laboratoires d'analyses de biologie médicale: exigences particulières concernant la qualité et la compétence, 2nd edn. ISO, Geneva
4. Plebani M, Lippi G (2010) Is laboratory medicine a dying profession? Blessed are those who have not seen and yet have believed. Clin Biochem 43:939–941
5. Lippi G, Blanckaert N, Bonini P, Green S, Kitchen S, Palicka V et al (2009) Causes, consequences, detection, and prevention of identification errors in laboratory diagnostics. Clin Chem Lab Med 47:143–153
6. Wiwanitkit V (2001) Types and frequency of preanalytical mistakes in the first Thai ISO 9002:1994 certified clinical laboratory, a 6-month monitoring. BMC Clin Pathol 1(1):5
7. Duchassaing D (1999) Pre-analytical phase in biochemistry: quality control process. Rev Fr Lab 1999(317):27–34
8. Danawade V, Jakate O, Yadav PV, Ghori M, Kattikar V (2016) IOT based stock verification system using raspberry PI, barcode scanner and android application. Int J Eng Sci Comput
9. Bartoli M, Berny C, Danel V et al (2012) Recommendations for the prescription, implementation and interpretation of medical examinations in biology in the context of severe poisoning. Ann Biol Clin 431–450
10. Gayatri MG (2020) Tracking of dementia patients using GPS & LORA WIFI
11. Rubio D (2017) Beginning Django. Apress
12. Django REST framework (2020) Web document. https://www.django-rest-framework.org/. Last accessed 11 Sept 2021
13. Create a Web API with ASP.NET core and visual studio for windows. Microsoft. Web Document. https://docs.microsoft.com/en-us/aspnet/core/tutorials/first-web-api?view=aspnetcore-5.0&tabs=visual-studio. Last accessed 11 Sept 2021
14. PostgreSQL—About. The PostgreSQL global development group. Web Document. https://www.postgresql.org/about/. Last accessed 11 Sept 2021
15. Wu D, Terpenny J, Schaefer D (2017) Digital design and manufacturing on the cloud: a review of software and services-RETRACTED. Artif Intell Eng Des Anal Manuf 31(1):104–118. https://doi.org/10.1017/S0890060416000305

The Use of GA and PSO Algorithms to Improve the Limitations of a Readout Circuit of an pH-ISFET Sensor

Abdelkhalak Harrak and Salah Eddine Naimi

Abstract A readout circuit described in this paper to enhance the temperature sensibility. An ISFET-based pH sensor type employed, with a temperature compensation had been improved and high linearity over a large range of pH changed from 1 to 12. In the beginning, the general optimization techniques used are the particle swarm algorithm and the genetic algorithm. Then a local optimization was applied to find a combination of geometry of the transistors for a good optimization result. With the optimization techniques used, the simulation results give a very low-temperature dependence of the circuit and good precision.

Keywords Evolutionary algorithm · pH-ISFET · Temperature insensitivity · Linearity

1 Introduction

In micro-nanoelectronics, we find a wide use of algorithms or of course artificial intelligence used to optimize the behavior of a circuit that constitutes from many transistors. The market need for sensors with high measurement accuracy requires good optimization of different physical quantities of circuits, hence the arrival of artificial intelligence [1]. Nowadays, the use of ISFETs chemical sensors is essential to confront the problems of detection of chemical and/or biological elements in several fields. Many of the devices are developed for the chemical and biological processing such as enzyme-substrate reaction, antigen-antibody bonding and protein-protein interaction. The adaptation and study of this circuit must have the implementation of optimization algorithms. Temperature is an important factor that affects the pH measurement. The working mechanism of a pH-ISFET is described by $I(V)$ equations similar to the MOS transistor. Studies that have been carried out

A. Harrak (✉) · S. E. Naimi
ICSR Team, ESETI Laboratory, Mohammed First University, Oujda, Morocco
e-mail: abdelkhalak.harrak@usmba.ac.ma

S. E. Naimi
e-mail: s.naimi@ump.ac.ma

by some research show that pH-ISFET is very sensitive to temperature, which gives unreliable measurements. For this, it is very important to find good methods and techniques to study the temperature behavior of ISFET sensors in order to improve their temperature stability [2–4]. In this work, we will study the temperature sensitivity of the circuit and linearity by using the genetic algorithm and particle swarm optimization.

2 The Main Blocks of the Proposed Circuit

The readout circuit has been proposed to improve the linearity of the output signal and to compensate the effect of temperature of the circuit. This readout circuit operates in a differential mode, as shown by the circuit structure in Fig. 1. The readout circuit consists of three main stages, and also of two power supply stages, these
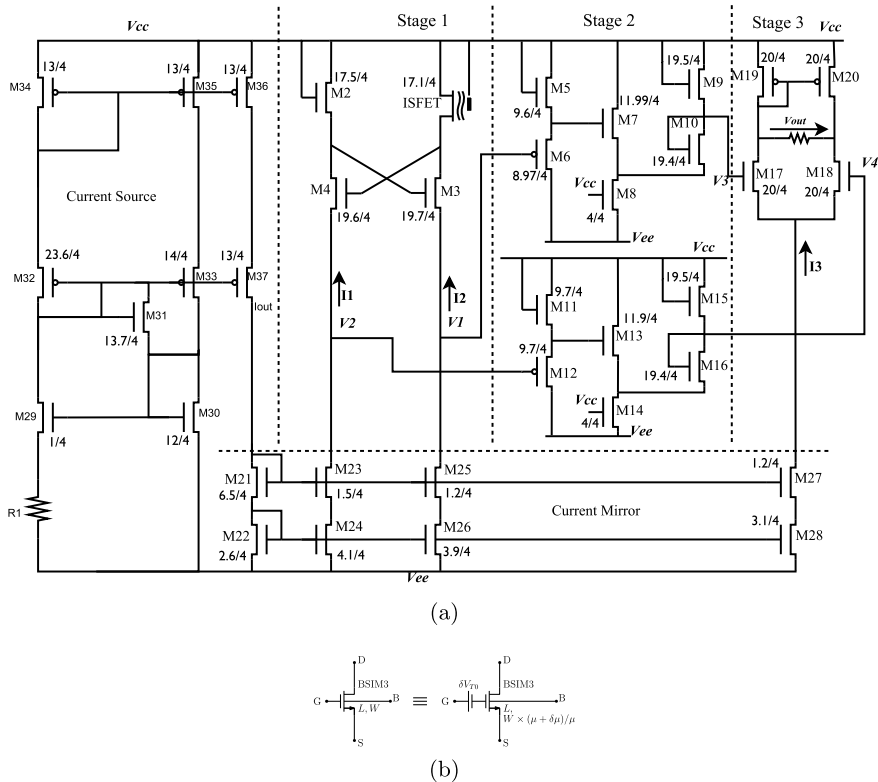
Fig. 1 **a** The proposed circuit: $R1 = 400\,k\Omega$, $R2 = 23.35\,k\Omega$ and $V_{cc} = -V_{ee} = 3\,V$ (the thermal coefficients of $R1$ are as: $TC1 = 1.510e-3$ and $TC2 = 510-7$). **b** Mismatch macro-model of MOSFET transistors

circuit's stages which contain the ISFET sensor, are designed, based on CMOS technology. In the first stage, which called Caprios quad, there are four transistors, the ISFET, M2, M3 and M4, which allow to extract a differential signal between $V1$ and $V2$, and this signal represents the threshold voltage variation of the pH-ISFET sensor. The second block is made of two attenuators, receive the two voltages $V1$ and $V2$, attenuate and shift them to be adapted to the inputs of the differential amplifier. The third stage is a differential amplifier, operates in weak inversion and allows the compensation of the temperature, of the difference of two signals $V3$ and $V4$, which are the outputs of the two attenuators.

2.1 The Mismatch Analysis

The sources of mismatch between devices are mainly limited by mathematical and experimental investigation into two, global and local variations (f_D and f_A, respectively).

$$\sigma^2(\Delta P) = f_L(W, L) + f_G(D) \approx \frac{A_P^2}{WL} + S_P^2 D^2 \tag{1}$$

where, $f_L(W, L)$ is the component depending on the transistor size. While $f_G(D)$ is associated with the dependence of the surface gradient (on the liquid) [5]. L and W are the geometry of the gate; D is the separation between the devices.

We had to think about how these changes affect the temperature sensitivity and linearity of the pH-ISFET readout circuit. Usually, the mismatch behavior modeled using two parameters, the mismatch of threshold voltage $\delta V_{T0} = V_{T01} - V_{T02}$, getting a standard deviation $\sigma_{V_{T0}}$, and the mismatch of current gain factor $\delta\beta/\beta = (\beta_1 - \beta_2)/\beta$ getting a standard deviation σ_β (with C_{ox} is the capacitance of gate, μ is the carrier mobility and $\beta = \mu C_{ox}$). Each model of the transistor in the readout circuit implemented in Spice appears in Fig. 1b. Based on the dependent variables, a simple approach is used in this work, taking into consideration the total density of sites constant ($N_s = N_{sil} + N_{nit}$). Modeling the mismatch behavior of the sensitive membrane, generally requires an assumption of Gaussian PDF for the random part of the mismatch. The simulation was done with $\overline{N_{sil}} = 4.5 \times 10^{18}\,\text{cm}^{-2}$, $\sigma(N_{sil}) = 5 \times 10^{17}\,\text{cm}^{-2}$ and $N_s = 4.6 \times 10^{18}\,\text{cm}^{-2}$. Even there is no experimental validation for the approach, so the simulation is not very precise, but in general, it gives an indication of its sensitivity to the variation of the process. The output distribution predicted by the method of Monte Carlo is performed by assigning the mobility variation $\delta\mu$ and the threshold voltage variation δV_{T0} as randomly generated:

$$\sigma(\delta V_{T0}) = \frac{A_{V_{T0}}}{\sqrt{2WL}}; \quad \sigma(\delta\mu) = \frac{A_\mu}{\sqrt{2WL}} \tag{2}$$

With, $A_\mu = 2.34 \times 10^{-4}$ cm^3/V s and $A_{V_{T0}} = 20.36 \times 10^{-9}$ V m, L and W being the size of the transistors.

3 Optimization Using Evolutionary Algorithms

3.1 The Genetic Algorithm

The change in temperature affects 37 transistors that make up the proposed read-out circuit. Not like most traditional search and optimization problems, the genetic algorithm works with a population of solutions [6]. An interval of the maximum and minimum width of the transistors defined, is chosen as an initial population. As in nature, the genetic algorithm evolves by selecting the best genes that can adapt to the existing environment, based on the principles of reproduction and mutation [7, 8]. The fixed population size was kept at 2000. The mutation rate was applied at 0.1, the crossover operator rate is 60%. At each production, 80% of promising individuals replace the population. As a general rule, we need to give the algorithm some stop criteria. Many criteria have been tested: the quality of the best result, the generation numbers and the population convergence. Generation number 500 gave us the results of this article. The simulation made in Xeon processor at @2.13 GHz in about 74 h. Three objectives define the cost minimization function:

- Minimize the quadratic error Δ_{T1} between the circuit output and the regression line of V_{out}, for a pH from 1 to 12 and at a temperature $T = 20\,^\circ$C. The error Δ_{T1} is expressed by (3), where β_1 and β_2 are the parameter estimators of the linear regression at $T = 20\,^\circ$C.
- The maximum residual error indicated in (4) must be minimized for a temperature range from 20 to 80 $^\circ$C and a pH from 1 to 12.
- Minimize the quadratic error Δ_{T2} between the circuit output V_{out} at $T = 80\,^\circ$C and the regression line of V_{out}, for a pH range from 1 to 12 and at temperature $T = 20\,^\circ$C. The error Δ_{T2} is expressed by Eq. (5).

$$\Delta_{T1} = \sqrt{\sum_{\text{pH}=1}^{\text{pH}=12} (V_{\text{out}}(\text{pH})_{T=20\,^\circ\text{C}} - (\beta_1 \times \text{pH} + \beta_2))^2} \tag{3}$$

$$\Delta_{\text{Max}} = \max[V_{\text{out}}(\text{pH}) - (\beta_1 \times \text{pH} + \beta_2)] \tag{4}$$

$$\Delta_{T2} = \sqrt{\sum_{\text{pH}=1}^{\text{pH}=12} (V_{\text{out}}(\text{pH})_{T=80\,^\circ\text{C}} - (\beta_1 \times \text{pH} + \beta_2))^2} \tag{5}$$

Let us define three objectives G_1, G_2 and G_3 for the above errors Δ_{T1}, Δ_{T2} and Δ_{Max}. A single error Δ to be reduced is as follow:

$$\Delta = \sqrt{w_1(\Delta_{T1} - G_1) + w_2(\Delta_{Max} - G_2) + w_3(\Delta_{T2} - G_3)} \qquad (6)$$

The dimensions w_1, w_2 and w_3 are width coefficients which must be adapted manually to obtain the best optimization.

3.2 Particle Swarm Algorithm

The particle swarm optimization (PSO) algorithm is a heuristic method inspired by the attitude of fish or bird populations [9]. A set of parameters is associated with a particle (in our problem we have associated the sizes of the channel of the transistor). It needs a population of random solutions to initiate the search. The particles circulate through the problematic space following the current optimal particles. The PSO is different from the GA algorithm because it does not have scalable operators. The PSO algorithm has shown its performance in various difficult optimization situations [10]. However, due to its information sharing action, all particles can rapidly converge toward a local minimum. On the other hand, the low rate of convergence is one of the drawbacks of GA. The PSO algorithm uses the recently introduced cost function.

4 Discussion and Simulation Results

The proposed circuit has been simulated using a standard BSIM3v3 provided by the MOSIS AMI $1\,\mu$m CMOS process. As a result of some simulations, the channel width of the transistors takes an approximate upper and lower limit. The overall optimization of the widths of the transistors was carried out by the PSO and the GA algorithms; the transistors lengths were fixed at $4\,\mu$m. The transistors width obtained by the simulation are presented on the circuit Fig. 1. As expected, the PSO and GA algorithms give two different optimal dimensions to the circuit. Indeed, the temperature sensitivity was approximately $2.56 \times 10^{-4}\,$pH/°C with the PSO and $2.89 \times 10^{-4}\,$pH/°C with the GA. According to the local optimization and as shown by the sensitivity curve in Fig. 2, the highest temperature drift is fewer than $2.39 \times 10^{-4}\,$pH/°C, while the precision of the circuit is approximately 0.014 pH. At a temperature of 20 °C, the determination coefficient is approximately 99.99 and 99.98% at a temperature of 80 °C. A Monte Carlo simulation was executed for a single final optimized conception. The result of about 2000 Monte Carlo analysis demonstrate good circuit behavior, even when a mismatch between devices is taken into account as described in Table 1. Finally, the proposed readout circuit achieves an efficiency of 15.7% for a temperature sensitivity of 0.001 pH/°C and a resolution of 0.05. Based on these results, in the future, the yield may be improved.
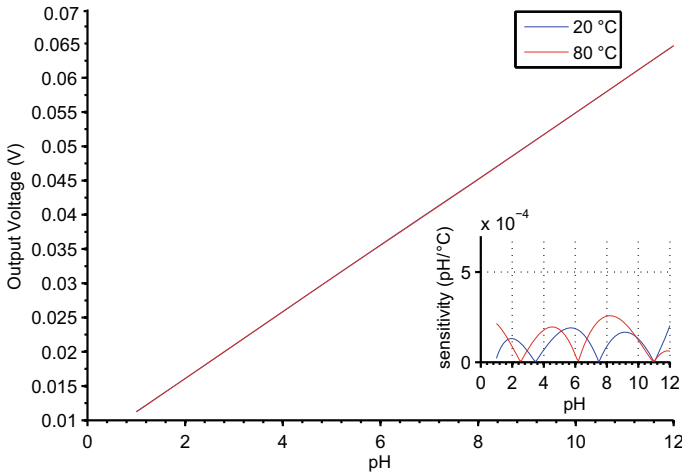
Fig. 2 The output of the proposed circuit as a function of the pH, at a temperature variation from 20 to 80 °C. In the bottom corner: the temperature sensitivity of the output pH/°C

Table 1 Efficiency of the readout circuit

Characteristics		2 point calibration (pH 2–4 or 7–10) (%)	3 point calibration (pH 4–7–10) (%)
Accuracy (pH)	Sensitivity (pH/°C)		
0.01	0.0005	1.1	1.1
0.05	0.0005	8.2	8.5
0.05	0.001	14.4	15.7
0.1	0.0005	8.5	8.5

5   Conclusions

A pH-ISFET sensor has been improved and linked with a simple and efficient circuit, compatible with standard CMOS technology. The temperature sensitivity of the sensor has been greatly improved. Through the use of site-binding theory, we modeled the temperature dependence of the sensitive membrane (SiO_2/Si_3N_4), and subsequently, we used the Verilog-a language to implement it, while the BSIM3v3 model of AMI CMOS $1\,\mu m$ supplied by MOSIS has been used for all transistors. Intensive simulation validates the robustness and the operating principle of the pH-ISFET readout circuit. In a wide range of temperature (from 20 to 80 °C) and pH (from 1 to 12), the circuit presents a good linearity, which has been proved by the coefficient of determination, was found around $R^2 = 0.9999$. Three isothermal zones were obtained at pH values (2.5, 6 and 11), with the temperature sensitivity equal to $2.4 \times 10^{-4}\,$pH/°C. The readout circuit accuracy was found to be 0.014pH. The temperature sensitivity was found to be less than 0.001 pH/°C for an unoptimized

efficiency interval between 8.5 and 15.7%, indicating that the probability of circuit manufacturing failure will be low.

References

1. Hsu WE, Chang YH, Huang YJ, Huang JC, Lin CT (2019) A pH/light dual-modal sensing ISFET assisted by artificial neural networks. ECS Trans 89(6):31
2. Gaddour A, Dghais W, Hamdi B, Ali MB, Development of new conditioning circuit for ISFET sensor
3. Firnkes M, Pedone D, Knezevic J, Doblinger M, Rant U (2010) Electrically facilitated translocations of proteins through silicon nitride nanopores: conjoint and competitive action of diffusion, electrophoresis, and electroosmosis. NANO Lett 10(6):2162–2167
4. Bhardwaj R, Majumder S, Ajmera PK, Sinha S, Sharma R, Mukhiya R, Narang P (2017) Temperature compensation of ISFET based pH sensor using artificial neural networks. In: 2017 IEEE regional symposium on micro and nanoelectronics (RSM). IEEE, pp 155–158
5. Pelgrom MJM, Duinmaiger ACJ, Welbers APG (1989) Matching properties of MOS transistors. IEEE J Solid-State Circuits SC-24, 1433–1439
6. Grefenstette JJ (2013) Genetic algorithms and their applications: proceedings of the second international conference on genetic algorithms. Psychology Press
7. Holland JH (1992) Adaptation in natural and artificial systems, vol 2. MIT Press, Cambridge, MA, 9780262082136
8. Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning, 1st edn. Addison-Wesley Longman Publishing Co., Inc., Boston, p 0201157675
9. Kennedy J, Eberhart C (1995) Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks, Australia, pp 1942–1948
10. Deng W, Yao R, Zhao H, Yang X, Li G (2019) A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm. Soft Comput 23(7):2445–2462

Modeling of Electromagnetic Field Effects on Interconnections Between High Frequency Deep Sub-micrometer CMOS Integrated Circuits Using FDTD Technique

Youssef Nadir, Khaoula Ait Belaid, Hassan Belahrach, Abdelilah Ghammaz, Aze-eddine Naamane, and Radouani Mohammed

Abstract This paper presents an improved FDTD method applied to nonuniform submicron CMOS interconnects to analyze the effect of a parasitic electromagnetic field on the signals carried at different points in the structure. As application we studied two structures of uniform and nonuniform interconnects loaded by nonlinear terminals. The proposed algorithms are implemented in the MATLAB tool. The results obtained are compared with those of PSPICE. A good agreement between the two results is achieved. The analysis of the obtained results shows that the proposed method is stable and precise. Also, it enables to simulate more complex structures.

Keywords Electromagnetic coupling · Nonuniform interconnections · CMOS driver · FDTD · MATLAB · PSPICE

1 Introduction

Nowadays, integrated electronics are becoming increasingly important, allowing circuits to be easily inserted into many applications such as telecommunications, transports and domotics. This progress has improved fabrication techniques by producing more efficient components embedded in complex systems [1]. The high integration density of circuits contributes to the increasing problems of electromagnetic compatibility. On the other hand, the increase in the number of functions inside the same chip leads to an increase in parasitic emissions in the form of radiation that can reach the different neighboring circuits. The resolution of electromagnetic interfer-

Y. Nadir (✉) · A. Naamane · R. Mohammed
L2MC Laboratory, National School of Arts and Crafts, Meknes, Morocco
e-mail: josse.nadir@gmail.com

Y. Nadir · H. Belahrach · A. Naamane
Electrical Laboratory, Royal School of Aeronautics, Marrakech, Morocco

K. A. Belaid · H. Belahrach · A. Ghammaz
Laboratory LSEEET, University Cadi Ayyad, Marrakech, Morocco

ence problems is difficult for systems integrating chips of different types. Let's take the example of a transmission-reception system, its front end stage corresponds to the analog circuits located between the antenna and the signal processing stage [2]. With the recent development, manufacturers have been able to integrate the transmission, reception, modulation, demodulation and signal processing stages in the same device. This large scale integration means that engineers have to find solutions in the design to avoid time consuming and costly redesign phases [3, 4].

The objective of this paper is to develop and propose mathematical models that represent the electromagnetic field effects, associated with signal and power integrity, in CMOS integrated circuits operating at high frequencies. Indeed, this work is divided as follows: Sect. 2 is dedicated to the state of the art of the existing and electrical interconnections problems in the presence of a parasitic electromagnetic field. The third section discusses examples of electromagnetic coupling between uniform and nonuniform interconnections and with linear and nonlinear loads. Finally, Sect. 4 concludes the paper.

2 Signal Integrity Analytical Study in Deep Sub-micrometer Interconnect Lines

With the development of embedded systems, the modules (analog, digital, microwave, power …) are more integrated. Therefore, the problems of signal integrity, electromagnetic compatibility (EMC) and reliability in different domain (microelectronics, automotive, aeronautics …) are increasing [5, 6]. Signal integrity is described as the management behavior of logic signals so as not to disturb the electronic board's functionality. This will require the study of the influence of passive elements (interconnects, packages and wires) on electronic systems. Electromagnetic emission is generated by electromagnetic disturbances, emitted intentionally or not by the source. This emission can be conducted through power and signal lines, or radiated. In the near field, the mutual coupling mechanism corresponds to inductive and capacitive coupling. The victim is characterized by its electromagnetic susceptibility, which means that it is sensitive to incident electromagnetic interferences. The robustness of the victim to electromagnetic disturbances is quantified by the concept of electromagnetic immunity [7, 8].

2.1 Electromagnetic Coupling Equations of Nonuniform Interconnect Lines

For nonuniform coupled interconnections, to study and analyze the phenomena mentioned above, the quasi transvers electromagnetic model is assumed [9].

Fig. 1 Cross-sectional view of multiple interconnects illuminated by an incident field $\vec{E}$
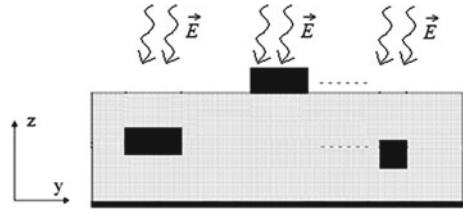
Consider M imperfect interconnect lines immersed in an incident field, as shown in Fig. 1, modeled by the following system of coupled partial differential equations (Telegrapher's equations) [10, 11]:

$$\begin{cases} \frac{\partial \mathbf{V}(x,t)}{\partial x} + \mathbf{R}(x)\mathbf{I}(x,t) + \mathbf{L}(x)\frac{\partial \mathbf{I}(x,t)}{\partial t} = \mathbf{V_s}(x,t) \\ \frac{\partial \mathbf{I}(x,t)}{\partial x} + \mathbf{G}(x)\mathbf{I}(x,t) + \mathbf{C}(x)\frac{\partial \mathbf{V}(x,t)}{\partial t} = \mathbf{I_s}(x,t) \end{cases} \tag{1}$$

where the voltages $\mathbf{V}$ and currents $\mathbf{I}$ are expressed in $M \times 1$ column vector form, and interconnect parameters $\mathbf{R}$, $\mathbf{L}$, $\mathbf{C}$ and $\mathbf{G}$ are expressed in $M \times M$ matrices per unit length distributed voltages and currents, respectively, and are written as follows [11]:

$$V_s(x,t) = -\frac{\partial E_T(x,t)}{\partial x} + E_L(x,t) \tag{2}$$

$$I_s(x,t) = -G(x)E_T(x,t) - C(x)\frac{\partial E_T(x,t)}{\partial x} \tag{3}$$

where $\mathbf{V_s}$ and $\mathbf{I_s}$ are expressed in $M \times 1$ column vector form, $\mathbf{E_T}$ and $\mathbf{E_L}$ are the transverse and longitudinal components of the incident electric field, respectively.

The cross-sectional view of modeled structure is shown in Fig. 1.

In this work we will present the finite difference method in the time domain (FDTD) to solve the equations system (1). This method is based on the discretization of the spatial domain into Nx cells each of length Δx, and the time domain into Nt intervals each of duration Δt. This discretization uses an interlaced leap-frog scheme such that the currents are computed at $(t + \Delta t/2)$ steps and $(x + \Delta x)$ positions, whereas the voltages are computed at $(t + \Delta t)$ steps and $(x + \Delta x/2)$ positions.

The recurrence relations of voltage and current along the line interconnects, are given by:

For $k = 2, 3, \ldots Nx$

$$\begin{aligned} V_{k+1/2}^n = A_1^{-1} \Big(& A_2 V_{k+1/2}^{n-1} - I_k^{n-1/2} + I_{k-1}^{n-1/2} \\ & + \Delta x \left(G_k E_k^n S + \frac{1}{\Delta t} C_k (E_k^{n+1} - E_k^n) A_k \right) \Big) \end{aligned} \tag{4}$$

For $k = 1, 2, 3, \ldots Nx$

$$I_k^{n+1/2} = A_3^{-1}\left(A_4 I_k^{n-1/2} - V_{k+1/2}^n + V_{k-1/2}^n + \frac{\Delta x}{\Delta t}D_k\left(E_k^{n+1} - E_k^n\right)\right) \quad (5)$$

With A_1, A_2, A_3 and A_4 are square matrices $M \times M$ dependent on the line parameters R, L, C and G. The $M \times 1$ column vectors A_k and D_k depend on the position of the interconnections, the characteristics of the incident field and the modal velocities propagated in the structure. So, $A_1 = \frac{\Delta x}{\Delta t}\mathbf{C} + \frac{\Delta x}{2}\mathbf{G}$; $A_2 = \frac{\Delta x}{\Delta t}\mathbf{C} - \frac{\Delta x}{2}\mathbf{G}$; $A_3 = \frac{\Delta x}{\Delta t}\mathbf{L} + \frac{\Delta x}{2}\mathbf{R}$; $A_4 = \frac{\Delta x}{\Delta t}\mathbf{L} - \frac{\Delta x}{2}\mathbf{R}$, and with E_k^n is the incident field at position k and time n.

To determine the voltages and currents at the ends of the structure investigated, the nature of the load circuits must be taken into account. In the following paragraph, nonlinear loads based on CMOS inverters are described in detail.

2.2 Coupled Interconnect Lines with Nonlinear Loads

We consider the structure of Fig. 2. The interconnections are driven by CMOS inverters and are loaded by parallel RC circuits.

In fact, the CMOS inverter can be modeled by several models. The most widely used model in submicron CMOS technology is the alpha power law model. The drain current of the PMOS and NMOS transistors can be given by:

$$I_P = \begin{cases} 0, & V_g \geq V_{DD} - |V_{TP}| \\ k_l p(V_{DD} - V_g - |V_{TP}|)^{\alpha p/2}(V_{DD} - V_1), & V_1 > V_{DD} - V_{SATP} \\ k_s p(V_{DD} - V_g - |V_{TP}|)^{\alpha p}, & V_1 \leq V_{DD} - V_{SATP} \end{cases} \quad (6a)$$
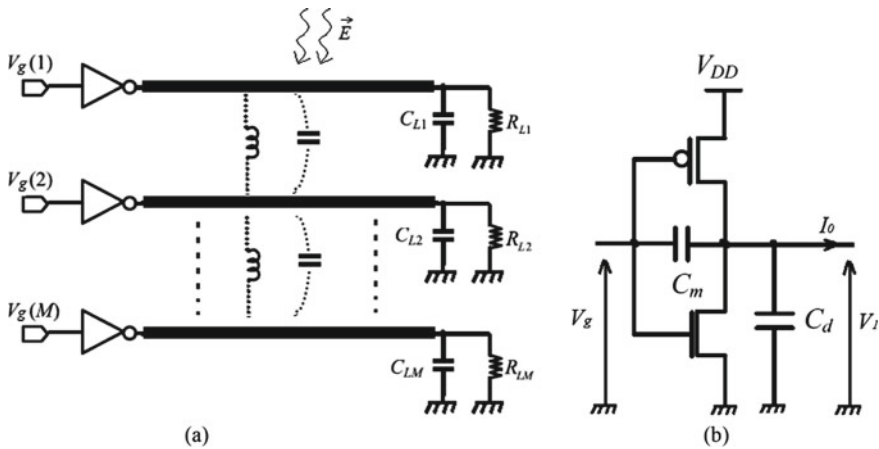
Fig. 2 **a** M-interconnects coupled with drivers and loads. **b** CMOS inverter model

$$I_N = \begin{cases} 0, & V_g \leq V_{TN} \\ k_{ln}(V_g - V_{TN})^{\frac{\alpha n}{2}} V_1, & V_1 < V_{SATN} \\ k_{sn}(V_g - V_{TN})^{\alpha n}, & V_1 \geq V_{SATN} \end{cases} \tag{6b}$$

where V_{DD} and V_g are the supply voltage and excitation source voltage, respectively. k_l and k_s are the parameters in linear and saturation states, respectively, α is the velocity saturation index and V_T is the threshold voltage.

The input voltage of the interconnections can be computed by the following recurrence relation:

$$V_1^{n+1} = Q_1 V_1^n + Q_2 \left(I_0^n - 2I_1^{n+1/2} + I_P^{n+1} - I_N^{n+1} + \frac{1}{\Delta t} C_m \left(V_g^{n+1} - V_g^n \right) \right)$$
$$+ P W_1^n \tag{7}$$

where
$Q_1 = P A_1^{-1} (A_2 + (C_m + C_d) / \Delta t)$, $Q_2 = P A_1^{-1}$, $P = \left[U + A_1^{-1} (C_m + C_d) / \Delta t \right]^{-1}$ and $W_1^n = \Delta x A_1^{-1} \left(G(1) E_1^n + \frac{1}{\Delta t} C(1) \left(E_1^{n+1} - E_1^n \right) \right) A(1)$, with E_1^n is the incident field at position $k = 1$ and time n.

The output voltage is given by the following recurrence relationship:

$$V_{Nx+1}^{n+1} = K_1^{-1} \left(K_2 V_{Nx+1}^n + I_{Nx}^n + K_3 E_{Nx}^n + K_4 E_{out}^{n-1} + K_5 E_{Nx}^{n+1} \right) \tag{8}$$

Which matrices K_1, K_2, K_3, K_4, and K_5 have the following formulas: $K_1 = \frac{\Delta x}{2\Delta t} C + \frac{\Delta x}{2} G + \frac{1}{\Delta t} C_L + G_L$; $K_2 = \frac{\Delta x}{2\Delta t} C + \frac{1}{\Delta t} C_L$; $K_3 = \frac{\Delta x}{2} GA + \frac{\Delta x}{2\Delta t} CA - \frac{\Delta x^2}{2\Delta t^2} CD + \frac{\Delta x^2}{4\Delta t} GD$; $K_4 = -\frac{\Delta x}{2\Delta t} CA + \frac{\Delta x^2}{4\Delta t^2} CD - \frac{\Delta x^2}{4\Delta t} GD$; $K_5 = \frac{\Delta x^2}{4\Delta t} GD$. With E_{Nx}^n is the incident field at position $k = N_x$ and time n. The relationships of the voltage and current at the input, output and along the line interconnects are implemented in MATLAB. Some examples will be studied in the following section.

3 Examples and Discussion

3.1 Uniform Interconnects Loaded with Nonlinear Circuits and Excited by Incident Field

The example in Fig. 3 considers three coupled interconnects with nonlinear terminations. The three lines are excited by a trapezoidal incident field of amplitude $1\,\text{kV/m}$ and rise and fall time $T_r = T_f = 0.1\,\text{ns}$ and pulse width $T_w = 5\,\text{ns}$. In addition, the 2nd interconnect is excited by a trapezoidal voltage source of amplitude 5V with rise and fall time $T_r' = T_f' = 0.1\,\text{ns}$ and pulse width $T_w' = 25\,\text{ns}$.

Fig. 3 Example of a three interconnect structure test circuit with nonlinear loads
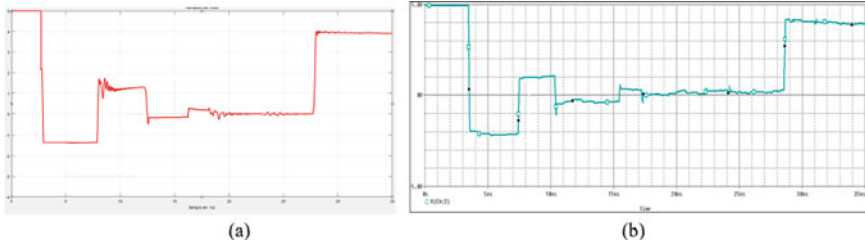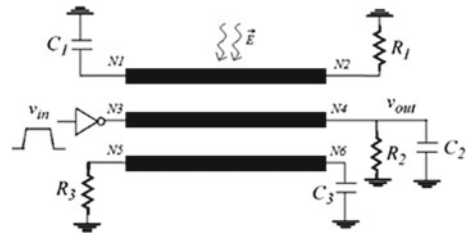
Fig. 4 Results at the output voltage of the second line, **a** MATLAB **b** PSPICE

After obtaining the input and output voltage recurrence relations, we implemented them on MATLAB and obtained the following result at the output of the second line (Fig. 4a). We also simulated the structure with the PSPICE tool (Fig. 4b).

As a conclusion, we found a good agreement between the two results obtained by the MATLAB and PSPICE tools. This proves the accuracy of our model.

3.2 Nonuniform Interconnects Loaded with Nonlinear Circuits and Excited by Incident Field

With the increasing flow of data in telecommunication systems and the increasing demand for faster data processing, there has been a renewed interest in modeling various non-ideal effects of the conductors that interconnect the different subsystems. The example discussed here is frequently found in reality. This example illustrates two nonuniform interconnect lines, the first is excited by a trapezoidal voltage source of delay $T = 1$ ns and the whole is excited by an external field of delay $T' = 3$ ns as shown in Fig. 5.

For this example, we performed simulations with the MATLAB tool. The induced voltage at the input of the 2nd victim line is shown in Fig. 6a. Also, the structure was simulated with the PSPICE tool (Fig. 6a).

Comparing both results obtained by MATLAB and PSPICE tools, we observe good agreement between them. Therefore, the proposed models of coupled and nonuniform interconnects are verified.
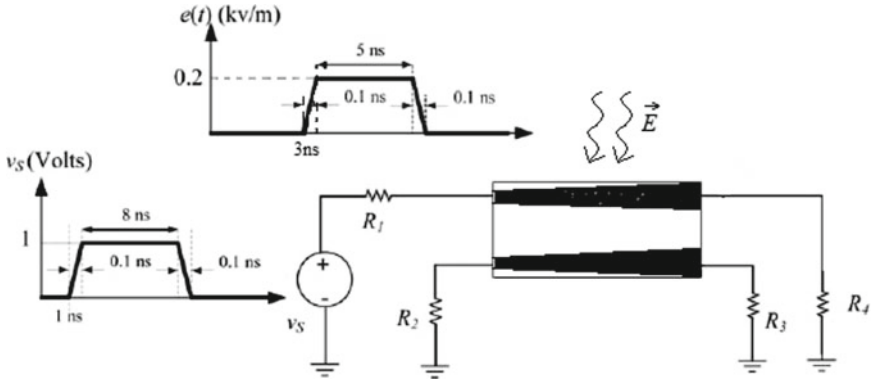
Fig. 5 Schematic of two nonuniform interconnect lines in the presence of an external field $\vec{E}$
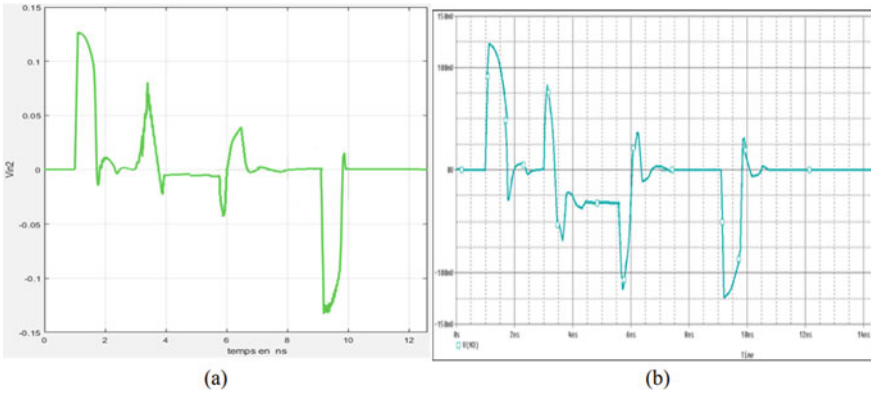
Fig. 6 Input voltage of the 2nd interconnect nouniform victim line, **a** MATLAB **b** PSPICE

4   Conclusion

Based on the FDTD method, we have modeled and simulated the electromagnetic coupling between nonuniform interconnect lines loaded by nonlinear circuits, using the telegrapher equations. The signal propagation mode is assumed to be quasi-TEM. Thus we have analyzed the validity of our algorithm implemented on MATLAB using PSPICE in the case of the presence of external conducted and radiated disturbances. The obtained results show a good agreement between the simulations performed with MATLAB and PSPICE. The proposed algorithm has been validated numerically with different examples.

References

1. Banerjee K (2001) 3-D ICs: a novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration. Proc IEEE 89(15):602–633
2. Nicolas D (2017) Conception de circuits RF en CMOS SOI pour modules d'antenne reconfigurables. Ph.D. micro et nanotechnologies/microélectronique. Université Paul Sabatier, Toulouse 3
3. Gade SH et al (2018) Millimeter wave wireless interconnects in deep submicron chips: challenges and opportunities. Integration VLSI J. https://doi.org/10.1016/j.vlsi.2018.09.004
4. Davis JA, Venkatesan R, Kaloyeros A et al (2001) Interconnect limits on gigascale integration (GSI) in the 21st century. Proc IEEE 89(3):305–324
5. Lee M, Pak JS, Kim J (2014) Electrical design of through silicon via. Springer, Dordrecht
6. Afrooz K (2012) Efficient method for time-domain analysis of lossy nonuniform multiconductor transmission line driven by a modulated signal using FDTD technique. IEEE Trans Electromagn Compat 54(2):482–494
7. Kumar VR et al (2014) An accurate model for dynamic crosstalk analysis of CMOS gate driven on-chip interconnects using FDTD method. Microelectron J 45:441–448
8. Kalpana AB, Hunagund PV (2012) Effect of driver strength on crosstalk in global interconnects. Int J Adv Comput Sci Appl 3(9):76–79
9. Wang W, Liu P-G, Qin Y-J (2013) An unconditional stable 1D-FDTD method for modeling transmission lines based on precise slipt-step scheme. Progr Electromagn Res 135, 245–260
10. Xu Q (2007) An efficient modeling of transmission lines with electromagnetic wave coupling by using the finite quadrature method. Trans Very Large Integr Syst 15:1289–1302
11. Rachidi F, Tkachenko S (2008) Electromagnetic field interaction with transmission lines from classical theory to HF radiation effects. In: Advances in electrical engineering and electromagnetics, p 29

Design and Experimentation of an Automotive Diagnostic Tool for Headlamp ECU Based on the UDS

Meryam El Mahri, Tarik Jarou, Ihssane Sefrioui, Sofia El Idrissi, and Jawad Abdouni

Abstract An Electronic Control Unit (ECU) diagnosis provides information about the behavior of the ECU, which in turn provides the overall health of the vehicle. In this paper, an effective Controller Area Network (CAN) diagnostic fault tool based on the Unified Diagnostic Services (UDS) protocol is presented. The diagnostic system, as the name implies, diagnoses the failures of the ECUs so that the maintenance personnel can accurately fix them. As the vehicle's network has become more complex, CAN diagnosis is now implemented as the most important feature of communication. The simulated diagnostic tool uses the UDS protocol to query the ECUs that are connected to the vehicle network, more precisely the headlamp ECU. The diagnostic data collected will then be saved in files.

Keywords The CAN protocol · The UDS protocol · Diagnostic tool · ECU

1 Introduction

The increasing application of in-vehicle electronic components makes it necessary to use diagnostic systems to monitor and control parameters. Development, industry and aftermarket are all areas that use diagnostic systems to perform their tasks. A diagnostic system must therefore contain a protocol to connect the diagnostic tools that designers, testers, and repairers use to verify diagnostic information. Each protocol can be tailored to a single diagnostic system and vehicle components. These systems require a great deal of effort to implement the protocol for a particular diagnostic system.

The application of communication protocols such as FlexRay, CAN, and LIN allows devices in the vehicle network to communicate with each other according to

M. El Mahri (✉) · T. Jarou · I. Sefrioui · S. El Idrissi · J. Abdouni
Advanced Systems Engineering Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco
e-mail: meryam.elmahri@gmail.com

J. Abdouni
e-mail: j.abdouni@enim.ac.ma

standards. Diagnostics identifies, verifies, and classifies symptoms to get a complete picture of the root cause of a problem in a car. The detection, improvement, and communication strategies applied to the abnormal operation of systems are controlled by electrical and electronic devices. Therefore, the goal of diagnostics is to identify the root cause of the abnormal operation so that a repair can be made. In the development cycle of an automotive system, diagnostic functions can be used in the development, manufacturing and service stages.

The purpose of this diagnostic tool is to detect faults that have occurred within the functionality (ECU). The UDS protocol has become the de facto standard in automotive diagnostics. It is standardized as ISO 15765-3 [1]. The diagnostic services available in the UDS are grouped into functional units and identified by a one-byte code (Service ID) [2]. Each service defines a request message, a positive response message and a negative response message. Since UDS uses messages of variable length, a transport protocol is required on layers with well-defined (short) message lengths, such as CAN. The transport protocol divides a long UDS message into chunks that can be transferred over the network and reassembles these chunks to recover the original message. One of the main diagnostic functions is the reading of diagnostic trouble codes (DTCs) [3].

1.1 Implementation of the UDS on the CAN Bus

The UDS protocol on CAN as per ISO 14229-1 and ISO 15765-3 is used in this paper. It sends the data packet to the UDS application layer based on the data received at the application layer. The corresponding response is sent over the CAN network [4]. The international standard ISO 15765-3 contains specifications of how, for example, a data frame, which cannot fit into a single CAN-frame, should be handled by using segmented messages [5]. The *Single Frame*, the *First Frame*, the *Consecutive Frame* and the *Flow Control Frame* all contain a data domain of 8 bytes. The data domain consists of information and control data for the transmission protocol. The structure of all kinds of CAN messages is shown in Table 1 [6].

Table 1 Structure of UDS frames on the CAN bus

Frame type	Byte 1	Byte 2	Byte 3	Byte 4–8
Single frame	SF_DL	SID	Data 1–6	
First frame	FF_DL		SID	Data 1–5
Consecutive frame	SN	Data 1–7		
Flow control frame	FS	BS	ST_{min}	–
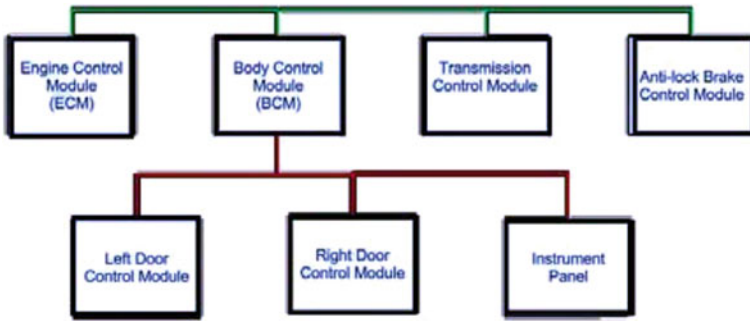
Fig. 1 Simplified CAN network in the automotive industry

The components of the different types of frames are:

- **SF_DL**: indicates the length of the *Single Frame* on 4 bits.
- **FF_DL**: indicates the length of the *First Frame* on 12 bits.
- **SN**: indicates the sequence number.
- **FS**: indicates the state of the data flow.
- **BS**: indicates the size of the block.
- **STmin**: is the minimum time interval.

1.2 Diagnostics with CAN

The **CAN** is an asynchronous serial network protocol using a bus topology. It is the most popular communication method in vehicles in recent years [7] (Fig. 1).

The CAN bus is a system for the current communication (real time) of control units. It is not used for the direct connection of sensors or actuators, but only to link controllers. In the case of the exchange of information via the CAN data bus, all data is transmitted only through two wires. Both bidirectional conductors will carry the same data, regardless of the number of control units and the amount of transmitted data. Transmission of information through the CAN bus has meaning only when we need to send a lot of information between multiple controllers [8].

2 Design of the Diagnostic Tool

2.1 Hardware Requirements

Our tool can be broadly divided into three parts represented in Fig. 2:

- Microcontroller board.
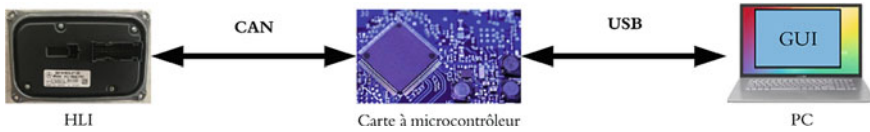
Fig. 2 General architecture of the tool

- PC allows the programming of the code, allowing the sending of the frames as well as the acquisition of the answers to these sent frames and the design of a graphic interface.
- HLI module, the *device under diagnostic*.

Essentially, the main board is the one that receives the input signals from the outputs via the CAN interface. These signals are transmitted via the USB port to the PC. An Human-Machine Interface (HMI) resides on the PC that processes the signals received from the main board or requests the main board to send various CAN messages back to the HLI.

2.2 Soft Design

More and more automotive electronic control modules are starting to use CAN instead of the traditional K diagnosis (ISO 14230) because of its higher speed and reliability. In order to unify the various on-board network diagnosis services, ISO established a general diagnosis communication protocol, ISO 14229, also called UDS. On the one hand, our tool, in the case of ECU failure, freezes the frame data and provides real time warning through the negative response. The tool can provide an accurate diagnosis in real time, as well as guarantee diagnostic information for after-sales diagnostic analysis. Our system achieves the goal of optimizing the repair time by facilitating the detection of faulty boards. Figure 3 summarizes the diagnostic steps.

As discussed before, the UDS protocol over CAN offers two types of transmission:

1. **Single frame**: for non-segmented messages that can fit into a single CAN frame. The message contains 8 bytes or less.
2. **Multiframe**: contains the first frame and one or more consecutive frames.

Figure 4 introduces the algorithms that explains those two sorts of transmission.

3 Experimentation and Results

The unit test can only guarantee that each function has the correct outputs. But it cannot guarantee the correct interaction between the different functions. So it is necessary to do a functional test of all the components. This test can verify the
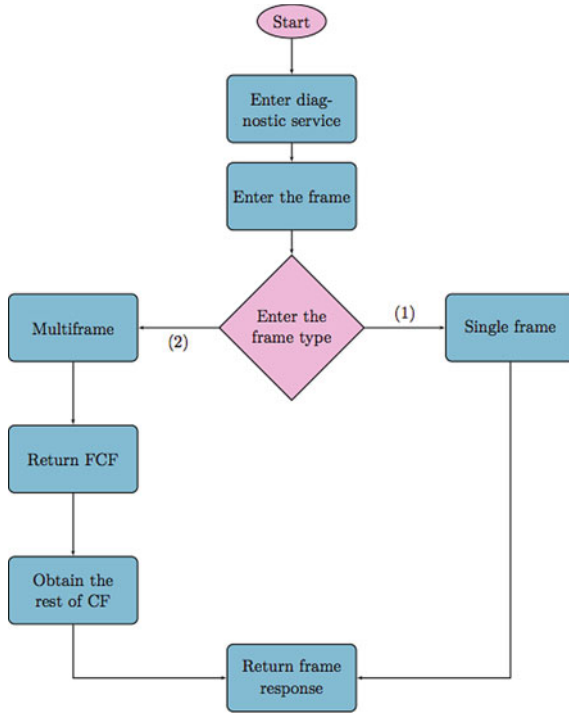
Fig. 3 Algorithm of the diagnosis of the system
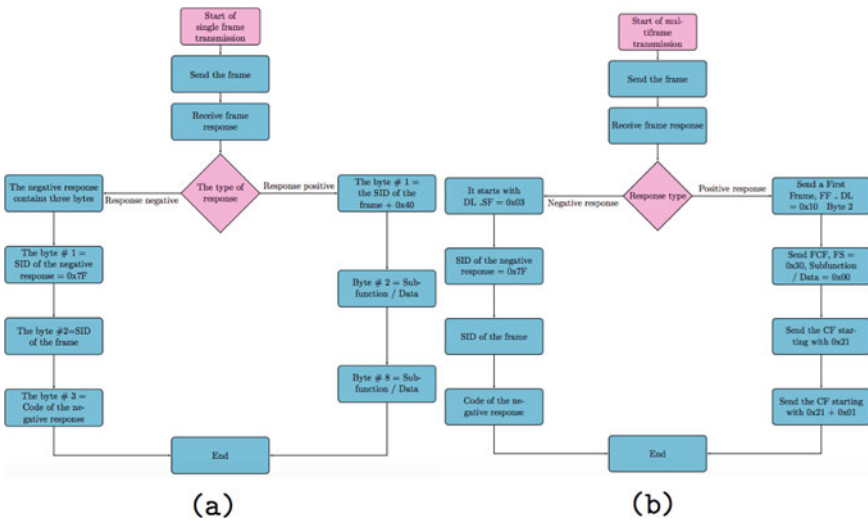
Fig. 4 Algorithms for the transmission of single frame (**a**) and multiframe (**b**)
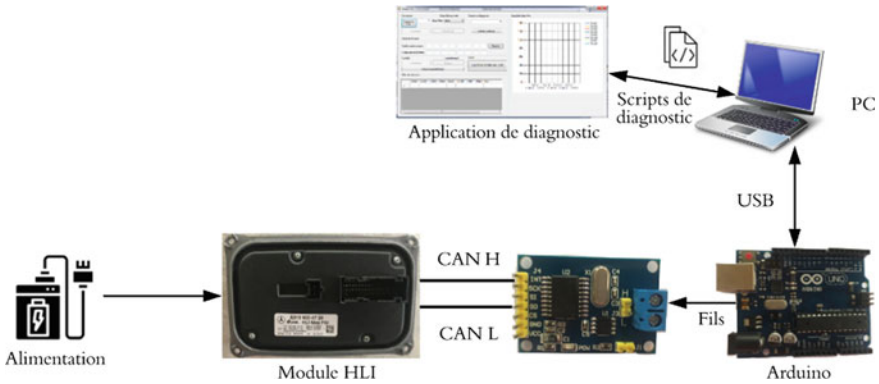
Fig. 5 Diagram of the diagnostic tool of the HLI module

operation of the system as defined. In order to test and verify the functionality of the developed tool, the system has been deployed and tested.

Figure 5 shows the real-world environment of this diagnosis tool.

As described in the diagram above, our system includes:

- Computer that houses the application that allows first to establish communication with the set consisting of the MCP2515 CAN Bus and the Arduino.
- Arduino assembly and the MCP2515 CAN Bus.
- Headlamp Interface (HLI) module considered as a device under diagnosis.
- Cables and wires.
- Module power supply which is fixed at 13 V.

Several aspects of the system have been tested, but due to confidentiality, only a few of these tests will be discussed in this section.

3.1 Read Serial Number

In order to read the *SerialNumber* of the ECU, it is necessary to pass by several services.

3.1.1 Activate L'*extendedDiagnosticSession*

The ECU gives positive answer so the *extendedDiagnosticSession* is activated (Fig. 6).
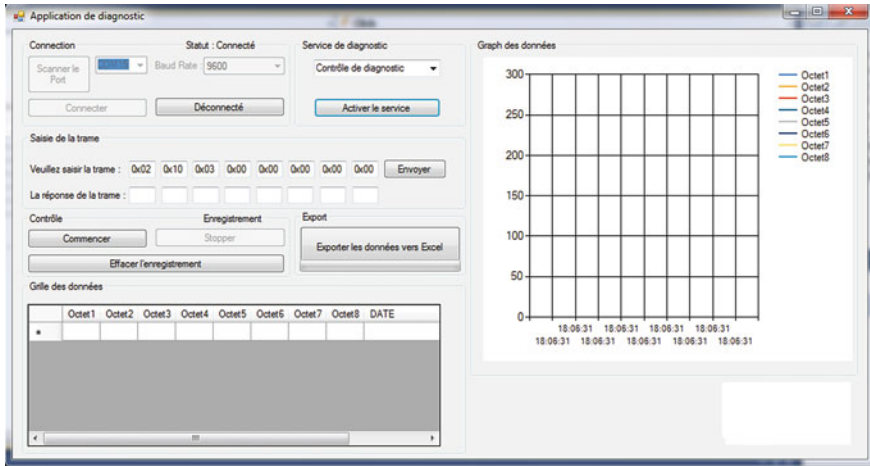
Fig. 6 Activation of the *extendedDiagnosticSession*

Table 2 Frames allowing the reading of the serial number

Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6	Byte 7	Byte 8
0x10	0x0F	0x62	0xFD	0x10	0x3F	0x04	0x00
0x30	0x00	0x00	0x00	0x00	0x00	0x00	0x00
0x21	0x30	0x05	0x4F	0x82	0x04	0x01	0x10
0x22	0x40	0x01	0xAA	0xAA	0xAA	0xAA	0xAA

3.1.2 Access to the Service *ReadDataByIdentifier*

We are now in the *extendedDiagnosticSession*, therefore we have the right to read the static information of the ECU, for example, the *Serial Number*. We must first access the *ReadDataByIdentifier* service.

To activate the service of *ReadDataByIdentifier*, we send the following frame:

0x03 0x22 0xFD 0x10 0x00 0x00 0x00 0x00.

Once we send this frame, we receive the following frames (Table 2).

The *First Frame* beginning with 0x10, is the *First Frame*.The second byte of the *First Frame* defines the size of the data that one wishes to read. As the second byte is equal to 15, our data will be represented on 15 bytes beginning from the next byte until the end of the *First Frame*. The second received frame is the *Control Flow Frame*. It begins with 0x30 and the remainder of this frame is put back to 0x00. The nine bytes which constitute the remainder of the *Serial Number* are introduced by the two frames begun by 0x21 and 0x22. The remainder of the last one is filled by 0xAA (Fig. 7).
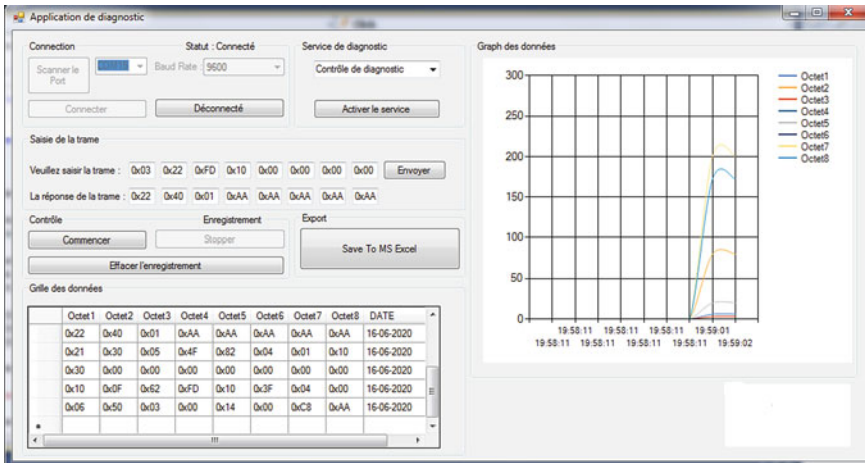
Fig. 7 Reading of the serial number of the HLI module

When we finish the diagnosis of our module, we can export the received frames to Excel. The Excel file generated will have the name of the day we performed the diagnosis.

4 Conclusion and Perspectives

The details of the hardware and software construction of the tool, which interfaces and communicates directly with the CAN bus and performs diagnostics based on the UDS protocol, were presented. The frames extracted from the ECU can be exported to an Excel file for saving. The purpose of this tool is to verify and validate the functionality of the control systems.

As an outlook, future work should enhance the ability to perform an enhancement of this diagnostic tool to perform HLI module testing and real time monitoring of test data in the testing process to further improve the automation capabilities of the system. It can also be enhanced by adding the diagnostic option for more ECUs connected to the virtual vehicle network by implementing more UDS diagnostic services, dedicated to each ECU.

References

1. Dekanic S, Grbic R, Maruna T, Kolak I (2018) Integration of CAN bus drivers and UDS on Aurix platform. In: Zooming innovation in consumer technologies conference (ZINC). IEEE
2. Slcianu M, Foalu C (2012) A new CAN diagnostic fault simulator based on UDS protocol. In: International conference and exposition on electrical and power engineering. IEEE

3. Wajape M, Elamana NB (2014) Study of ISO 14229-1 and ISO 15765-3 and implementation in EMS ECU for EEPROM for UDS application. In: International conference on vehicular electronics and safety (ICVES). IEEE
4. Nayak SR, Bagubali A (2019) Study on diagnosis of automotive network. In: International conference on vision towards emerging trends in communication and networking (ViTECoN). IEEE
5. Pendrill R (2016) Automation of UDS-based flashing for software testing purposes in CANoe
6. Wang J, Zhou Y, Li Q (2014) Research on fault diagnostic system in CVT based on UDS. Hindawi Publishing Corporation Advances in Mechanical Engineering
7. Prodanov W, Valle M, Buzas R (2009) A controller area network bus transceiver behavioral model for network design and simulation. IEEE Trans Ind Electron
8. Kubia M, Beno P (2019) Realization of communication via the CAN bus. In: The 13th international scientific conference on sustainable, modern and safe transport. IEEE

Design and Experimentation of an Automatic Communication Tool for Automotive Cards

Ihssane Sefrioui, Tarik Jarou, Meryam El Mahri, Sofia El Idrissi, and Jawad Abdouni

Abstract The present study has allowed the realization of an automatic communication tool for electronic modules such as BCM and smart junction boxes. We propose a platform with a graphic interface developed by the C# language to achieve communication via the Controller Area Network (CAN) and Unified Diagnostic Services (UDS) based on vector devices. This tool is designed to connect or disconnect to the electronic automotive part and to allow sending periodic frames to visualize the responses of the card in real time.

Keywords Controller area network (CAN) · UDS protocol · Automotive card

1 Introduction

In passenger cars, the automation industry uses CAN for vehicle engine management as the in-vehicle engine management network (IVN). This application includes managing the engine, the body electronics such as the door, the roof control, the AC, and the car lightning. The European and the American car manufacturers are using CAN-based IVNs [1].

The industrialization of the Body Control Module goes through several steps and tests; among these tests, we have the functional test, which allows testing the good functioning of these BCM before integrating them into the vehicle. To realize this test, it is necessary first of all to succeed in the communication with our BCM, to be able to diagnose it and to carry out the tests of several functionalities of the module.

This paper presents a realization of a communication tool that allows communication with the BCM to diagnose and test it. We started by describing our proposal tool, both hardware and software parts. After that, we detailed the design part and we finished by representing the experimentation results of our communication tool.

I. Sefrioui (✉) · T. Jarou · M. El Mahri · S. El Idrissi · J. Abdouni
Advanced Systems Engineering Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco
e-mail: ihssane.sefrioui@gmail.com

The communication protocol CAN, ISO 11898, describes how information is passed between devices on a network and conforms to the Open Systems Interconnection (OSI) model defined in terms of layers [2]. The actual communication between devices connected by the physical medium is defined by the model's physical layer [3]. The ISO 11898 architecture defines the lowest two layers of the seven-layer OSI/ISO model as the data-link layer and physical layer [4, 5].

The UDS protocol is the most universal and suitable for the widest possible range of projects. In particular, it allows the ECU reprogramming [6] and is designed for offline diagnosis of vehicle malfunctions in a service station. For our tool, we choose to implement the UDS protocol.

The protocol UDS is a communication protocol specified in the ISO15764-2 and ISO14229 for ECU diagnostics [7]. Diagnostics of the ECU based on UDS are performed using standardized communication over the CAN bus. Basically, the client sends a predefined request, and the ECU responds with a predefined answer. The request and answer consist of a series of bytes (byte string) [8].

2   Description of the Proposed Tool

The proposed communication tool represents a connection of hardware devices and software applications to manage messages exchange. The hardware part of the communication tool comprises the following devices:

- CANCaseXL is a device from the company "Vector". Using the CAN protocol allows: the sending and receiving of data frames, detection and generation of error frames, and time synchronization.
- The role of the fictive load is to transform the frames that arrive from the CANCase into signals to communicate with the automotive card.
- The power supply is used to supply electricity to other electrical devices. It is fixed on 12.5 V.
- We aim to realize a generic application that can communicate with several automotive cards of brands such as BCM and smart junction box (Fig. 1).

Our system should initialize the driver, set up channels, and transmit and receive messages to succeed in the communication. Figure 2 describes the global software architecture of our communication system.

Figure 3 describes the steps followed to connect the automotive card and the application using CANCase.

Fig. 1 Hardware part of the experimentation
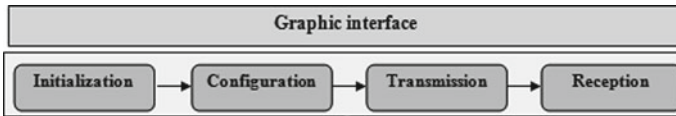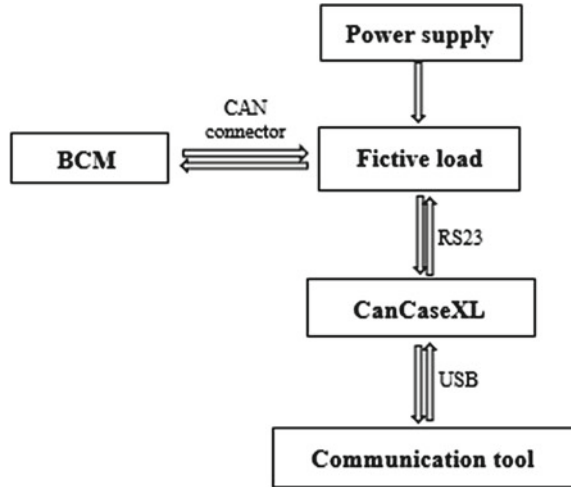
Fig. 2 Software part of the experimentation

3 Design of the Communication Tool

The logic part consists of developing a well-structured algorithm broken down into interfaces that inherit from the classes; these last ones contain methods of initialization, communication, and deactivation of the drivers….

Concerning the driver, we will work either with the CANCase tester driver or with the ESD tester driver.

Initialization: During the development phase, the first part of generating a communication module is to initialize the driver with which we will launch the communication.

We started by developing a function responsible for the initialization phase. Figure 4 shows the different objectives of the function.

Transmission of a message: A message is a frame of data. Since we have developed a communication algorithm via the CAN protocol, this message will consist of several fields.

Our transmission system will be managed by a function called "Send_Msg" which includes the following tasks as Fig. 5 presents.

Reception of a message: The reception message is also a frame of data to allow receiving a frame of the electronic automotive part; we developed a function that allows recovering all the data sent by the part.

Fig. 3 Communication
diagram by the CAN
protocol
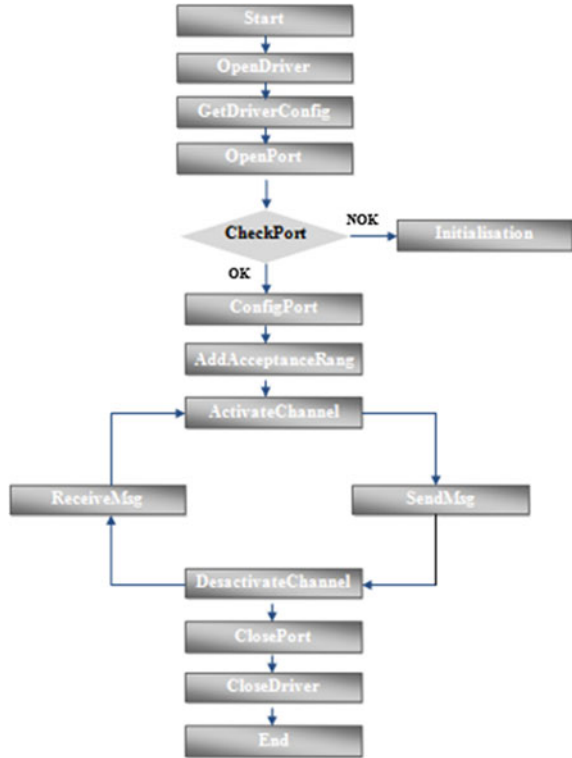
Fig. 4 Initialization of the
communication tool
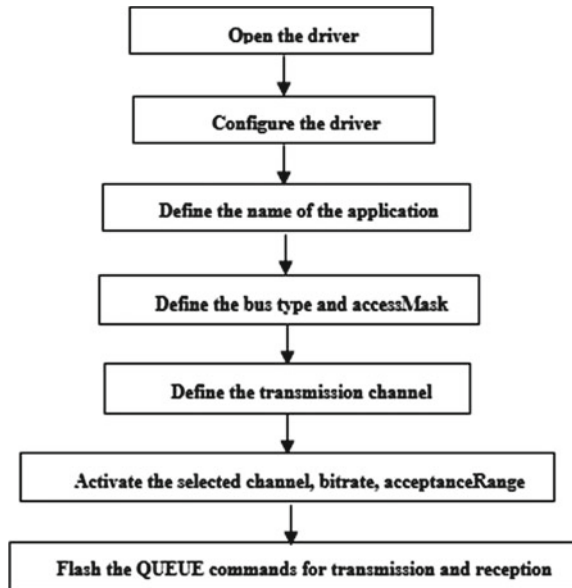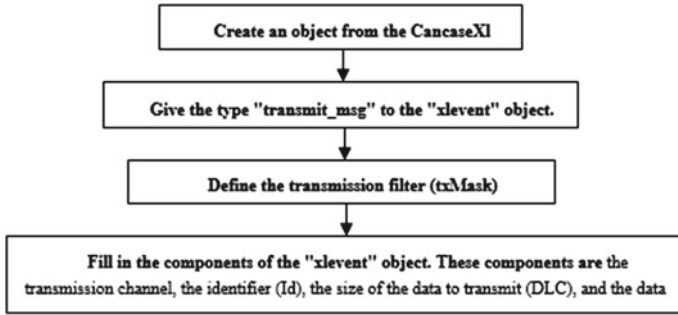
Fig. 5 Transmission of a message on the communication tool

We have developed several functions that allow retrieving the message from the transmission according to the need. The first two functions developed are "ReceiveMsg()" and "ReceiveFiltredMsg".

The difference between them is the result of the return and the input parameters. One contains an identifier filter of the frames of the reception, and the other one contains no input parameters.

Generally, the two functions perform the following tasks described in Fig. 6.

Our tool should, first of all, communicate with the CANCase hardware. It needs a software interface that acts as an intermediary between the system and the external device; it is the CANCaseXL driver, also called "Vector Hardware Configuration". Using this driver, we can configure our application and the transmission channels.
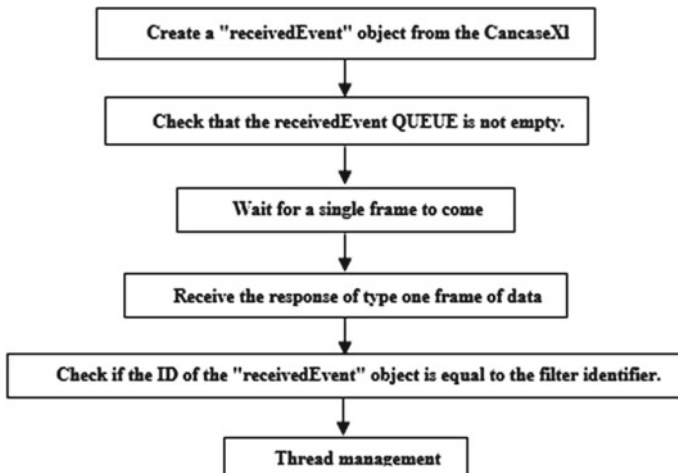
Fig. 6 Reception of a message on the communication tool

4 Experimentation and Results

4.1 Experimentation Procedure

To test the reliability of the application, we have to go through several steps. This part of the test procedure will describe the steps for connecting and powering the hardware to make the application ready for communication. Then, a list of test modules will be listed in the verification phase to demonstrate the software connection scenario. Below are the tasks that must be followed to launch the communication tool:

- Power the automotive electronic part by linking these connectors with the ground and the power source.
- Connect the CANCaseXl equipment with the CAN bus.
- Connect the two wires of the CAN bus to the connectors of the C1A automotive card.
- Connect our communication tool with the CANCaseXl equipment via its USB port.

4.2 Experimentation Result and Interpretation

During the realization of our application, we have dealt with two main parts (hardware and software), each of which generates its part of errors, so we have to perform a set of tests and verifications to find the right solution.

This part will present a list of test modules, test scenarios, and results obtained from Table 1.

Configuration interface: This interface connects or disconnects the application to the electronic automotive part. It allows displaying the type of the initialization result, either OK or NOK (Fig. 7).

Communication interface: This interface allows sending and receiving frames. It ensures communication with the automotive card. It manages the threads that allow avoiding the blocking of the application at the time of the test to visualize the results in real time. In addition, it allows sending periodic frames to prevent the automotive card from entering the sleep mode (Fig. 8).

Table 1 Testing and verification

Test module	Test scenarios	Status
Application and drivers	Configure the driver	OK
Data management	Extract and display the message on the graphical interface	OK
The connectors	Test the connectors	OK
CAN connection	Activate and choose the appropriate channel	OK

Fig. 7 "Configuration" interface
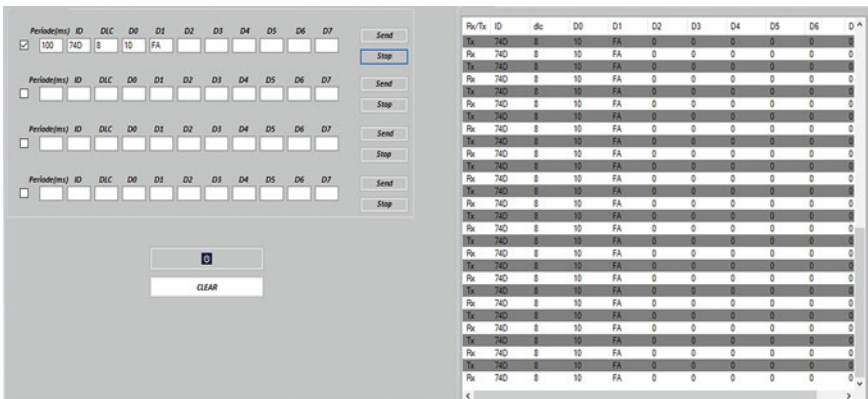
Fig. 8 "Communication" interface

5　Conclusion

Successfully communicating with the automotive card and allowing messages to be sent and received is the initial step in diagnosing and testing the automotive card when it is out of adjustment.

One of the main steps in the production of automotive cards is the functional testing of these cards. It is carried out by specific machines for each type of card which is too expensive. This test is performed by executing and simulating functions such as window opening, rain sensor, central locking, windshield wiper…. For most factories, it is difficult to perform this test for start-up projects. Therefore, a test and diagnostic application suitable for any card will be cost-effective. The tool realized in this paper represents the initial phase of communication that can be completed to achieve a successful functional test and diagnostic tool for automotive cards.

References

1. Othman HF, Aji YR, Fakhreddin FT, Al-Ali AR (2006) Controller area networks: evolution and applications. In: IEEE 2006 2nd international conference on information and communication technologies—Damascus, Syria, 24–28 Apr 2006
2. Chen H, Tian J (2009) Research on the controller area network. In: IEEE 2009 international conference on networking and digital society (ICNDS)—Guiyang, Guizhou, China, 30 May 2009–31 May 2009
3. Kubisa M, Beno P (2019) Realization of communication via the CAN bus. In: The 13th international scientific conference on sustainable, modern and safe transport. IEEE
4. Jang D, Han S, Kang S, Choi, J-W (2015) Communication channel modeling of controller area network (CAN). In: IEEE 2015 seventh international conference on ubiquitous and future networks (ICUFN)—Sapporo, Japan, 7–10 July 2015
5. Dekanic S, Grbic R, Maruna T, Kolak I (2018) Integration of CAN bus drivers and UDS on Aurix platform. In: IEEE 2018 zooming innovation in consumer technologies conference (ZINC)—Novi Sad, Serbia, 30–31 May 2018
6. Slcianu M, Foalu C (2012) A new CAN diagnostic fault simulator based on UDS protocol. In: International conference and exposition on electrical and power engineering. IEEE
7. Wajape M, Elamana NB (2014) Study of ISO 14229-1 and ISO 15765-3 and implementation in EMS ECU for EEPROM for UDS application. In: International conference on vehicular electronics and safety (ICVES). IEEE
8. Xing W, Chen H, Ding H (1999) The application of controller area network on vehicle. In: IEEE international vehicle electronics conference (IVEC'99)—Changchun, China, 6–9 Sept 1999

Portable Ultrasound Sensors System for Breast Cancer Early Diagnosis

G. Zaz, M. Zekriti, and L. Fakri-Bouchet

Abstract This paper gives a description a new portable system for the detection and prediction of breast cancer prognosis. This system is composed of a set of ultrasonic transducers mounted on a flexible support. The transducers measure the Speed of Sound (SoS) and attenuation from the density and stiffness of the tissues. Due to this property, the proposed system can accurately detect the nature of the breast tumor. As known, the speed of sound in the tissues is temperature-dependent; thus, to ensure the reliability and the accuracy of the measurement, the proposed system will also measure the local temperature of the tissue based on the spectral components of the same ultrasonic transducers. This device uses artificial intelligence in cancer diagnosis, which reduces relatively the margin of error.

Keywords Breast cancer · Ultrasonic transducers · Artificial intelligence (AI)

1 Introduction

Breast cancer is one of the major health problems in the world. In Morocco, it is the most prevalent cancer in women. It is the first cancer among women both in terms of incidence and in terms of mortality and the third of all registered cancer cases.

G. Zaz (✉)
SIGER Laboratory, Faculty of Science and Technology (FST) of Fez, University of Sidi Mohamed Ben Abdellah (USMBA), 30000 Fez, Morocco
e-mail: ghita.zaz@usmba.ac.ma

M. Zekriti · L. Fakri-Bouchet
Euromed Research Center, Euromed University of Fez, UEMF Campus, 30070 Fez, Morocco
e-mail: m.zekriti@ueuromed.org

L. Fakri-Bouchet
e-mail: latifa.fakri-bouchet@insa-lyon.fr

L. Fakri-Bouchet
Institute of Applied Sciences of Lyon (ISA), University Lyon 1, 69622 Villeurbanne, France

INSA of Lyon, Lyon, France

According to the 2020 edition of Oncohighlights [1], in Morocco, over 48% of new cancer cases were diagnosed in 2019. Sixty-five percentage of cancers affect women versus 35% for men.

Breast cancer is the most common form of cancer in women in Morocco 35%, followed by cervical cancer (11.2%), thyroid cancer (8.6%), colorectal cancer (5.9%), and ovarian cancer (4.3%). In 2019, 10,414 new cases of breast cancer were identified in the kingdom.

In 2030, this figure will rise to 16,018, according to forecasts by the Greater Casablanca Registry (RCGC). In 2004, this figure was 5465. Women aged 45–49 are the most affected, and it is the leading cause of their death. Breast cancer is the leading cause of cancer mortality among women in Morocco, causing nearly 3700 deaths per year [2].

The increase in mortality is also due to limited access to medications; lack of access to diagnosis and lack of wider use of multi-modality treatment from preventive examinations and medical imaging such as mammography (MMG), ultrasound tomography (UT) and computed one (UCT), or magnetic resonance imaging (MRI), etc. These most common breast cancer detection modalities are generally limited by radiation exposure, discomfort, high costs, inter-observer variabilities in image interpretation, and low sensitivity in detecting cancer in dense breast tissue. All of these imaging systems simply provide images of pathologies in the human body. They do not characterize any features.

Moreover, in rural areas of Morocco where access to mammography is limited, a portable breast ultrasound screening will be able to offer the benefits of tomography at low cost, without ionizing radiation, without contrast agents and with minimal impact on the clinical workflow.

2 Background

2.1 Techniques of Breast Cancer Detection

There are several techniques for detecting breast cancer. The first is X-ray mammography, which provides images of the entire breast volume after compression of the breast [3]. This examination has been considered the most effective way to detect and diagnose breast cancer. However, the number of cancers that escape detection by mammography is significant, especially in young women with dense breasts. In addition, mammography may lead to unnecessary biopsies due to its low specificity [4]. In this case, biopsies not only increase costs, but also put patients under emotional pressure. Magnetic resonance imaging (MRI) has also been recommended by the American Cancer Society (ACS) for breast cancer screening in women at very high risk. Although very effective, MRI is expensive and can be risky because of the contrast media required.

As a complementary screening modality to X-ray mammography and MRI, the ultrasound tomography can be used. This technique allows the volume of the breast to be inspected with an acoustic probe manipulated by an operator in order to establish a more precise diagnosis of the nature of the objects detected. The use of this technique is constantly being improved. Studies have also shown that the use of ultrasound images increases the detection rate of any type of cancer by 17% compared to conventional techniques [5]. Unlike mammography, ultrasound imaging is essential for the examination of young women under 35 years of age (glandular and therefore dense breasts) [6]. It also has the advantage of being inexpensive (compared to MRI), non-restrictive (no compression of the breast) and non-irradiating (no ionizing radiation). However, ultrasound imaging is limited by various factors: partial views, difficulty in reproducing the results accurately by the same (operator-dependent nature of the examination) and in conferring a quantitative character to the images.

In addition to the most commonly used diagnostic breast imaging examinations (MMG, MRI and US), there are other imaging modalities under development, which will play an increasingly important role not only in diagnosis but also for therapy [7].

To the same end, the authors of this paper propose a new non-invasive, painless, and three-dimensional imaging system to improve breast cancer detection. Using ultrasonic probes, this system allows accurate and real-time 3D imaging of the entire breast interior, independent of the operator. By extracting quantitative parameters from the breast tissue, this system can distinguish a malignant and benign one.

2.2   Ultrasonic Solutions for Breast Cancer Detection

The most common form of ultrasonic transducers is a piezoelectric crystal. These elements can be used either as a transmitter or as a receiver. These are available with operating frequencies ranging from 20 kHz to 2 GHz.

In reflection mode, an ultrasonic pulse is transmitted through ultrasonic transducer. The pulse generates acoustic pressure waves that propagate through the medium and interacts with changes in medium. The reflected signals are returned to the transducer where the received echoes can be back projected to create an image of the medium [8]. In transmission mode, the transmitter generates a continuous output, the echo of which is detected by a separate receiver, on opposing sides of the media of interest [9].

The transmission of acoustic signal can be used to extract certain quantitative characteristics of breast tissue that help in diagnostic evaluation. For this purpose, new practices have emerged, for example: sono-elasticity and elastography techniques extract elastic properties of tissues from Young's modulus, shear modulus, or shear wave velocity by measuring the velocity displacement of tissues subjected to a low-frequency vibration [10]. A disadvantage of these techniques is that the diagnosis is strongly related to the mechanical properties of the tissue [11]. An

alternative solution is currently recommended. This technique is based on the propagation of longitudinal waves in the tissue to measure the Speed of Sound (SoS) [12]. The SoS of the acoustic wave between the transmitter and receiver is obtained by counting the time elapsed between the beginning of the transmission and the end of the reception, defined as the time-of-flight (TOF) [13]. In addition to this technique, changes in ultrasonic wave amplitude can indicate tissue attenuation [14]. It has been shown that the SoS and attenuation measurements combined to the reflected ultrasonic image have great potential to distinguish benign and malignant tumors with high accuracy. The ultrasonic image provides some criteria of breast nodule, such as surrounding tissue, shape, margin contour, and lesion boundary [15]. The SoS and attenuation measurements provide meaningful acoustic characteristics to estimate breast tissue density and stiffness which can help the clinicians to detect abnormalities and anticipate future development of breast cancer. Indeed, tumors are stiffer than their surroundings, especially when cancerous: SoS is about 3% faster through malignant tumors than through surrounding healthy tissue and 1.5% faster through benign tumors [14].

Several research projects are therefore being conducted to develop and improve this technique. For example, the ultrasound computer tomography (USCT) system reconstructs the SoS and attenuation maps from transmission measurements based on a large number of transducers located around the inspected tissue and immersed in a tank of degassed water. The ultrasound waves are transmitted along several angular directions to obtain a 3D image of the observed tissue [16, 17]. Apart from this technique, there are few ultrasonic systems that are currently undergoing clinical trials or news preclinical research. These include Quantitative Transmission Ultrasound Tomography (QT) Ultrasound's QT Scanner 2000 [18], Delphinus Medical Technologies' SoftVue System [19], and Mastoscopia's Multimodal Ultrasound Tomography (MUT) Mark II System [16, 20]. The objective of all these systems is the same: the complete detection of breast cancer. However, they operate with different center frequencies, with different numbers and positions of the transducer elements and different ways to construct 3D ultrasonic image.

2.3 Artificial Intelligence for Breast Cancer Detection

Artificial intelligence (AI) is at the heart of the medicine of the future with numerous applications such as telemedicine, intelligent prostheses, personalized treatments, and diagnostic assistance. For the last ten years, AI has experienced a considerable growth thanks to the emergence of new deep learning algorithms, made possible by the massive amount of data and the computing power available through GPU computing. This approach is based on neural networks (NN), a large amount of data and proceeds by machine learning. The objective is to search for regularities in the data in order to extract knowledge. Very recently, remarkable improvements in breast cancer risk prevention have been achieved [21] using a mammography-based deep learning algorithm [22]. Note that the relevance of these approaches in breast cancer

risk prevention is highlighted in [23]. Considering the performances of deep learning algorithms using mammography images, it seems to us very relevant to develop a new approach based on ultrasound imaging using the new medical devices presented above.

3 Novel System for Breast Cancer Early Diagnosis

3.1 The Interest of the Ultrasonic System

Common disadvantages of previous systems include limited penetration distance of the ultrasonic wave and inflexible scanning methods. Thus, these solutions are primarily applicable for standard breast shapes and cannot be adapted to other parts of the human body which are not easily submersible. In addition, the temperature variation aspect is not taken into account, which reduces the accuracy of the measurement. All of these disadvantages, however, can be mitigated and even overcome.

This solution, presented in this paper, is composed of a set of very thin ultrasound transducers deposited on a flexible substrate. The flexible aspect allows the device to adapt to different breast shapes. Thanks to the SoS and the attenuation maps, combined with a reflected 3D image, the device detects the presence of a nodule and extracts its features (malignant or benign).

Thanks to the multi-physics aspect of the device, the local temperature of the tissue is also measured in order to improve the reliability of the diagnosis.

This innovative system allows collecting, communicate remotely and interpret the available information in real time. Thanks to an intelligent algorithm based on neural network, the device can help in the clinical decision-making.

3.2 Transduction Elements Design

The proposed system uses an array of transmitted and reflected ultrasonic transducers fixed on flexible support. Conventional ultrasonic transducers are usually made of several layers of material: an active element (piezoelectric disk), a backing element, and a delay line, all acoustically coupled together. In our case, the transducers possess the same multilayer structures composed of piezoelectric elements welded on one surface to a golden electrode. An aluminum layer is deposited on the second surface as an electric input and output electrodes. Then, a thin layer of glue welds the whole system to a delay line, which is a cylindrical silica rod (6 mm in diameter and 2 mm in height). Finally, an epoxy resin acts as a backing medium [24] (Fig. 1).

The transducers have a diameter of 1 cm and a thickness up to 4 mm. The transmitting transducer emits a broadband pulse with a center frequency of up to 2 MHz (f_1). At the same time, the reflected elements operate with a center frequency of

Aluminum Electrode — Piezoelectric element

Gold & chrome — Silica
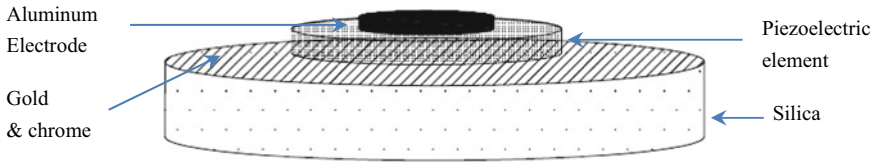
Fig. 1 Structure of the proposed transducer

Table 1 Thickness of the transducer layers

Parameters	Thickness
Piezoelectric (LiNbO$_3$)	$d_1 = 1.5$ mm for $f_1 = 2$ MHz; $d_2 = 750$ μm for $f_2 = 4$ MHz
Silica (SiO$_2$)	$d_3 = 2$ mm
Aluminum	20 μm
Gold and chrome	Some μm

up to 4 MHz (f_2). The center frequency is an important component of ultrasonic systems as a higher central frequency improves the resolution of the reflection mode. The central frequency of the transducer changes according to the thickness of the piezoelectric element (Table 1).

3.3 Operating Principle

The ultrasonic system uses transmitted transducers to generate the SoS and attenuation maps, while the reflected transducers generate the reflection image. To perform these measurements, transducers are excited with a pulse of greater bandwidth to increase the SoS image accuracy (Figs. 2 and 3). Indeed, bandwidth is essential in ultrasonic systems as a larger bandwidth will result in a shorter transmission pulse (time domain), which is more easily recognized for travel time estimation in SoS imaging [16]. After excitation, transducers emit a planar wave one by one. When only one element is exited the others record to generate reflection, SoS, and attenuation images and provide an image of the breast volume.

Before manufacturing the transducers, the development of accurate numerical tools is important to model the behavior of the acoustic wave in the multilayer structure. Nevertheless, interesting results can be obtained by modeling the samples as infinite plates subjected to incident plane waves. Using the transfer matrix related to each layer considerably simplifies the modeling [25].

Figure 4 shows series of echoes; the first echo corresponds to the piezoelectric excitation. Then, for each reflection of the acoustic echo at the interface (delay line/tissue), the wave undergoes a reduction of its amplitude and a phase shift (of $\pi/2$). The time interval between echoes corresponds to the delay line thickness. The second series of echo is a reflection of the first series of echoes on the surface of

Fig. 2 Normalized pulse
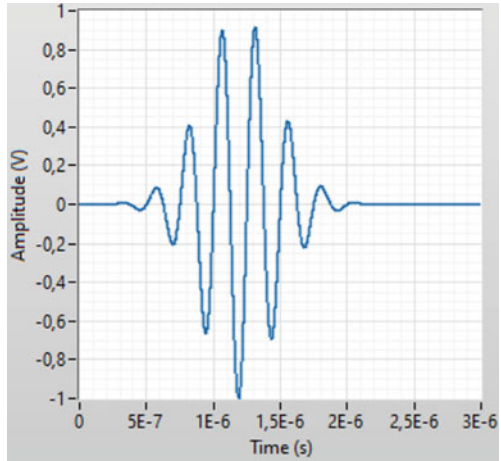
Fig. 3 Frequency spectrum of the pulse
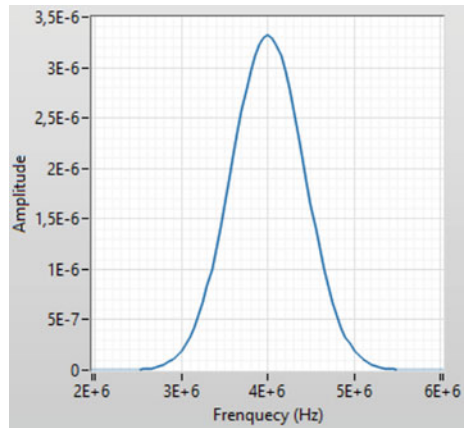
the nodule. A primary cause of the first echo attenuation is mainly related to the contact (gold/chrome/glue layers) between the piezoelectric element and the delay line. From the time interval between different series of echoes, it can be possible to locate the position of the nodule and measure the velocity and attenuation within the breast nodule.

The temperature dependence of the SoS is an important parameter. To ensure the reliability of the measurement, the device measures the local tissue temperature based on the spectral components of the same ultrasound transducers [26]. The diagnostic protocol thus takes into account the global parameters influencing the diagnosis. This aspect of the device will be discussed in detail in a future paper.

To help in the decision, an AI algorithm for early diagnosis of breast is being developed. More precisely, we propose a nodule classification system based on an artificial neural network. The artificial neural network can classify nodules based
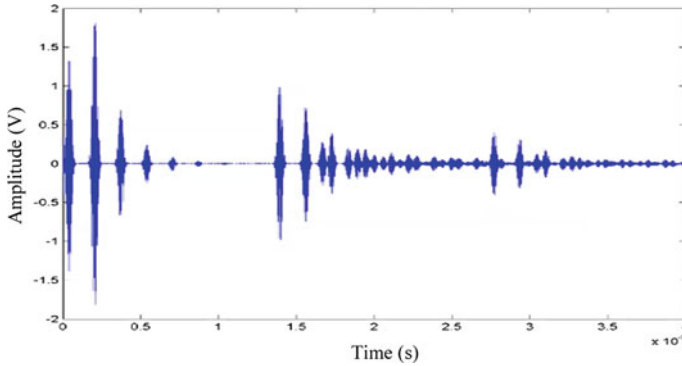
Fig. 4 Propagation of the ultrasonic waves in the transducer/tissue mediums

on three morphological features representing the shape, edge characteristics, and darkness of a nodule, as well as the speed of sound in the nodule and its temperature. The constructed model will be tested on a set of simulated data collected from the ultrasonic device and from a database of images of histologically confirmed cases, containing benign and malignant nodules. This database will be provided by the radiology department of the Centre Hospitalier Universitaire Hassan II (CHU) of Fez, Morocco.

4   Discussion and Conclusion

This research work aims to contribute to the efforts led more globally and more specifically in Morocco with the help of the Lala Salma Foundation (RCGC: Registre des Cancers du Grand Casablanca) to find a solution for the early diagnosis of breast cancer accessible to all patients regardless of geography or socio-economic status.

This paper focuses on the development and miniaturization of a flexible ultrasound detection system. The proposed system is an array of ultrasound transducers (transmission and reflection) fixed on a flexible support. This system is non-invasive, painless, non-ionizing, and inexpensive. The transmission transducer emits a broadband pulse with a center frequency up to 2 MHz. At the same time, the reflected elements operate with a center frequency of up to 4 MHz. Thanks to the multiphysics aspect of the device; the local temperature of the tissue is also measured to improve the reliability of the diagnosis.

The developed system will use all strategies offered by AI tools: artificial neural networks: CNN used for US images and RNN used for signal processing.

This system represents a promising challenge and needs more research to improve its performance. In the future, it will play an important role in the early detection of breast cancer.

References

1. https://oncohighlights.com/. Last access 25 Feb 2022
2. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2018) Global cancer observatory: cancer today. International Agency for Research on Cancer, Lyon, France
3. Cheng HD, Shi XJ, Min R, Hu LM, Cai XP, Du HN (2006) Pattern Recognit 39(4):646–668
4. Jesneck JL, Lo JY, Baker JA (2007) Radiology 244(2):390–398 (2007)
5. Chang R-F, Wu W-J, Moon WK, Chen D-R (2003) Ultrasound Med Biol 29(5):679–686
6. Maev RG (ed) (2002) Acoustical imaging, vol 26. Springer US, Boston
7. Hill CR (1995) Br J Radiol 8
8. Wells PN (2006) Ultrasound imaging. Phys Med Biol 51:R83
9. Marioli D, Narduzzi C, Offelli C, Petri D, Sardini E, Taroni A (1992) Digital time-of-flight measurement for ultrasonic sensors. IEEE Trans Instrum Meas 41(1):93–97
10. Goksel O, Eskandari H, Salcudean SE (2013) IEEE Trans Med Imaging 32(2):408–418
11. Cosgrove D et al (2013) Ultraschall Med Eur J Ultrasound 34(03):238–253
12. Glozmanet T, Azhari H (2010) J Ultrasound Med 29(3):387–398
13. Dai H, Zhao S, Jia Z, Chen T (2013) Low-cost ultrasonic distance sensor arrays with networked error correction. Sensors 13(9):11818–11841
14. Li C, Duric N, Huang L (2008) Breast imaging using transmission ultrasound: reconstructing tissue parameters of sound speed and attenuation. In: Proceedings of the 2008 international conference on biomedical engineering and informatics. Sanya, China, 27–30 May 2008, pp 708–712
15. Guo R, Lu G, Qin B, Fei B (2018) Ultrasound imaging technologies for breast cancer detection and management: a review. Ultrasound Med Biol 44(1):37–70. https://doi.org/10.1016/j.ultrasmedbio.2017.09.012
16. Kratkiewicz K, Pattyn A, Alijabbari N, Mehrmohammadi M (2022) Ultrasound and photoacoustic imaging of breast cancer: clinical systems, challenges, and future outlook. J Clin Med 11(5):1165. https://doi.org/10.3390/jcm11051165
17. Gemmeke H, Hopp T, Zapf M, Kaiser C, Ruiter NV (2017) 3D ultrasound computer tomography: hardware setup, reconstruction methods and first clinical results. Nucl Instrum Methods Phys Res Sect A Accel Spectrom Detect Assoc Equip 873:59–65
18. Malik B, Terry R, Wiskin J, Lenox M (2018) Quantitative transmission ultrasound tomography: imaging and performance characteristics. Med Phys 45:3063–3075
19. Sandhu G, Li C, Roy O, Schmidt S, Duric N (2015) Frequency domain ultrasound waveform tomography: breast imaging using a ring transducer. Phys Med Biol 60:5381
20. Forte S, Dellas S, Stieltjes B, Bongartz B (2017) Multimodal ultrasound tomography for breast imaging: a prospective study of clinical feasibility. Eur Radiol Exp 1:27
21. Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ (2019) Clin Radiol 74:5
22. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R (2019) Radiology 292(1):60–66
23. Sitek A, Wolfe J (2019) Radiology 292(1)
24. Zaz G, Le Clézio E, Dekious A, Chrifi Alaoui M, Calzavara Y, Despaux G (2018) In-situ high-resolution measurement of RHF nuclear fuel plates' spacing. IEEE Trans Nucl Sci 65(11):2776–2783. https://doi.org/10.1109/TNS.2018.2875310

25. Folds DL, Loggins CD (1977) Transmission and reflection of ultrasonic waves in layered media. J Acoust Soc Am 62(5):1102–1109
26. Zaz G, Calzavara Y, Le Clézio E, Despaux G (2015) Adaptation of a high frequency ultrasonic transducer to the measurement of water temperature in a nuclear reactor. Phys Proc 70:195–198. https://doi.org/10.1016/j.phpro.2015.08.116

Cough Detection for Prevention Against the COVID-19 Pandemic

Btissam Bouzammour, Ghita Zaz, Malika Alami Marktani, Ali Ahaitouf, and Mohammed Jorio

Abstract This paper proposes a novel and robust technique for remote cough recognition for COVID-19 detection. This technique is based on sound and image analysis. The objective is to create a real-time system combining artificial intelligence (AI) algorithms, embedded systems, and network of sensors to detect COVID-19-specific cough and identify the person who coughed. Remote acquisition and analysis of sounds and images allow the system to perform both detection and classification of the detected cough using AI algorithms and image processing to identify the coughing person. This will give the ability to distinguish between a normal person and a person carrying the COVID-19 virus.

Keywords COVID-19 · AI algorithms · Cough detection · Embedded systems · Sensor fusion · Real-time processing

---

B. Bouzammour (✉) · G. Zaz · A. Ahaitouf · M. Jorio
SIGER Laboratory, Faculty of Sciences and Technology, Sidi Mohamed Ben Abdellah University, Fez, Morocco
e-mail: btissam.bouzammour@usmba.ac.ma

G. Zaz
e-mail: ghita.zaz@usmba.ac.ma

A. Ahaitouf
e-mail: ali.ahaitouf@usmba.ac.ma

M. Jorio
e-mail: mohammed.jorio@usmba.ac.ma

M. Alami Marktani
SIGER Laboratory, National School of Applied Sciences, Sidi Mohamed Ben Abdellah University, Fez, Morocco
e-mail: malika.alamimarktani@usmba.ac.ma

1   Introduction

The world has suffered greatly from the consequences of the coronavirus on our lives. The rapid spread of this pandemic continues to disrupt the balance of the world, making any attempt to limit the spread of the virus one of the most important priorities to be taken seriously.

Many research groups are trying to find a solution to this pandemic by using AI to recognize the cases of COVID-19 quickly. The main goal is to distinguish this virus from other similar pathologies by listening to a person's voice when coughing.

Imran et al. [1] conducted a preliminary study to detect COVID-19-related coughs collected with smartphone applications, where a combination of deep patterns was trained from 48 patients who tested positive. Chloe et al. [2] used Web-based applications to download the population's coughing sounds along with their demographic data and medical history to develop a machine learning algorithm based on voice, breath, and cough sounds. Gökcen et al. [3] develop an AI-based mobile application to COVID-19 by real-time cough measurement. A public data set was used, features (MFCC features, status, gender, respiratory condition, fever muscle pain, and status) were selected, and they applied deep learning algorithm for classification. The model provided an accuracy of 79%. Erdoğan et al. [4] propose a study to develop a system able to detect COVID-19(+) patients from the acoustic data of cough. Data has been selected from free access site. The feature extraction has been done by a traditional approach using the empirical mode decomposition and the discrete wavelet transform, and the feature selection was applied with the ReliefF algorithm. An accuracy of 97.8% was obtained. Tena et al. [5] have developed a model based on the automatic diagnosis of COVID-19 from automatic extraction of cough characteristics. Autoencoder was implemented for extraction features, and supervised machine learning algorithm was applied. The model provided an accuracy close to 90%.

This paper presents a novel technique to detect the presence of COVID-19 cough through smart technologies using visual and sound methods to be able to detect and identify any person carrying this virus. The rest of the paper is organized as follows. In Sect. 2, coughing detection audio estimation is described. Section 3 presents the coughing detection pose estimation technique. Section 4 gives experimental results of the proposed solution. Finally, the paper is concluded in Sect. 5.

2   Coughing Detection Audio Estimation

The proposed algorithm is illustrated in Fig. 1. The system consists of four main components: data segmentation, features extraction, classification, and data separation.

(a) Data Segmentation

(b) Features extraction

(c) Classification

(d) Data separation
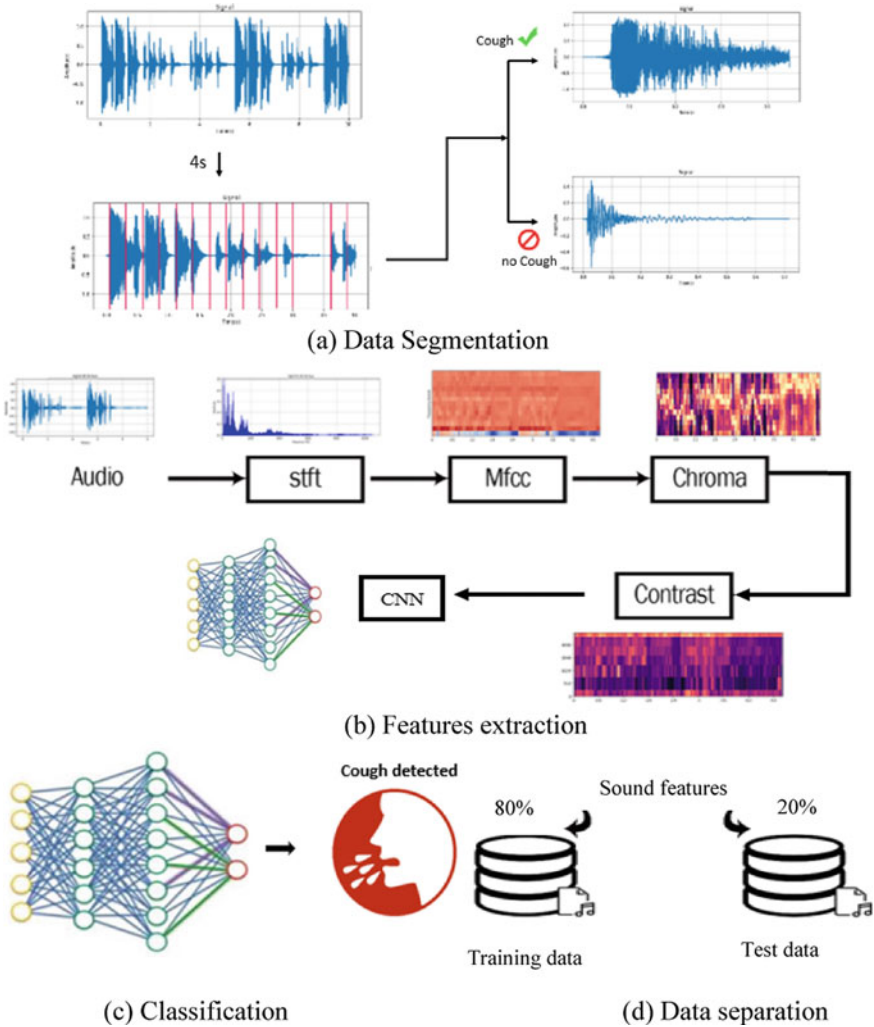
Fig. 1 Cough sound detection process

2.1 Data Segmentation

The dataset is labeled as cough and no-cough with a duration of 4 s that was chosen experimentally. The cough class contains the sound of pure cough, and the no-cough includes any sound except cough (environment, clear noise, speech …).

2.2 Feature Extraction

Using the librosa python library**,** four features of the audio files were extracted. These features are Mel frequency cepstral coefficients (MFCC), Short-Time Fourier Transform (STFT), Chroma, and Contrast.

- **Mel frequency cepstral coefficients (MFCC)**: It is a widely used feature in automatic sound recognition. It is the result of the real short-term log-cosines transformation of the energy spectrum, expressed by the Mel frequency scale [6]. The original sound is pre-processed by a pre-emphasis filter and a bandpass filter. Then, the pre-processed signal is segmented into frames and a window is added.
- **Short-Time Fourier Transform (STFT)**: It is performed on each frame, and the spectrum is squared. Then, the result is filtered by a bank of filters Mel filters to obtain an energy adapted to human frequencies (Mel Energy). The logarithm of the Mel energy is used. Then, a Discrete Cosine Transform (DCT) is performed. Finally, the MFCC are obtained.
- **Chroma**: Chroma or chrominance vector is a 12-element feature vector indicating the amount of energy in each pitch class, which identifies the property that allows the classification of the sound on a frequency-related scale, and the energy of each frequency is represented by a color [7]. The blue color corresponds to a low amplitude, and the more vivid colors (such as red) corresponds to amplitudes progressively stronger.
- **Contrast**: The contrast characterizes the light distribution of an image [8]. Visually, it can be interpreted as a spread of the histogram of brightness of the image. A high contrast image has a good dynamic distribution of gray values over the entire range of possible values, with clear whites or deep blacks. On the contrary, a low contrast image has a low dynamic range, with most pixels having very close gray values.

2.3 Classification

Convolutional Neural Network (CNN) is a machine learning technique inspired by the structure of the brain. It comprises a network of learning units called neurons. These neurons learn to convert input signals (in our case the spectrogram image of the cough) into corresponding output signals (the label "cough"), forming the basis for automated recognition.

A CNN architecture is made of a succession of processing blocks allowing to extract the features that discriminate the image class from the others. A treatment block consists of [9, 10]:

- Convolution layer (CONV), which processes the data from a receiver field, it is used to extract the different characteristics of the input images.

- Activation layer (ReLU), it is a non-linear activation function, which replaces all negative values received as inputs by zeros.
- Pooling layer (POOL), it allows to compress the information by reducing the size of the intermediate image, to improve network efficiency and avoid overlearning.
- Grouping layer (Flatten), which allows the grouping of feature maps in vector columns.
- Fully connected layer (FC), which allows to classify the input image of the network. It returns a vector where each element indicates the probability for the input image belongs to a class.

2.4 Data Separation

The general principle of audio classification systems includes two stages [11]:

1. A learning stage which can be seen as a development phase leading to the implementation of a classification strategy.
2. A testing stage by which the performance of the classification system is evaluated.

In general, a system is ready for real use only after a succession of learning and testing steps that allow the implementation of an efficient classification strategy.

2.5 Performance Evaluation

After training the model, the results obtained are observed and the training characteristics are varied in order to increase the accuracy rate and decrease the error rate.

The model should perform as well on the training data as on the validation data. This is the ideal case; it means that the model is efficient and recognizes the images it knows as well as those it has never seen.

3 Coughing Detection Pose Estimation

To indicate the movements of a person with the camera, this paper proposes to use the "multi-person pose estimation" model that detects the main points in the human body, knowing that in general a person who coughs places his hand or his elbow facing his mouth [12]. The developed algorithm calculates two indexes indicating whether a person is coughing or not. All the details are given bellow.

Fig. 2 Presentation of the "multi-person pose estimation" model indices [12]
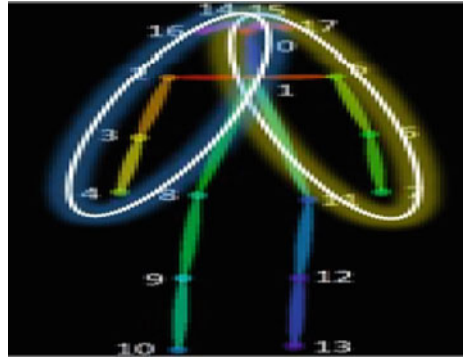
3.1 Multi-person Pose Estimation

The multi-person pose estimation model is a model that estimates the position of the 18 points (x, y) of $P0$–$P17$ in a 2D plane [13], where we can distinguish different points in the body such as elbows, knees, neck, shoulders, hips, and chest.

A person who coughs makes specific gestures and movements and then makes a coughing sound. The first reaction is to move the hand toward the mouth (right or left hand) and sometimes the whole arm toward the same sound outlet (the mouth) (Fig. 2).

We defined two main indices:

$$\text{Index\_R for the right side: } \frac{d_{5,6} + d_{6,7}}{2 * d_{0,7}} \tag{1}$$

$$\text{Index\_L for the left side: } \frac{d_{2,3} + d_{3,4}}{2 * d_{0,4}} \tag{2}$$

Each index uses the distances between different points. These are calculated from their coordinates (x, y), in order to propose an equation that allows us to define a cough threshold.

3.2 Threshold Validation

The idea was taken from a project on wink detection by Soukupova and Cech [14] who were able to validate the results by using the SVM model to have a threshold. The latter was also based on distances between specific points in the eye.

The same steps are used to validate the final index threshold in the next section.

To find a threshold that indicates the existence of cough, we first collect data to build a dataset, then apply a classification to obtain the threshold of detection of cough that we will integrate in our program, and finally make tests.

3.3 Dataset

To collect a real database, the first step was to use a video of a person coughing and another video where the person is not coughing, and then we store the values of the two indices (Index_R and Index_L), as shown in Figs. 3 and 4. For this purpose, the stored videos were cut in 60 frames per second in order to process each frame (Frame by Frame), as we aim to increase the accuracy of the results. The results are stored in an Excel sheet (.xlsx) to facilitate classification.

From Figs. 3 and 4, we can visualize the margin or the person coughs, and for both indices, the next step is to apply a classification algorithm to properly indicate and validate the chosen threshold, based on the Support Vector Machine.

Fig. 3 Presentation of Index_L
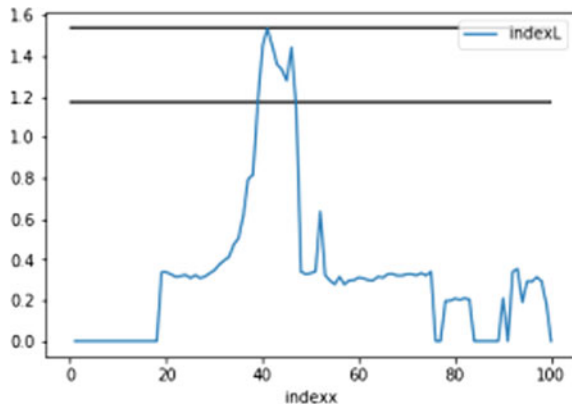
Fig. 4 Presentation of Index_R
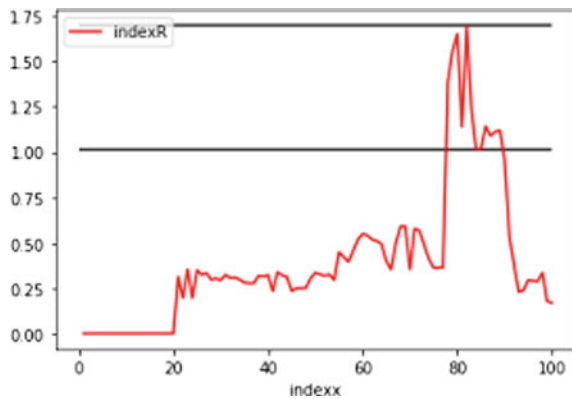
3.4 Support Vector Machine

Support Vector Machines (SVM) are a class of learning algorithms initially defined for discrimination. They have been then generalized to the prediction of a quantitative variable. In the case of discrimination of a dichotomous variable, they are based on the search of the optimal margin hyperplane which, when possible, correctly classifies or separates the data while being as far as possible from all the observations. The principle is therefore to find a classifier, or a discrimination function, with the highest possible generalization capacity (predictive quality) [15]. The choice of this model is motivated by non-negligible technical constraints, and SVM present in practice very good performances, it is able to provide good classification performances from a reduced number of learning examples while acting in very high dimensional spaces. We apply the SVM algorithm and repeat it until the final accuracy increases to 1.0, where the two index thresholds are: Index_L = 1.23 and Index_R = 1.23.

3.5 Performance Evaluation

To detect the person who coughs, the program must indicate the values of the two indexes (Left, Right):

- If Index_L $\geq$ 1.23 OR Index_R $\geq$ 1.23 $\rightarrow$ it's a coughing person.
- Else if Index_L < 1.23 and Index_R < 1.23 $\rightarrow$ it's a non-coughing person.

4 Results

4.1 Audio Detection

The proposed architecture consists of four convolution layers, followed by pooling layers. The activation function (ReLU) [10] is performed with the convolution and finally a fully neural network layer for classification. The accuracy of learning and validation increases with the number of epochs, i.e., the number of times an algorithm uses the dataset, this reflects that at each epoch the model learns more information. Similarly, the learning and validation error decreases with the number of epochs. There is no evidence of over-training or under-training, so the model has done the training well and can generalize on audio files that it has never seen.

The result of our model is as follows:

- A "waiting for detection" message is displayed when the program is executed.
- A "cough detected" message is displayed when the cough is identified.
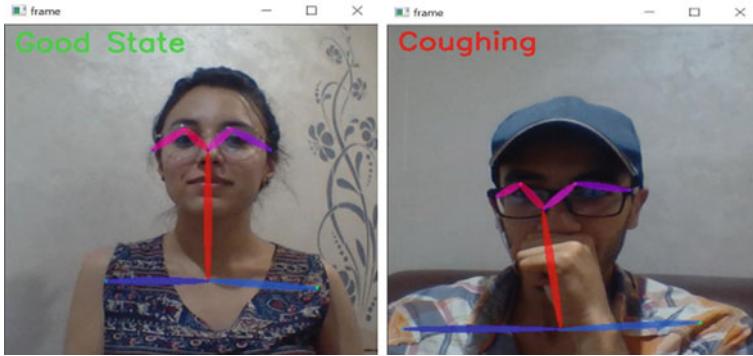
Fig. 5 Real-time results

This model has been developed using a TOSHIBA PROTÉGÉ laptop. It is characterized by: Windows 10 Professional x64, Intel(R) Core (TM) i7 CPU 2.70–2.90 GHz, 16 GO, and 237 Go SSD.

The model achieved an accuracy of 95.90% in one hour, 31 min, 34 s, with a learning error rate of 3.9472%.

4.2 Image Detection

To visualize the program results, the desktop camera is used. And the result is displayed in the video in real time, either Coughing or Good State as shown in Fig. 5.

The estimated time during the processing of a single image was between 0.3 and 0.5 s depending on the number of persons on the image. The estimated time during processing of each frame of the video was between 0.7 and 0.85 s depending on the condition of the person as well as the quality of PPI (Pixel Per Inch) and the resolution of the video (most of the time the tests are done by our computer camera).

5 Conclusion

This paper proposes an intelligent system capable of identifying one of the most common symptoms of COVID-19 (cough). The design of this system was carried out in several stages based on 2 main components: The first one allows to detect the sound of the cough, and the second one allows to locate the person who coughs. During this study, significant results are obtained. These results were presented and interpreted to show the effectiveness of the proposed methods. This progress gives the possibility to integrate this system in another more powerful system which includes

the detection of other symptoms of COVID such as body temperature and respiratory rate detection in order to give a more accurate diagnosis for carrying the COVID-19 virus.

References

1. Imran A, Posokhova I, Qureshi HN, Masood U, Riaz MS, Ali K, Nabeel M et al (2020) AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Inform Med Unlocked 20:100378
2. Brown C, Chauhan J, Grammenos A, Han J, Hasthanasombat A, Spathis D, Mascolo C et al (2020) Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. arXiv preprint arXiv:2006.05919
3. Gökcen A, Karadağ B, Riva C, Boyacı A (2021) Artificial intelligence-based COVID-19 detection using cough records. Electrica 21(2):203–208
4. Erdoğan YE, Narin A (2021) COVID-19 detection with traditional and deep features on cough acoustic signals. Comput Biol Med 136:104765
5. Tena A, Clarià F, Solsona F (2022) Automated detection of COVID-19 cough. Biomed Signal Process Control 71:103175
6. McFee B, Raffel C, Liang D, Ellis D, McVicar M, Battenberg E, Nieto O (2015) Librosa: audio and music signal analysis in python. In: Presented at the python in science conference. Austin, Texas, pp 18–24. https://doi.org/10.25080/Majora-7b98e3ed-003
7. Srivastava A, Jain S, Miranda R, Patil S, Pandya S, Kotecha K (2021) Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease. PeerJ Comput Sci 7:e369. https://doi.org/10.7717/peerj-cs.369
8. Lee D, Lee J, Ko J, Yoon J, Ryu K, Nam Y (2019) Deep learning in MR image processing. Investig Magn Reson Imag 23:81. https://doi.org/10.13104/imri.2019.23.2.81
9. Ghimire A, Thapa S, Jha AK, Kumar A, Kumar A, Adhikari S (2020) AI and IoT solutions for tackling COVID-19 pandemic. In: Presented at the 4th international conference on electronics, communication and aerospace technology (ICECA). IEEE, Coimbatore, India, pp 1083–1092. https://doi.org/10.1109/ICECA49313.2020.9297454
10. OpenClassrooms (2021) Découvrez les différentes couches d'un CNN—Classez et segmentez des données visuelles. Last accessed 12 June 2021
11. Affonso C, Rossi ALD, Vieira FHA, de Leon Ferreira ACP (2017) Deep learning for biological image classification. Exp Syst Appl 85:114–122
12. Chen S, Demachi K (2020) A vision-based approach for ensuring proper use of personal protective equipment (ppe) in decommissioning of fukushima Daiichi nuclear power station. Appl Sci, 10(15), 5129
13. Faber M (2019) https://github.com/michalfaber/keras_Realtime_Multi-Person_Pose_Estimation.github. Last accessed 09 July 2021
14. Soukupova T, Cech J (2016) Eye blink detection using facial landmarks. In: 21st computer vision winter workshop. Rimske Toplice, Slovenia
15. Drugman T, Urbain J, Dutoit T (2011) Assessment of audio features for automatic cough detection. In: 19th European signal processing conference. IEEE, pp 1289–1293

Power Electronics and Control Systems

Impact of Grid-Connected Photovoltaic System in the Power Quality of a Transmission Network

M. Dib, A. Nejmi, and M. Ramzi

Abstract This article presents the results of the impact study on the high-voltage (HV) transmission network of a 40 MW PV plant connected to the high impedance Moroccan HV network. The analysis of the power plant's behavior at the connection point includes the calculation of the active and reactive power flux and the harmonics. The simulation results from the Electrical Transient Analyzer Program (ETAP) software showed the influence of the short-circuit power of the network and the type of inverter on the harmonic ratio at the connection point of the photovoltaic (PV) plant with the HV network. The study recommends installing a low-pass filter to reduce harmonic distortion at the short-circuit power (PCC) connection point.

Keywords Renewable energy · Photovoltaic power plant · Harmonic · HV network

1 Introduction

Renewable energy sources are becoming an increasingly important part of the global energy mix. Morocco is no exception to this trend because it has a particularly rich potential compared to other countries.

A series of medium-sized PV plants (20–40 MW) are planned in Morocco's solar energy program.

However, the structure of the existing HV and medium voltage (MV) networks needs to be changed in order to effectively manage the integration of these renewable energy sources [1–6].

High penetration of photovoltaic systems will certainly have consequences for the operation of the electricity grid. However, the performance of PV systems connected to the grid can be affected due to the intermittent nature of PV, its high penetration, and the presence of nonlinear loads in the distribution system [7–9].

M. Dib (✉) · A. Nejmi · M. Ramzi
Laboratory of Automation, Energy Conversion and Microelectronics, Faculty of Science and Technology, Sultan Moulay Slimane University, B.P. 523, Beni Mellal, Morocco
e-mail: med3.dib@gmail.com

The energy quality problems encountered in the grid-connected PV system are slow and fast voltage fluctuations, overvoltages, imbalance, and harmonics [10–12].

These parameters must be in the range specified by the standards IEC and IEEE [10–12].

It is, therefore, important to analyze the PV plant's impact on the power quality parameters associated with the power system before connecting to the grid. The size and location of the plant are important parameters for evaluating the influence of the photovoltaic system on the HV network.

This part of the study evaluates the compatibility of a PV installation with the requirements concerning the harmonic disturbance generated at the PCC connection point with the HV network, i.e., the 60 kV voltage for our case study.

Harmonics are caused by the introduction on the network of nonlinear loads such as equipment incorporating power electronics. More generally, all materials incorporating inverters, rectifiers, and switching electronics deform currents and create voltage fluctuations on the low voltage distribution network.

Harmonic currents flowing through the impedances of the network give rise to harmonic voltages. The frequency domain that corresponds to the study of harmonics is generally between 100 and 2000 Hz, from the harmonic of rank 2 to the harmonic of rank 40. The maximum levels rank by rank are defined in the standard IEC 61000-3-6. The harmonic distortion rate (THD) is limited to 3% in 60 kV and 6.5% in 22 kV.

As far as their effects are concerned, harmonic currents and voltages may, in the long term as well as in the short term, have harmful consequences for equipment and network elements.

Further induced heating may reduce the service life of rotating machines, capacitors, power transformers and neutral conductors. Temporary or even permanent malfunctions of sensitive equipment may be recorded [13].

2   Description of the Study System

A harmonic study is necessary because the photovoltaic plant with inverters will be connected to the Moroccan low-power HV short-circuit network.

The study center will be a central site of one of the sites to be built and connected to the Moroccan HV network. The chosen 35 MVA PV plant consists of 130,560 panels, 17 inverters, 17 transformers 0.4/22 kV, and 2 transformers 60/22 kV one in service and the other of reserve.
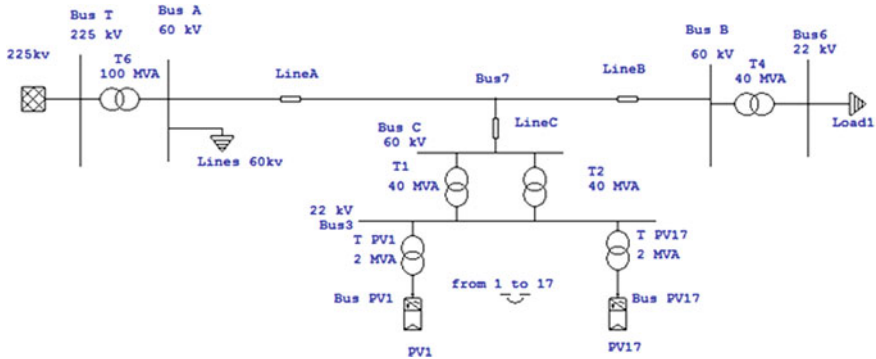
Fig. 1 Electrical diagram by ETAP

3 Modeling and Simulations

The representation of the photovoltaic installations in the ETAP calculation software is carried out white generators. Figure 1 represents a simplified diagram of connection of the plant to the 60 kV network.

4 Results and Discussions

PV inverters are modeled as power sources. The spectral behavior used is taken from manufacturer data. The individual voltage harmonics and the harmonic distortion rate (TDH) are calculated by injecting a harmonic current into the frequency scan of the impedances (Fig. 2).

The simulation results by ETAP show that the distortion rate is of the order of 3.47%. This rate exceeds the value of the standard which is 3%.

Figure 3 shows the individual harmonic rates THDi.

Note that the higher harmonics of ranks H23, H25, and 26 have fairly high values which are responsible for this value of 3.47% of the overall rate.

The harmonic distortion depends on the types of inverters that are harmonic generators and the short-circuit power of the network to which the PV plant is connected. The smaller this value, the higher the distortion rate. As the voltage distortion exceeds the limits defined in IEC 61000-3-6, the harmonic processing must be taken into consideration.

Two solutions to reduce the harmonic rate will be presented.
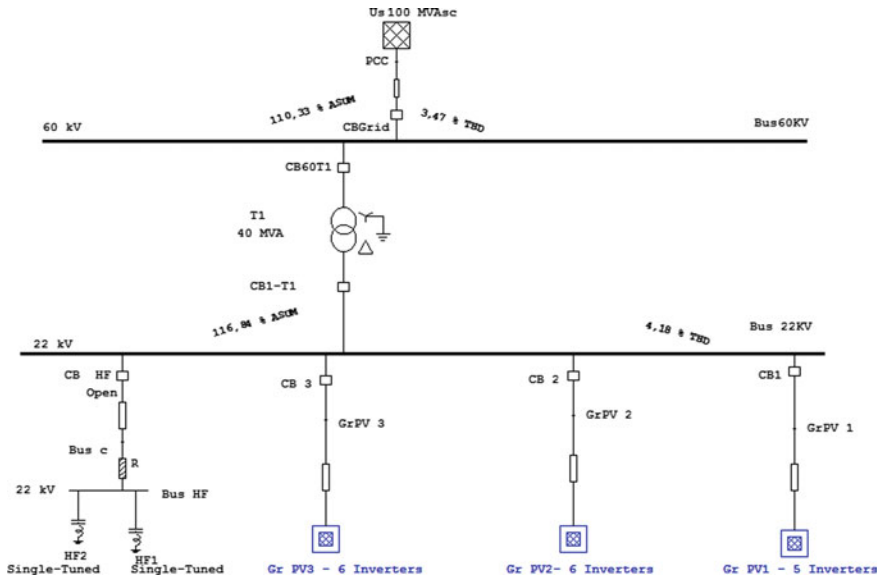
Fig. 2 Harmonic simulation result without filter
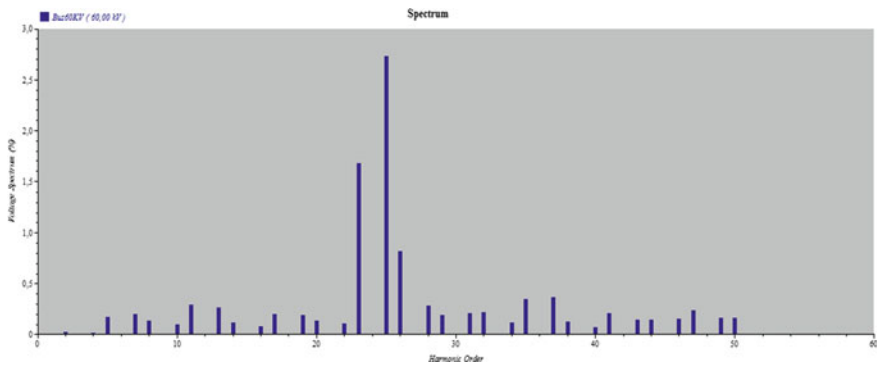
Fig. 3 % the harmonics individuals-without filter

4.1 Harmonic Reduction by Increasing Short-Circuit Power

The short-circuit power is inversely proportional to the impedance of the network. More than the short-circuit power is large, the network impedance is low.

To reduce the THD values, the level of short circuit in the appropriate sectors of the network can be increased. This can be done by additional connections to the main branch or by new connections to other parts of the network.
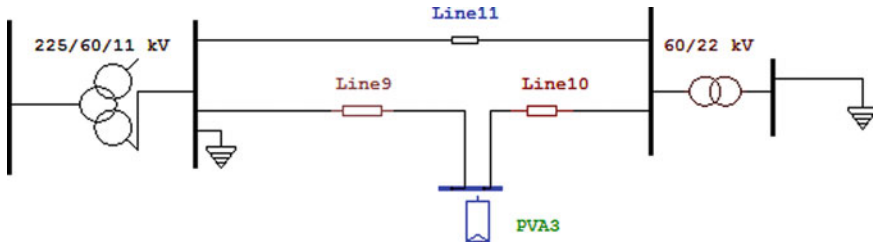
Fig. 4 Connection method

In order to respect the criterion of operation of the network in situation $N - 1$ and to increase the power of short circuit, the variant consists of producing a double-dotted line with two departure spans at the elevator station and to make another line between the existing posts. In this case, the connection mode of the control panel will be as follows (Fig. 4).

With this configuration, the short-circuit power will become 130 MVA.

For this short-circuit power, the VDH will be 2.92%.

For a short-circuit power equal to 125MVA, the THD is the 3% limit value indicated by IEC 61000-3-6.

With the increase in short-circuit power, it was possible to reduce the harmonic distortion at the connection point. However, these network enhancement measures would be very expensive since the distances are very long and are not required from the point of view of power flow. So it is not recommended to strengthen the network just to reduce the THD.

The distortion is influenced mainly by the type of inverter installed which is a source of higher order harmonic currents that must be filtered.

4.2 Reduction of Harmonics by Installing the Low-Pass Filter

To limit the current harmonics injected by the inverters, it is recommended to install a low-pass filter.

The values of the low-pass filter components are as follows: $C = 19.73\ \mu F$, $L1 = 0.51$ H, $R = 10\ \Omega$.

The results of harmonic distortion of the voltage at the point of connection, after the addition of the low-pass filter, are presented in Fig. 5.

THD became 0.89% well below the limit of IEC 61000-3-6.

Figure 6 represents the spectrum of the individual harmonics.

The higher harmonics object of the filtering is well eliminated. On the other hand, some harmonics of lower ranks have become significant.

Without a filter, the upper ranks H23, H25, H26, and H28 subjected to filtering respectively represented 1.68%, 2.7%, 0.82, and 0.28%. After filtration, these ranks
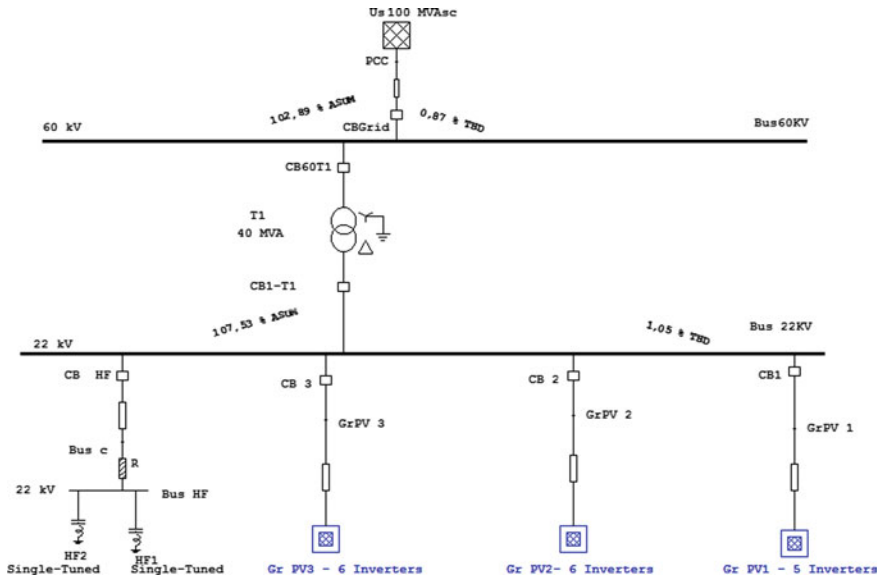
Fig. 5 % harmonics with filter
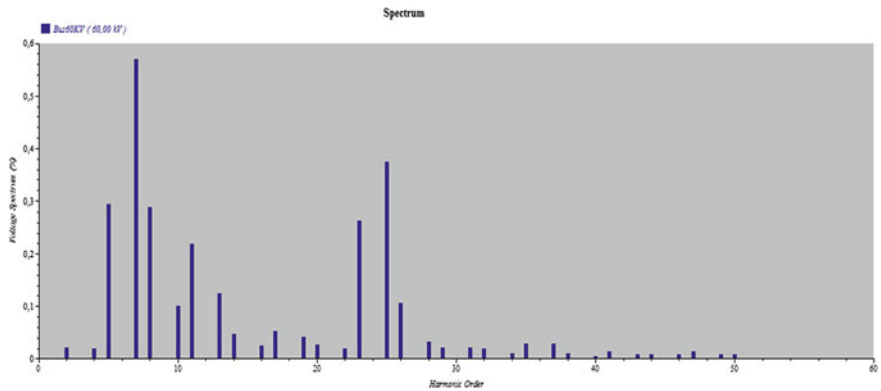
Fig. 6 % harmonics individuals with filter

respectively represent 0.26%, 0.37%, 0.11, and 0.03%. These ranks have fallen below the values of the norm.

By tale, the lower rank harmonics H5, H7, and H8 have undergone an increase compared to their states before filtering and exceed the limits of the norm. After filtering, the lower ranks H5, H7, and H8, respectively, represent 0.29%, 0.57%, and 0.28% against the values 0.17%, 0.20%, and 0.13 before filtering. Despite this increase, the overall rate became 0.89% after filtering against 3.47% before filtering.

A peak is observed at harmonic level 7. The resonance problem can appear.

The resonance phenomenon is at the origin of the most important harmonic distortions in the network.

5   Conclusions

The simulations carried out on the network showed the influence of the photovoltaic system on the quality of the energy of the Moroccan transmission network. The results show that, without harmonic filtering, total harmonic distortion (THD) exceeds the limit. The reason is the low level of short circuit at the end of the antenna network, where the PV plant will be connected. With the installation of filters, total harmonic distortion (THD) became acceptable.

To maintain harmonic levels below acceptable thresholds in order to limit their impact, network operators can use several levers:

- The installation of filters that can absorb some of the harmonic currents emitted by the installations.
- The limitation of disturbances at the source, by setting emission limits for the devices and for the installations.
- Increased the short-circuit power at the connection point by creating the loop lines of the source stations.

In addition, an impact study of all the PV power plants that will be granted on the same HT artery of the Moroccan HV network is necessary in order to evaluate the influence of one on the others on the one hand and their impact on the network at the connection points on the other hand.

References

1. Bhattacharyya S, Myrzik JMA, Kling WL (2007) Consequences of poor power quality e an overview. In: Proceedings of the universities power engineering conference, vol 1, pp 651–656
2. Muhamad MI, Manm N, Radzi MAM (2007) The effects of power quality to the industries. In: 5th student conference on research and development, pp 1–4
3. Wang YJ, O'Connell RM, Brownfield G (2001) Modeling and prediction of distribution system voltage distortion caused by nonlinear residential loads. IEEE Trans Power Deliv 16(4):744–751
4. Patsalides M, Stavrou A, Efthymiou V, Georghiou GE (2012) Towards the establishment of maximum PV generation limits due to power quality constraints. Int J Electr Power Energy Syst 42(1):285–298
5. Alvarez-Herault MC, Caire R, Raison B, Hadjsaid N, Bienia W (2012) Optimizing traditional urban network architectures to increase distributed generation connection. Int J Electr Power Energy Syst 35(1):148–157
6. Leão RPS, Barroso GC, Sampaio RF, Almada JB, Lima CFP, Rego MCO et al (2011) The future of low voltage networks: moving from passive to active. Int J Electr Power Energy Syst 33:1506–1512

7. Eltawil MA, Zhao Z (2010) Grid connected photovoltaics power systems: technical and potential problems—a review. Renew Sustain Energy Rev 14(1):112–129

8. Eftekharnejad S, Vittal V, Heydt GT, Keel B, Loehr J (2013) Impact of increased penetration of photovoltaic generation on power systems. IEEE Trans Power Syst 28(2):893–901

9. González P, Romero-Cadaval E, González E, Guerrero MA (2011) Impact of grid connected photovoltaic system in the power quality of a distribution network. In: Camarinha-Matos LM (eds) Technological innovation for sustainability. DoCEIS 2011. IFIP advances in information and communication technology, vol 349. Springer, Berlin

10. Vinayagam A, Swarna K, Khoo S, Stojcevski A (2016) Power quality analysis in microgrid: an experimental approach. J Power Energy Eng 4:17–34

11. Ayub M, Gan CK, Kadir AFA (2014) The impact of grid-connected PV systems on harmonic distortion. In: IEEE innovative smart grid technologies—Asia (ISGT ASIA). Kuala Lumpur, pp 669–674

12. Meshram S, Agnihotri G, Gupta S (2013) Performance analysis of grid integrated hydro and solar based hybrid systems. Adv Power Electron. https://doi.org/10.1155/2013/697049

13. Girgis AA, Nims J, Jacomino J, Dalton JG, Bishop A (1992) Effect of voltage harmonics on the operation of solid-state relays in industrial application. IEEE Trans Ind Appl 28(5):1166–1173

Nonlinear Control of a Three-Phase, Double-Stage Grid-Tied Photovoltaic System

Salwa Naddami, Najib Ababssi, and Mohcine Mokhlis

Abstract In this paper, a robust backstepping control technique is proposed to enhance the power quality of a double-stage, three-phase grid-connected photovoltaic (PV) system. A two-step control process is used in the research. On the PV array side, P&O-backstepping MPPT is employed to extract the maximum power by controlling a boost DC converter. On the grid side, the backstepping controller is designed to ensure a quasi-total transmission of the extracted PV power to the grid with a unity power factor and reduced harmonic distortion by controlling the *d*- and *q*-axis grid currents. The DC link voltage is regulated to maintain a constant value under various conditions, which is employed to obtain a reference value of the direct current vector, that is responsible for active power control. The performance of the backstepping controller in terms of robustness under different operating conditions and disturbances is validated in the MATLAB/Simulink environment.

Keywords PV modules · Grid · Boost converter · Inverter · Backstepping control · P&O · MPPT

1 Introduction

In today's world, as conventional energy resources are limited, there is a significant increase in demand for renewable energy because it produces clean, pollution-free, and renewable energy. For the same purpose, the Kingdom of Morocco has adopted since 2009 under the High Royal Directions an energy map as an objective, the rise in power of renewable energy, strengthening energy efficiency and regional integration [1].

S. Naddami (✉) · N. Ababssi
IMII Laboratory, Faculty of Science and Technology, Hassan First University, 26000 Settat, Morocco
e-mail: s.naddami@uhp.ac.ma

M. Mokhlis
Department of Electrical Engineering, Mohammadia School of Engineers, Mohammed V University, Rabat, Morocco

The integration of wind and solar energy source into the power grid at the distribution level is a new development. These energy resources inject electric power directly through a solar system based on power electronic converters [2]. Consequently, solar photovoltaic systems are composed of an array of solar modules, a DC–DC power converter, and an AC–DC inverter as the final interface [3, 4].

DC converter employed in this work is a boost, which is stepping up the voltage to the required level through the Maximum Power Point Tracking strategy selected.

In the literature, we find some techniques based on the perturbation and observation (P&O) algorithm [5], others on artificial neural networks such as the fuzzy algorithm [6], the ripple correlation control (RCC) [7], and the sliding mode control [8].

The first suggested controller in this study comprises the backstepping command combined with the P&O MPPT technique, which serves a good performance. A classic P&O provides an optimal voltage reference that is employed on a backstepping controller and adjusts the duty cycle of the switching pulse [9]. Several studies were conducted for injecting the maximum active power with a unit power factor. In this research [10], the author proposed a traditional method to control active power by implementing a PI regulator. However, the inverter is a highly nonlinear system, and the use of a linear controller may not follow the reference value across a large operating range.

Nonlinear control of the AC inverter is presented in this paper to resolve the problem of linearity and enhance the global stability of the system by implementing a backstepping approach to ensure the injection of three-phase current into the grid while maintaining a unit power factor (UPF).

2   Description of the System Architecture

As shown in the system (see Fig. 1), the considered system contains subsystems such as PV array, DC converter, VSI, RL filter, grid, and control blocs.

2.1   *Photovoltaic Array*

Modeling PV array starts with the solar cell. Providing a good ratio simplicity/accuracy, the simple model of solar cell contains five parameters as in [11].

The equation that governs the voltage–current characteristic is as follows:

$$I_{pv} = I_{ph} - I_s \left(e^{\frac{V_{pv} + R_s I_{pv}}{a V_t}} - 1 \right) - \frac{V_{pv} + R_s I_{pv}}{R_{sh}} \tag{1}$$
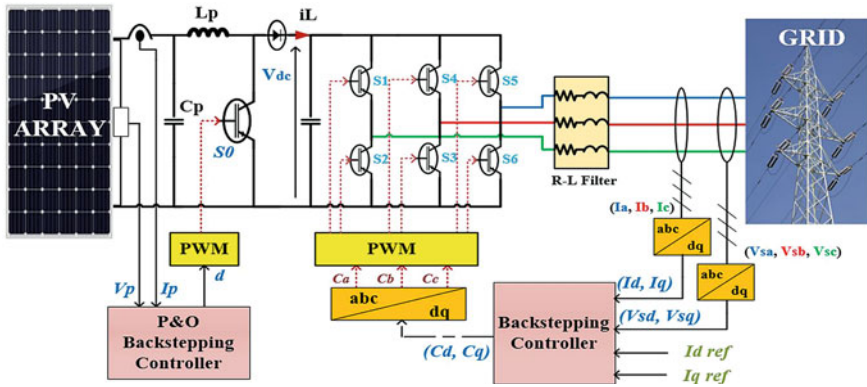
Fig. 1 Grid-tied PV system configuration

where I_{pv}, photo current; I_s, cell saturation of dark current; V_t, Thermal voltage which is / $V = kT/q$; q, 1.6×10^{-19} C charge of an electron; T, the cell's working temperature; A, an ideality factor; R_p, shunt or parallel resistance; and R_s, series resistance.

2.2 DC Converter Model

The DC converter, also known as a chopper, is a switching converter with an output voltage that can be smaller or higher than the input. In the study system, a DC converter is used to level up the output voltage. By application of the basic theory of boost converter (see Fig. 1), the dynamic equations can be formulated as follows:

$$\frac{dVp}{dt} = \frac{1}{Cpv}I_p - \frac{1}{Cp}I_L \tag{2}$$

$$\frac{diL}{dt} = \frac{1}{L}Vp - \frac{1}{L}(1-d)V\mathrm{dc} \tag{3}$$

$$\frac{dV\mathrm{dc}}{dt} = \frac{1}{C\mathrm{dc}}(1-d)iL - \frac{1}{C\mathrm{dc}}I\mathrm{dc} \tag{4}$$

2.3 Voltage Source Inverter Model

The inverters are the most widely used power converters when interfacing DC sources like PV generators to the power grid (see Fig. 1). The basic architecture of the

inverter consists of six bidirectional switches (S1–S6). Three-phase PWM signals are synchronized to the grid by applying PLL control. The control of VSI will also assure regulation of the input DC voltage by using the PI regulator.

The inductive filter (see Fig. 1) is used to reduce the majority of harmonics from the inverter producing a pure sine wave and due to its simplicity of manufacturing.

In this research, the inverter provides only active power to achieve unity power factor (UPF).

From the state space equations of the inverter stage and after the PARK transformation, we obtain the representation in the DQ frame.

$$\frac{dVdc}{dt} = \frac{1}{Cdc}(1-d)iL - \frac{3}{2Cdc}(Cd\,id) - \frac{3}{2Cdc}(Cq\,iq) \tag{5}$$

$$\frac{di_d}{dt} = wi_q - \frac{Rf}{Lf}i_d - \frac{1}{Lf}Vs_d + \frac{Vdc}{Lf}Cd \tag{6}$$

$$\frac{di_q}{dt} = -wi_d - \frac{Rf}{Lf}i_q - \frac{1}{Lf}Vsq + \frac{Vdc}{Lf}Cq \tag{7}$$

3 Control Strategy

In this work, the proposed control strategy is based on the backstepping nonlinear technique. The first control acts on the PV array side, to extract the maximum power and to boost the voltage to the required level dictated by the power grid voltage.

The second control is applied on the grid side, to regulate the active and reactive power injected into the grid and obtain a unit power factor (UPF).

3.1 Maximum Power Point Tracking (MPPT) Control on the PV Array Side

The P&O-backstepping approach is used to control the DC converter in the considered system to optimize the extracted power. The voltage reference (V_{ref}) from the P&O MPPT block is tracked by this controller. Taouni et al. [12] describes the design of the boost converter's backstepping controller. The backstepping controller must generate the following control low:

$$d = \frac{L}{V_{dc}}\left(-C_2\mathcal{E}_2 + \frac{1}{C_p}\mathcal{E}_1 - \frac{1}{L}(V_p - V_{dc}) + \dot{\alpha}\right) \tag{8}$$

where

$$\mathcal{E}_1 = V_{pv} - V_{\text{ref}} \tag{9}$$

$$\mathcal{E}_2 = I_L - \alpha \tag{10}$$

$$\alpha = I_{pv} + C_{pv}\left(C_1\mathcal{E}_1 - \dot{V}_{\text{ref}}\right) \tag{11}$$

With α is the desired value of inductive current, C_1 and C_2 are gain parameters of the controller.

3.2 Power Controller Design on the Grid Side

The objective of this control strategy is to inject the power extracted from the PV array into the grid with lower harmonic distortion, by controlling direct and quadrature currents of inverter.

First Step. To control the reactive power, we set the tracking error and its derivative:

$$\mathcal{E}_{iq} = i_q - i_{q\text{ref}} \tag{12}$$

$$\dot{\mathcal{E}}_{iq} = wi_d - \frac{R_f}{L_f}i_q - \frac{1}{Lf}Vs_q + \frac{V_{dc}}{L_f}C_q \tag{13}$$

Lyapunov function and its derivative are expressed as follows:

$$V_{iq}\left(\mathcal{E}_{iq}\right) = \frac{1}{2}\mathcal{E}_{iq}^2 \tag{14}$$

$$\dot{V}_{iq}\left(\mathcal{E}_{iq}\right) = \mathcal{E}_{iq}\left[wi_d - \frac{R_f}{L_f}i_q - \frac{1}{Lf}Vs_q + \frac{V_{dc}}{L_f}C_q\right] \tag{15}$$

The control low C_q is chosen so as to obtain the following expression:

$$wi_d - \frac{R_f}{L_f}i_q - \frac{1}{Lf}Vs_q + \frac{V_{dc}}{L_f}C_q = -C_{iq}\mathcal{E}_{iq} \tag{16}$$

Then the expression of the control low is as follows:

$$C_q = \frac{L_f}{V_{dc}}\left[-C_{iq}\mathcal{E}_{iq} + wi_d + \frac{R_f}{L_f}i_q + \frac{1}{Lf}Vs_q\right] \tag{17}$$

Second Step. This step aims to regulate the direct current to the desired value $i_{d\text{ref}}$ to inject the maximum active power into the grid.

$$\mathcal{E}_{id} = i_d - i_{dref} \tag{18}$$

$$\dot{\mathcal{E}}_{id} = wi_q - \frac{R_f}{L_f}i_d - \frac{1}{Lf}Vs_d + \frac{V_{dc}}{L_f}C_d \tag{19}$$

The Lyapunov function and its derivative are as follows:

$$V_{id}(\mathcal{E}_{id}) = \frac{1}{2}\mathcal{E}_{id}^2 \tag{20}$$

$$\dot{V}_{id}(\mathcal{E}_{id}) = \mathcal{E}_{id}\left[wi_q - \frac{R_f}{L_f}i_d - \frac{1}{Lf}Vs_d + \frac{V_{dc}}{L_f}C_d\right] \tag{21}$$

The control low C_d is chosen so as to obtain the following expression:

$$wi_q - \frac{R_f}{L_f}i_d - \frac{1}{Lf}Vs_d + \frac{V_{dc}}{L_f}C_d = -C_{id}\mathcal{E}_{id} \tag{22}$$

$$C_d = \frac{L_f}{V_{dc}}\left[-C_{id}\mathcal{E}_{id} + wi_q + \frac{R_f}{L_f}i_d + \frac{1}{Lf}Vs_d\right] \tag{23}$$

where C_{id}, C_{iq} are parameters of the controller that determine tracking performance. The tracking error converges asymptotically to the origin, which ensures the global stability of the system.

After obtaining the control low C_d and C_q, the park transformation from DQ to ABC is used to get the three modulating signals C_A, C_B, C_C. Using a PWM generator, these three modulating signals will create six switching pulses for the inverter as illustrated (see Fig. 1).

4   Simulation and Results

In this paper, the simulation of the grid-connected PV system is performed under the MATLAB/Simulink platform, with a real-time reduced profile of solar irradiance and temperature (see Fig. 2). The values are taken from PVGIS tool [13], for the geographic location of Settat City in Morocco during a normal day of April.

The PV generator can produce up to 100 KW (under STC), and the inverter receives the PV power through the boost converter, which regulates the DC voltage at the value of 600 V. Finally, the inverter injects the active power into a 400 V line to line grid.

The power injected into the grid, compared to the produced PV power, indicates that the tracking of the maximum power is achieved with accuracy and rapidity (see Fig. 3).
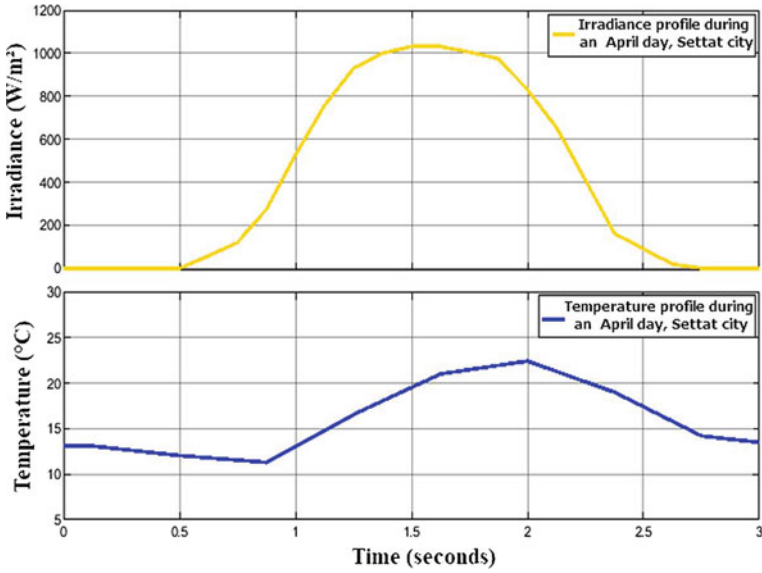
Fig. 2 Real profile of irradiation and temperature of the Settat City, Morocco, reduced in time
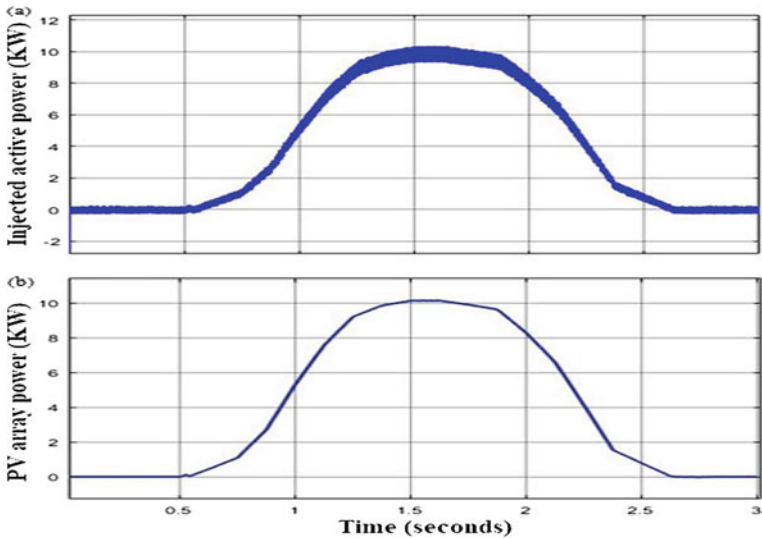
Fig. 3 **a** Active power injected into the grid, **b** the PV array power
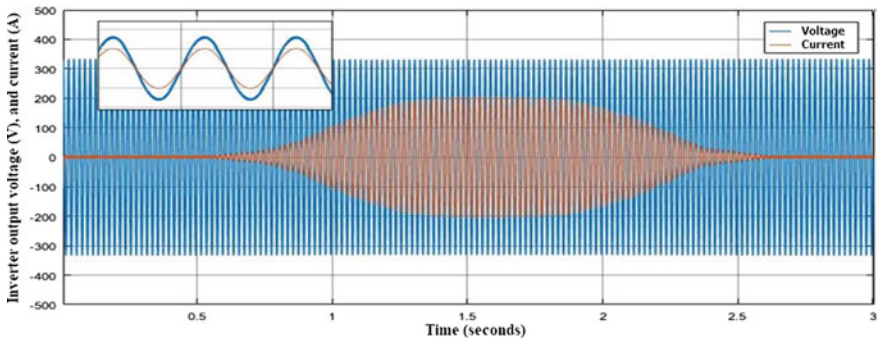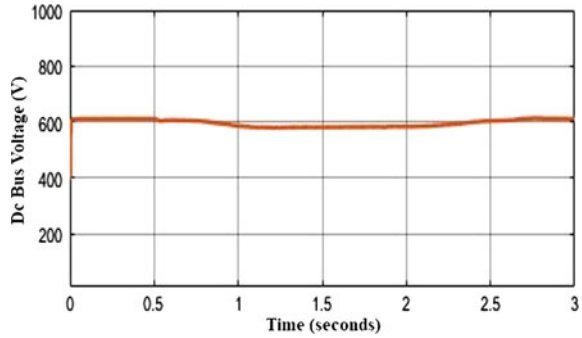
Fig. 5 Inverter output voltage and current

The simulation takes environmental and load variations. The DC link voltage is regulated to the required value of 600 V, and a small drop of 5% of DC bus voltage is noticed from 0.6 to 2.5 s, due to irradiation increase (see Fig. 4).

The simulation shows that the output inverter current is sinusoidal and in phase with the grid voltage (see Fig. 5), which means that the unit power factor is achieved.

5 Conclusion

This paper presents the backstepping control of a two-stage three-phase grid-connected PV system, where the whole system is modeled by mathematical equations. For the first stage, the boost regulator consists of a P&O-backstepping MPPT to extract the maximum power generated by the PV array. This power will be injected into the grid with a unit power factor (UPF) and low harmonic distortion. A second backstepping regulator is designed to control the inverter current i_d and i_q in such a way that the grid current tracks the reference $i_{d\text{ref}}$ and $i_{q\text{ref}}$.

The simulation results show that all the set objectives are achieved under variable irradiation and temperature conditions, respecting the parameters of accuracy and rapidity that qualify the good quality of the adopted control.

Future work will focus on hybrid generation systems, power flow, and energy storage integration.

References

1. Observation et Prospective. https://www.mem.gov.ma/Pages/secteur.aspx?e=10. Accessed 14 Nov 2021
2. Blaabjerg F, Yang Y, Yang D, Wang X (2017) Distributed power-generation systems and protection. Proc IEEE 105:1311–1331. https://doi.org/10.1109/JPROC.2017.2696878
3. Santos OL Contribution to the DC–AC conversion in photovoltaic systems: module oriented converters, p 249
4. Baharlou S, Yazdani MR (2017) A non-isolated high step-up DC/DC converter with low EMI and voltage stress for renewable energy applications. J Electr Eng Technol 12:1187–1194. https://doi.org/10.5370/JEET.2017.12.3.1187
5. Kamran M, Mudassar M, Fazal MR, Asghar MU, Bilal M, Asghar R (2020) Implementation of improved perturb & observe MPPT technique with confined search space for standalone photovoltaic system. J King Saud Univ Eng Sci 32:432–441. https://doi.org/10.1016/j.jksues.2018.04.006
6. Algarín CR, Giraldo JT, Álvarez OR (2017) Fuzzy logic based MPPT controller for a PV system. Energies 10:2036. https://doi.org/10.3390/en10122036
7. Sahu P, Sharma A, Dey R (2020) Ripple correlation control maximum power point tracking for battery operated PV systems: a comparative analysis. In: 2020 IEEE international IOT, electronics and mechatronics conference (IEMTRONICS). IEEE, Vancouver, BC, Canada, pp 1–6
8. Chinnappan R, Logamani P, Ramasubbu R (2019) Fixed frequency integral sliding-mode current-controlled MPPT boost converter for two-stage PV generation system. IET Circ Dev Syst 13:793–805. https://doi.org/10.1049/iet-cds.2018.5221
9. Vazquez JR, Martin AD (2015) Backstepping control of a buck-boost converter in an experimental PV-system. J Power Electron 15:1584–1592. https://doi.org/10.6113/JPE.2015.15.6.1584
10. Boudaraia K, Mahmoudi H, El Azzaoui M (2016) Modeling and control of three phases grid connected photovoltaic system. In: 2016 international renewable and sustainable energy conference (IRSEC). IEEE, Marrakech, pp 812–816
11. Banu I-V, Asachi G (2012) Modeling and simulation of photovoltaic arrays, p 7
12. Taouni A, Abbou A, Akherraz M, Ouchatti A, Majdoul R (2016) MPPT design for photovoltaic system using backstepping control with boost converter. In: 2016 international renewable and sustainable energy conference (IRSEC). IEEE, Marrakech, pp 469–475
13. JRC Photovoltaic Geographical Information System (PVGIS)—European Commission. https://re.jrc.ec.europa.eu/pvg_tools/fr/. Accessed 15 Dec 2021

Experimental Test of a Three-Phase Inverter Using a Launchpad TMS320F28379D Card

Chaymaâ Boutahiri, Ayoub Nouaiti, Aziz Bouazi, and Abdallah Marhraoui Hsaini

Abstract This paper presents an experimental test of a three-phase inverter feeding an induction motor. The studied system is composed of a didactic three-phase inverter, and a Launchpad TMS320F28379D card. A sinusoidal pulse width modulation (SPWM) control technique is applied to generate alternating sinusoidal currents. The simulation tests through MATLAB Simulink are verified experimentally at the laboratory. The obtained results prove the efficiency of the adopted control.

Keywords Three-phase inverter · Induction motor · SPWM · DSP

1 Introduction

In the industrial areas, electrical drives require more and more variable speeds. For this purpose, the progress of power electronics has allowed the development of several techniques to realize variable speed AC drives [1, 2].

Inverter is one of the solutions provided by power electronics to get AC currents without being limited to conventional electricity generators, and also to improve energy quality. It is based on fast and robust control devices. The applications of inverters differ according to their work field. There are mainly used to feed induction motors and standalone systems [3].

Several switching techniques are used to control the DC-AC converter such as hysteresis controller, sinusoidal pulse width modulation(SPWM), discontinuous pulse width modulation, and space vector pulse width modulation techniques [4].

Digital signal processing cards (DSP) are used to control converters; there are based on microprocessors to run digital applications as fast as possible. Indeed, these processors allow providing not only very important computational efforts in order to reach real time performances but also flexibility at the programming level [5–7].

C. Boutahiri (✉) · A. Nouaiti · A. Bouazi · A. M. Hsaini
Laboratory of Computer Science, Applied Math and Electrical Engineering (IMAGE), IEVIA
Team, EST, Moulay Ismail University, Meknes, Morocco
e-mail: c.boutahiri@edu.umi.ac.ma

The presented work exposes the test of the SPWM switching method for a three-phase inverter. The over system is based on a three-phase inverter from Semikron and a Texas instrument TMS320F28379D Launchpad card. To observe the behavior of voltages and currents, the simulation and experimental results are exposed.

2 Description of the Studied System

The studied system, as shown in Fig. 1, consists of an induction motor supplied with a three-phase Semikron inverter. This latter is powered through a direct voltage and controlled by SPWM signals generated with a development kit from Texas instruments.

2.1 Semikron Inverter

The Semikron inverter consists of a diode rectifier, a three-phase voltage inverter, a chopper, four drivers for the four IGBT arms, two capacitors to filter the DC bus voltage, and a cooling system with external sockets. Figure 2 and Table 1 present some specifications of the Semikron inverter (Max current 30 A), while Fig. 3 exposes its main structure [8].

2.2 Induction Motor

The induction motor (IM) is composed of a stator winding and the rotor winding (or squirrel-cage). Figure 4 shows the electrical model of the IM. X_1 and X_2 are the reactance of stator and rotor, respectively, X_m the magnetizing inductor, R_s the stator resistance, and S is the velocity slip [9].

Fig. 1 Structure of the studied system
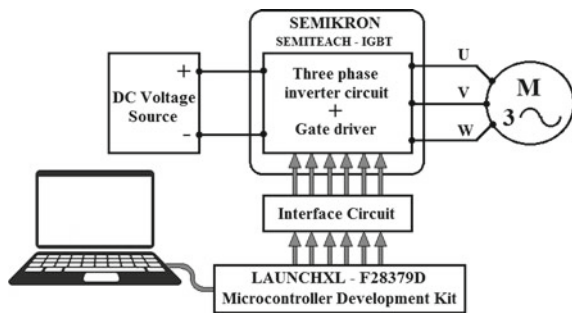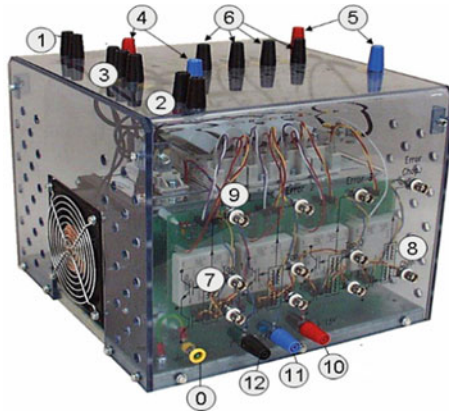
Fig. 2 Overview of the Semikron inverter

Table 1 Some specifications of the used Inverter

No	Function	Voltage
3	Rectifier input	230/400 V
5	DC IGBT inverter inputs	600 VDC
6	AC IGBT inverter + chopper outputs	400 VAC/600 VDC
7	PWM input of inverter	C-MOS logic 0/15 V
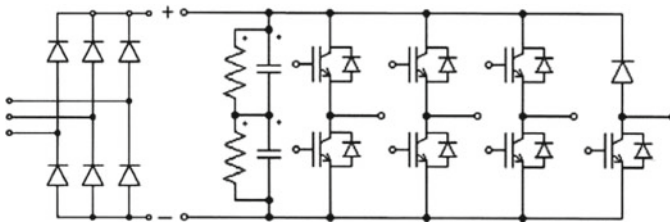
Fig. 3 Structure of the Semikron inverter
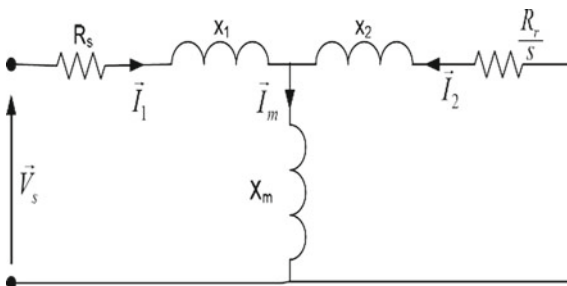
Fig. 4 Steady state IM model

2.3 LAUNCHXL-F28379D Microcontroller Development Kit

Texas instrument TMS320F28379D Launchpad (see Fig. 5) is a 32-bit dual core. Its main components are general purpose input/output (GPIO), analog to digital converter (ADC), digital to analog converter (DAC), and enhanced pulse width modulator (ePWM) module. Each ePWM module is identical with two PWM outputs, ePWMA and ePWMB, and multiple modules can be synchronized to work together as required by the system application, to generate complex pulse width waveforms [10, 11].

3 Programmed SPWM Technique

The conventional sine-triangle pulse width modulation (SPWM) in the three-phase version is performed by comparing a low frequency modulating signals (three sinusoidal references waves phase-shifted by 120° with each other) with a high-frequency carrier signal (triangular wave) (see Fig. 6) [12, 13]. Two parameters characterize this technique:

The modulation index m which is the frequency ratio of the carrier f_c to the reference f_{ref} (Eq. 1), and the modulation rate r which is the ratio of the voltage amplitudes of the reference V_{ref} to the carrier V_c (Eq. 2).

Fig. 5 TMS320F28379D card
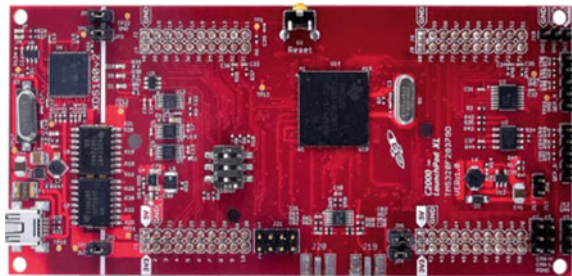
Fig. 6 Three-phase SPWM technique
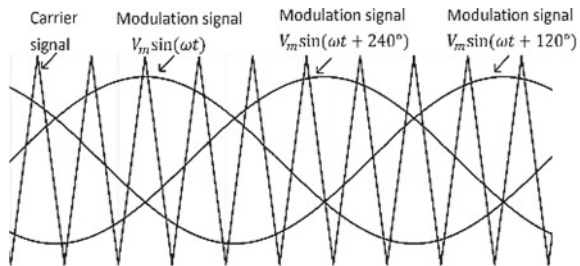
Fig. 7 SPWM control technique
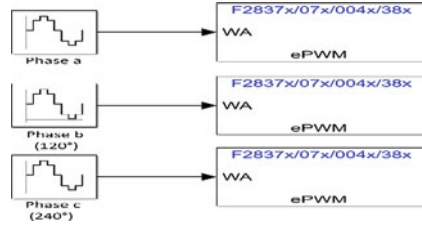
Fig. 8 Counting modes of the CPU timer
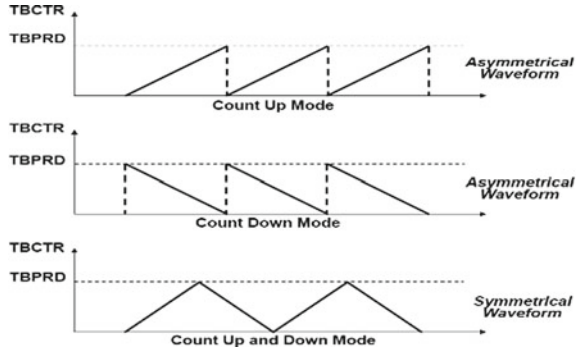
$$m = \frac{f_c}{f_{\text{ref}}} = \frac{T_{\text{ref}}}{T_c} \tag{1}$$

$$r = \frac{V_{\text{ref}}}{V_c} \tag{2}$$

The SPWM technique is applied to control the three arms of the Semikron inverter using the TMS320F28379D card. The ePWM modules are programmed with MATLAB Simulink to obtain the gate pulses of the inverter (Fig. 7) [14]. The ePWM module units compare the reference wave with the internal carrier signal. This latter is programmed using the counting modes of the CPU timer (Fig. 8). The reference signals are three positives sine waves shifted by 120° and configured to adjust the modulation rate r.

4 Simulation and Experimental Results

4.1 Simulation Results

The studied system supplied with a DC voltage (300 V) is tested with MATLAB Simulink to feed an induction motor of 1.5 Kw with a torque load of 3 Nm (see

Fig. 9). The SPWM technique is applied with a modulation rate $r = 1$ and a carrier frequency fc = 4 kHz. Figures 10, 11 and 12 expose the obtained results.

Figure 10 shows the AC voltage and current across the IM with a frequency of 50 Hz. Figure 11 indicate the AC currents per phase, while Fig. 12 expose the obtained speed and the electromagnetic torque. The RMS value of the voltage per phase is 128 V and per line is 222 V, while the RMS value of the AC currents per phases is 2.9 A. The developed speed and torque of the IM are 1380 Rpm and 3.8 Nm, respectively.
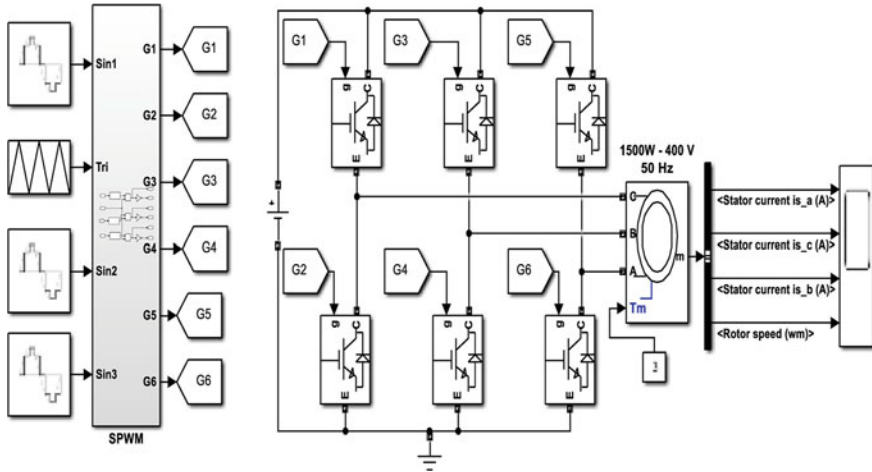
Fig. 9 Simulink blocks diagram
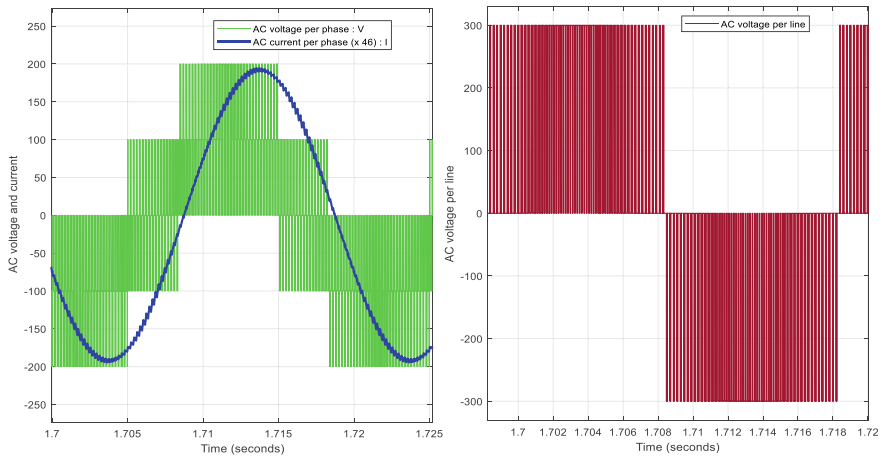
Fig. 10 AC voltage and current ($\times 46$)

Fig. 11 AC currents per phase
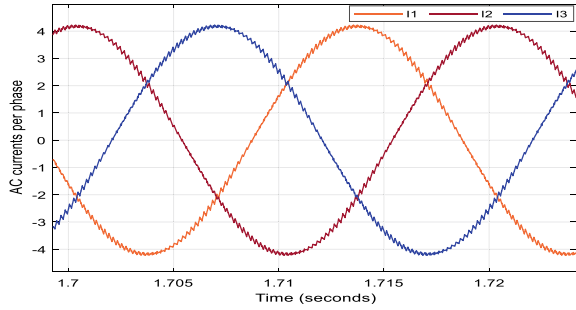
Fig. 12 Speed and torque
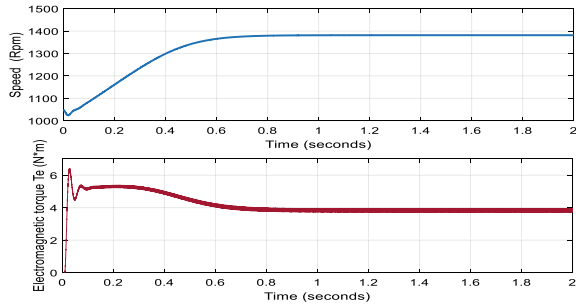
It can be seen from the results that the waveforms of the currents are sinusoidal with low THD, using just a low SPWM frequency. This proves the correctness of the simulation results and the efficiency of the used control.

4.2 Experimental Results

The experimental test bench at the laboratory is shown in Fig. 13. It consists of the Launchpad card, the Semikron inverter, DC voltage source (300 V), and a 1.5 Kw three-phase induction motor (see Table 2). The gate pulses of the inverter must be in a range of 0–15 V, while the used card generates pulses with 0–3 V. For that, an interface circuit based on transistors is adopted. The obtained results were analyzed using a C.A 8336 three-phase electrical networks analyzer (Qualistar+).

Figures 14, 15 and 16 present the recorded results with a modulation rate $r = 1$ and frequency $fc = 4$ kHz.

From the above results, the RMS value of the voltages per phase is 114 V and per line is 198 V. The RMS value of the AC currents per phase is 2.8 A, while the THD is 1.3%. This latter is very low without using filters, which leads to a normal operation, without mechanical noise, and torque ripples for the induction motor.

The experimental results are similar to the simulation ones. This proves the efficiency of the tested system.

Fig. 13 Experimental test bench

Table 2 Characteristics of three-phase IM

Power	1.5 Kw
Voltage	Δ220 V/λ 380 V
Current	Δ6 A/λ 3.5 A
Speed	1400 Rpm

Fig. 14 THD of the AC currents
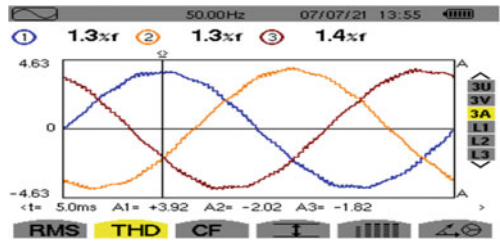
Fig. 15 AC voltage and current

Fig. 16 AC voltage per line
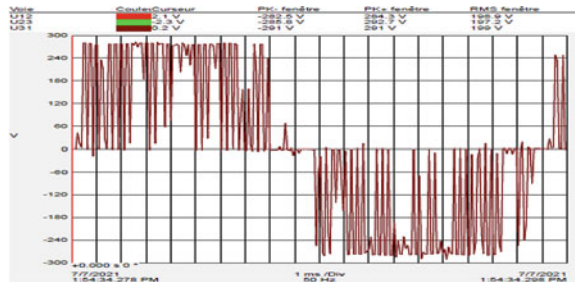
5 Conclusions

In this paper, the SPWM control technique is applied to a three-phase inverter supplying an induction motor. The SPWM was generated using the TMS320F28379D Launchpad programmed with MATLAB Simulink. The experimental results confirm the efficiency of the system to reduce THD. The future work will be focused on more performant controls.

References

1. Zhou D, Zhao J, Li Y (2016) Model-predictive control scheme of five-leg ac–dc–ac converter-fed induction motor drive. IEEE Trans Ind Electron 63(7):4517–4526
2. Xu D, Wang B, Zhang G, Wang G, Yu Y (2018) A review of sensorless control methods for AC motor drives. CES Trans Electr Mach Syst 2(1):104–115
3. Usmani AA, Shahrukh M, Mustafa A (2017) Comparison of different three phase inverter topologies: a review. In: 2017 International conference on innovations in electrical, electronics, instrumentation and media technology (ICEEIMT), pp 19–24
4. Usama M, Kim J (2020) Vector control algorithm based on different current control switching techniques for Ac motor drives. ArXiv preprint arXiv:2005.04651
5. Perach B, Weiss S (2018) SiMT-DSP: a massively multithreaded DSP architecture. In: IEEE transactions on very large scale integration (VLSI) systems, vol 26, no 8, pp 1413–1426
6. Yang C, Chen S, Zhang J, Lv Z, Wang Z (2019) A novel DSP architecture for scientific computing and deep learning. In IEEE Access 7:36413–36425
7. Hu GS (2005) Introduction to digital signal processing. Tsinghua University Press, Beijing, pp 126–135
8. Sargos F (2008) IGBT power electronics teaching system principle for sizing power converters. Application Note AN-8005, September, 2008
9. Rai T, Debre P (2016) Generalized modeling model of three phase induction motor. In: 2016 International conference on energy efficient technologies for sustainability (ICEETS), 2016, pp 927–931
10. Lokesha H, Veena S, Safiya A, Ashwin Kumar M, Basavaraju KS (2019) Evaluation of Delfino (TMS320F28379D) processor for Helmet ANC application. In: 2019 Global conference for advancement in technology (GCAT), 2019, pp 1–4
11. Boulmane A, Zidani Y, Belkhayat D (2018) Implementation of modulation techniques on TMS320F28379D launchpad using the BOOSTXL-DRV8305EVM. In: 2018 International symposium on advanced electrical and communication technologies (ISAECT), 2018, pp 1–6
12. Ismail B et al (2006) Development of a single phase SPWM microcontroller-based inverter. In: 2006 IEEE international power and energy conference. IEEE
13. Kabashi Q et al (2018) Low order harmonic analysis of 3-phase SPWM and SV-PWM inverter systems driving an unbalanced 3-phase induction motor load. Model Simul (IREMOS) 134
14. Guedouani R, Fiala B, Boucherit MS (2017) Implementation of multi-carrier PWM using a DSP TMS320F28335. Application to series multicellular single-phase inverter. In: 2017 10th International conference on electrical and electronics engineering (ELECO). IEEE

Efficient Fuzzy Logic MPPT Controls for Sudden Change in Load

Mohammed Boutaybi, Yamina Khlifi, and Hajji Bekkay

Abstract The objective of this paper is to study the performance of a photovoltaic (PV) system controlled by two types of Maximum Power Point Tracking (MPPT) commands under a rapid change in load. These commands are, respectively, perturb and observe (P&O) and improved fuzzy logic (FL) based on two types of algorithms: Mamdani and Takagi-Sugeno (TS). These improved FL commands are based on two inputs which represent the slope of the power-current curve and changes of this slope instead of the power-voltage curve slope. The designed PV system consists of a typical PV panel (1Soltech 1STH215-P) feeding a resistive load via a (DC/DC) converter where the MOSFET is controlled by PWM signal generated by the MPPT control. The simulation results obtained using MATLAB/Simulink environment show that the performance of the boost converter is very satisfactory (97%) in efficiency and that FL-Mamdani, and FL-TS MPPTs controls are more efficient than the classical P&O. Indeed, the comparison of different performance parameters, such as PV system efficiency and response time to achieve MPP, shows that the PV system efficiency obtained by fuzzy logic method is about 99.60% for FL-Mamdani, 99.64% for FL-TS and 99.20% for P&O method under the rapid change in load. In addition, the response time is shorter for FL methods (25.7 ms for FL-Mamdani and 21.9 ms for FL-TS) than in P&O method which records 31.5 ms. Furthermore, the proposed FL controls (FL-Mamdani and FL-TS) reduce the oscillations obtained by the P&O method and converge quickly to the maximum power point (MPP) regardless the load change.

Keywords Boost converter · Efficiency · Fuzzy logic · Mamdani · Takagi-Sugeno · MATLAB/Simulink · MPPT · P&O command · Photovoltaic system

M. Boutaybi (✉) · Y. Khlifi · H. Bekkay
Energy, Embedded System and Data Processing Laboratory, ENSAO, Mohamed First University, Oujda, Morocco
e-mail: boutaybi1994@gmail.com

1 Introduction

The energy of photovoltaic systems is a very important source, free and non-polluting. The photovoltaic panel has nonlinear characteristics. It varies continuously according to the atmospheric conditions and the connected load.

To study the performance of the PV system, photovoltaic models are used to determine the PV characteristics under different weather or load conditions. The photovoltaic model usually consists of a photocurrent source, diodes and resistors. The most common photovoltaic designs are the single-diode and dual-diode designs. For the MPPT methods, there are classical techniques such as perturbation and observation (P&O) [1, 2], and incremental conductance (IC) [3] controls. They present disadvantages such as a slow tracking speed and/or oscillations around the MPP and misinterpretation of the MPP location during rapid changes in weather conditions or load. Therefore, to improve the performance of PV systems and overcome these drawbacks, more robust controls and techniques have been suggested, such as fuzzy logic [4] and ANN algorithms [5] in order to increase the efficiency of PV system. For fuzzy logic control, two improved types of algorithms are proposed in this study: FL-Mamdani and FL-Takagi-Sugeno (TS) to enhance the efficiency of the PV system under rapid change in load. Most of published works [6–8] relating to fuzzy logic MPPT commands are based on two inputs error and change of error which represent the slope of the power-voltage $P(V)$ curve and changes of this slope. In the proposed study, the inputs of FL control are defined instead as the slope of the power-current P(I) curve and changes of this slope in order to achieve an efficient control of PV system under sudden change in load. In this context, our work consists in modeling, in MATLAB/Simulink environment, an autonomous PV system including a typical photovoltaic panel connected to a resistive load through a boost converter (DC/DC) whose operation is controlled by the MPPT command and operating at a switching frequency of 5 kHz. The aim of this work is to analyze the performance and evaluate the dynamic response of the PV system controlled by three types of MPPT controls. These simulations mainly concern the different electrical parameters (current, voltage and power), the efficiency of the MPPT and the response time under rapid changes in load.

2 Mathematical Model of PV Cell

The most common PV cell model is the single-diode model. This model consists of five parameters, namely the photocurrent (I_{Ph}), the saturation current of the diode (I_s), the ideality factor (a), the series resistance (R_s) and the shunt resistance (R_{sh}) [9]. This model is shown in Fig. 1.

The current generated by the photovoltaic cell is given by following relation [9]:

$$I_{\mathrm{pv}} = I_{\mathrm{ph}} - I_{\mathrm{s}}\left[\exp\left(\frac{V_{\mathrm{pv}} + I_{\mathrm{pv}}R_{\mathrm{s}}}{aV_t}\right) - 1\right] - \frac{V_{\mathrm{pv}} + I_{\mathrm{pv}}R_{\mathrm{s}}}{R_{\mathrm{sh}}} \tag{1}$$
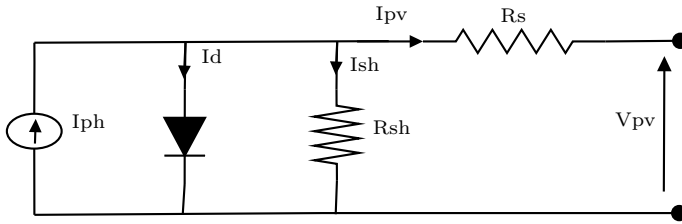
Fig. 1 Equivalent circuit of single-diode model for PV cell

Table 1 Electrical characteristics of 1Soltech 1STH-215-P PV panel under STC

Parameter-variable	Value	Parameter-variable	Value
Number of cell (N_{cell})	60	Rated power	213.15 (W)
Short-circuit current (I_{sc})	7.84 (A)	Optimum current (I_m)	7.35 (A)
Open-circuit voltage (V_{oc})	36.3 (V)	Optimum voltage (V_m)	29 (V)
T coefficient of I_{sc} (K_i)	0.102(%°C)	Shunt resistance (R_{sh})	313,3991 (Ω)
T coefficient of V_{oc} (K_v)	$-360,099$ (%°C)	Series resistance (R_s)	0.39383 (Ω)

where $V_t = \frac{K.T}{q}$ the thermal voltage, T: the PV cell temperature in K, q: Charge of the electron (1.6×10^{-19} C), K: Boltzmann Constant (1.38×10^{-34} J/K), a: Diode ideality factor ($a = 0.98117$).

The essential parameters, required for PV panels modeling, are given in Table 1 under standard test conditions (STC) where $T = 25\,°C$ and $G = 1000\,W/m^2$.

3 Photovoltaic System

MATLAB/Simulink software is used to model, simulate the whole PV system under variable load and fixed temperature and irradiation level values. The crucial problem of the PV system is to design an adaptation approach (MPPT control) that allows the automatic MPP tracking and fast convergence to the optimal conditions regardless of sudden changes in load. Figure 2 shows the synoptic diagram of a photovoltaic system feeding a resistive load.

The DC/DC converter included in the PV system is of boost type where its various components (inductance L, the input and output capacitors: Ce and Cs) are sized by considering the following conditions: the converter provides an adaptable voltage to the chosen value (46 V) from the PV panel, with residual ripple of 1% in the input and output voltage. The converter operates in continuous conduction at a switching frequency of 5 KHz. The outputs voltage and current (V_{pv}, I_{pv}) of PV panel are linked to the outputs (V_o and I_o) of the boost converter by the following relationships [10, 11]:
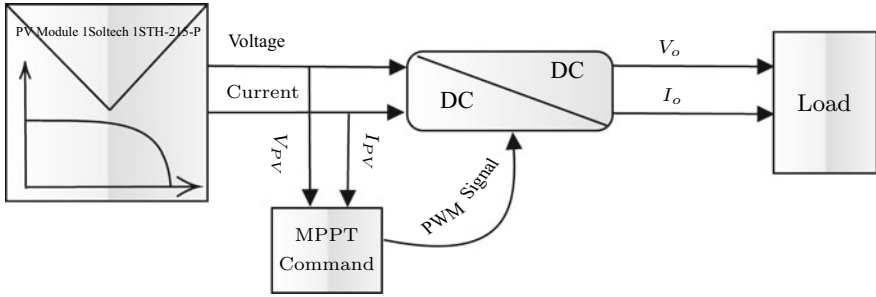
Fig. 2 Synoptic diagram of the photovoltaic system whose operation is controlled by the MPPT command

Fig. 3 Typical plot of power-voltage characteristics P_{pv} (V_{pv}) at STC for PV panel and different values of load
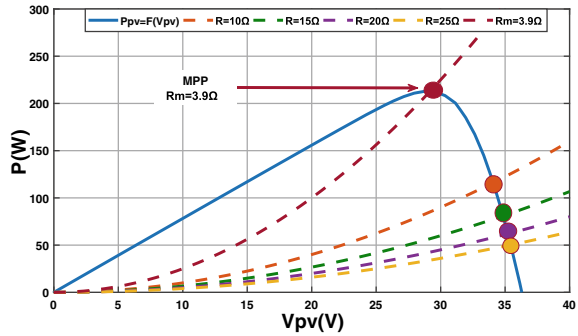
$$V_o = \frac{V_{PV}}{1 - D} \quad I_o = (1 - D)I_{PV} \tag{2}$$

where: D is the Duty cycle.

Figure 3 shows the typical plots of power-voltage characteristics at STC for both the PV panel and different resistive load values.

4 Proposed Maximum Power Point Tracking Commands

Two types of MPP tracking controls are studied in this work: the P&O algorithm and fuzzy logic method of Mamdani and Takagi-Sugeno types. Each of these techniques has its own advantages and disadvantages in terms of simplicity, efficiency, robustness and response time to reach the MPP.

Fig. 4 Organigramme
"P&O" algorithme
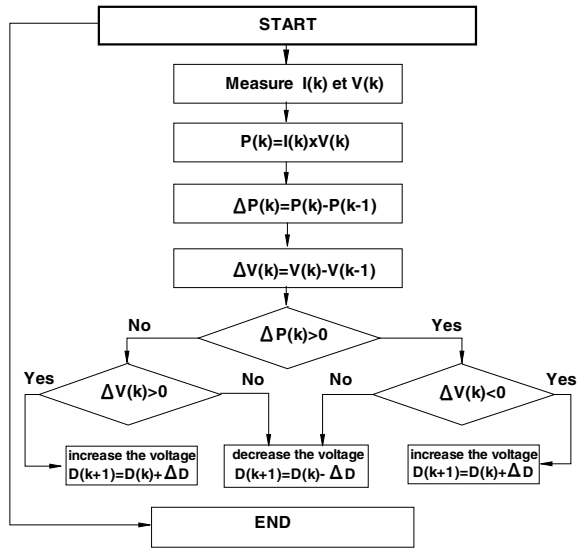
4.1 Perturb and Observe (P&O) MPPT Control

The Perturb and Observe method is one of the most popular and widely used classical algorithms. The principle of this method is to perturb the voltage of the photovoltaic panel and observe its effect on the power variation. Figure 4 presents the flowchart of this algorithm [12, 13].

4.2 Fuzzy Logic MPPT Control

Fuzzy logic control has been used in PV systems to track the MPP [14]. This method consists of three essential steps: fuzzification, inference procedure and defuzzification as shown in Fig. 5.
Where:
E(k): is the error at simpled time k. It is given by:

$$E(k) = \frac{P(k) - P(k-1)}{I(k) - I(k-1)} \tag{3}$$

where: $P(k)$ and $I(k)$ are the power and current of the photovoltaic panel, respectively.

CE(k): is the variation of error defined as follow:

$$CE(k) = E(k) - E(k-1) \tag{4}$$
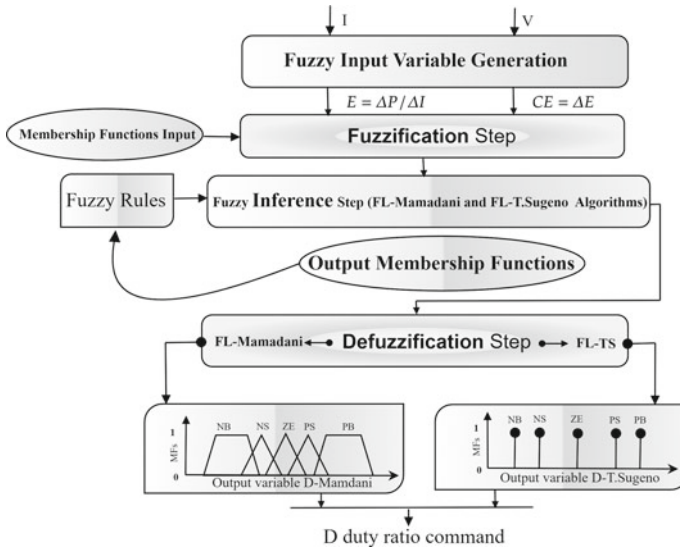
Fig. 5 Flowchart of fuzzy controller

Fig. 6 Fuzzy member functions (MFs) used in fuzzification process
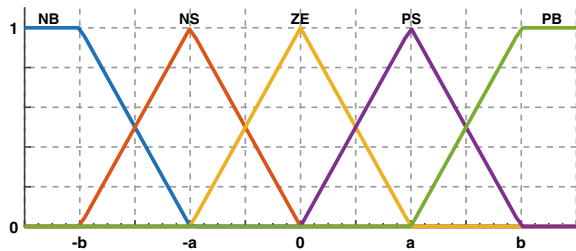
Fuzzification step consists to transform the input variables (E, CE) into linguistic variables or fuzzy variables [15]: They are expressed in terms of five linguistic variables which are Negative Big (NB), Negative Small (NS), Zero (ZE), Positive Small (PS) and Positive Big (PB). Their typical member functions are presented in Fig. 6.

In the **inference step**, decisions are made. Indeed, we establish logical relations (rules) between the inputs E and CE and the output D as shown in the inference Table 2. Figure 7 shows the member functions surface for FL-Mamdani and FL-TS MPPT commands.

The **defuzzification step** concerns the conversion of the output fuzzy sets into a numerical value. This defuzzification step depends on the type of FL command: In the case of FL-Mamdani method, it is based on the center of gravity which is the most widely used method. However, in the case of the FL-TS method, it uses a weighted average method to determine the output D. This output is not in the form of a fuzzy set, but is instead a constant or a linear equation.

Table 2 Fuzzy rules relating the inputs and output variables

E	CE				
	NB	NS	ZE	PS	PB
NB	ZE	ZE	NB	NB	NB
NP	ZE	ZE	NS	NS	NS
ZE	NS	ZE	ZE	ZE	PS
PS	PS	PS	PS	ZE	ZE
PB	PB	PB	PB	ZE	ZE
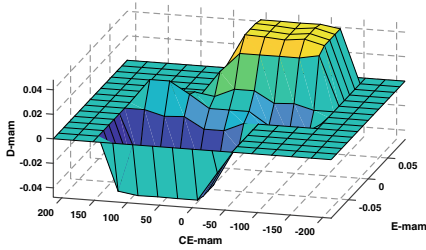
MFs surface view of FL-Mamdani MPPT   MFs surface view of FL-TS MPPT
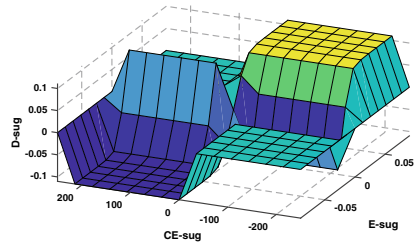
Fig. 7 MFs surface view of FL-Mamdani and FL-TS MPPT controls

5 Simulation Results and Discussions

5.1 Effect of Rapid Change in Load

The PV system is simulated in the MATLAB/Simulink environment under STC and a rapid change in load (R) as shown in Fig. 8a.

Figures 8b and 9 show the simulation results, at STC, of voltages, currents and powers values at the input and output of the boost converter (DC/DC) as a function of time when the load varies according to Fig. 8a.
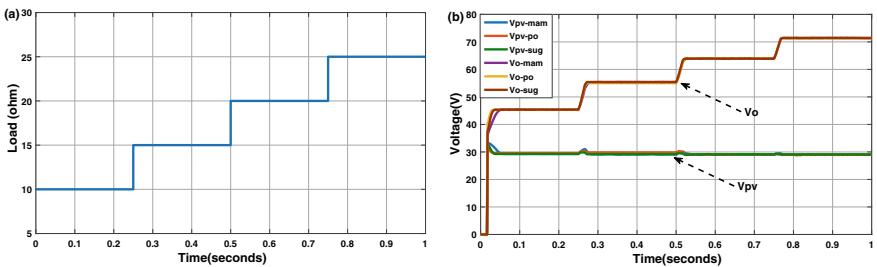
Fig. 8 **a** Waveform of rapid changes in load (R) and **b** input and output voltages (V_{pv}, V_o) of boost converter controlled by (PO and FL) MPPT commands according to the fast change of load
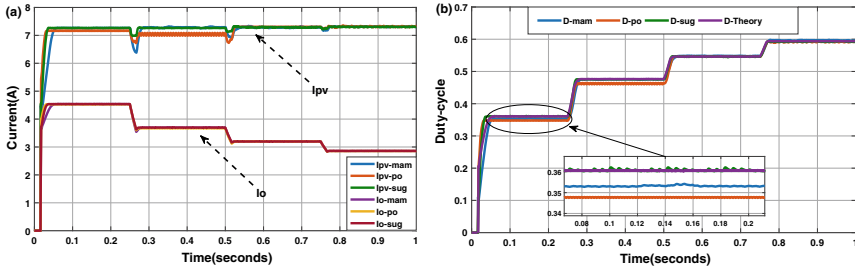
Fig. 9 **a** Input and output currents (I_{pv}, I_o) of boost converter, **b** duty cycle, according to the fast changes of load under STC
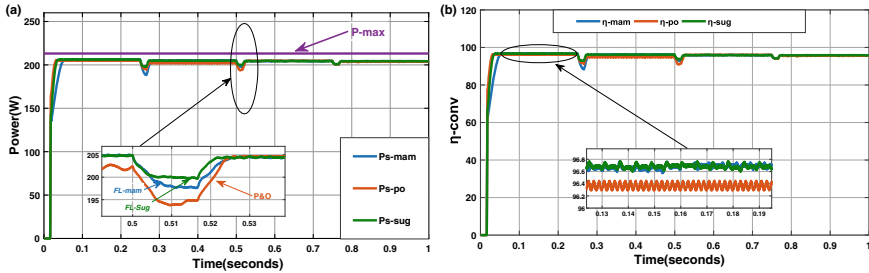
Fig. 10 Waveforms of: **a** calculated maximal power (P_{max}), input (P_{PV}) and output power (P_o). **b** Boost efficiency (η_{Conv}), according to the fast changes of load

It should be noted that the converter plays its roles correctly: increasing voltage, decreasing current as shown in Figs. 8b and 9a and duty cycle value is adjusted to the adequate value corresponding to MPP, with high accuracy for FL-TS method as shown in Fig. 9b. On the other hand, as shown in Fig. 10a, the maximum power delivered by PV panel is almost transferred to the load using either PO or FL commands.

Furthermore, the proposed FL controls reduce the oscillations obtained by the P&O method and converge rapidly to the maximum power point (MPP). It should be noted that the boost efficiency is very satisfactory. Its value is in order of 97% regardless the load changes as shown in Fig. 10b.

5.2 Validation and Evaluation of MPPT Controls

The efficiency values of the MPPT commands are defined as [16]:

$$\eta_{MPPT} = \frac{P_{PV}}{P_m} \tag{5}$$

Fig. 11 MPPT efficiency (η_{MPPT}) according to the load changes
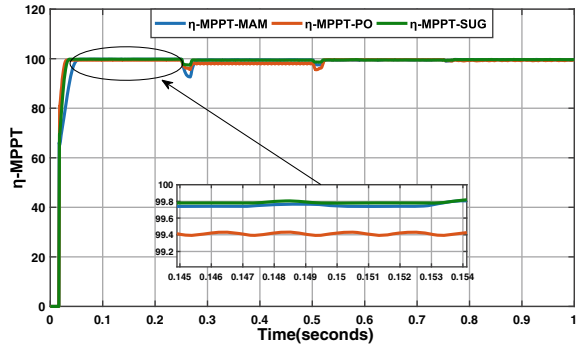
Table 3 Comparison of the performance of P&O and FL MPPT commands

MPPT methods	Response time (ms)	Efficiency (%)
FL-Mamdani	25.7	99.60
FL-TS	21.9	99.64
P&O	31.5	99.20

The obtained results in Fig. 11 show that the fuzzy logic command based on Mamdani or TS methods exceeds the classical P&O control in terms of response time and efficiency. Indeed, as shown in Table 3, very satisfactory efficiency value of 99.64% and response time of 21.9 ms are obtained for FL-TS.

6   Conclusion

In this work, a comparative study between the conventional P&O and intelligent MPPT algorithms such as FL-Mamdani and FL-Takagi-Sugeno is presented. These improved FL commands are based on two inputs which represent the slope of the power-current curve and changes of this slope instead of the power-voltage curve slope.

This study is based on the comparison of different performance parameters, such as: PV system MPPT efficiency and response time to achieve maximum power point under rapid changes in load. Among the evaluated methods, the intelligent techniques FL-Mamdani and FL-TS have a good performance compared to P&O command. Furthermore, FL-TS command is simple, show reduced oscillations around MPP and records the highest MPPT efficiency (99.64%) and the lowest response time 21.9 ms compared to other commands. Thus, the intelligent FL-TS MPPT command can be considered as an efficace solution to control PV systems under rapid changes in load.

References

1. El Motahhir S, El Hammoumi A, Ghzizal A (2018) Photovoltaic system with quantitative comparative between an improved MPPT and existing INC and P&O methods under fast varying of solar irradiation. Energy Rep 4:341–350. https://doi.org/10.1016/j.egyr.2018.04.003

2. Gupta AK, Chauhan YK, Maity T (2018) Experimental investigations and comparison of various MPPT techniques for photovoltaic system. Indian Acad Sci 43:132. https://doi.org/10.1007/s12046-018-0815-0

3. Singh K, Anand A, Mishra AK, Singh B, Sahay K (2021) SEPIC converter for solar PV array fed battery charging in DC homes. J Inst Eng (India) 102:455–463. https://doi.org/10.1007/s40031-020-00522-0

4. Guenounou O, Boutaib D, Chabour F (2014) Adaptive fuzzy controller based MPPT for photovoltaic systems. J Energy Convers Manage 78:843–850. https://doi.org/10.1016/j.enconman.2013.07.093

5. Suganthi L, Iniyan S, Samuel AA (2015) Applications of fuzzy logic in renewable energy systems. J Renew Sustain Energy Rev 48:585–607. https://doi.org/10.1016/j.rser.2015.04.037

6. Radjai T, Gaubert J-P, Rahmani L (2014) The new FLC-variable incremental conductance MPPT with direct control method using Cuk converter. In: 2014 IEEE 23rd international symposium on industrial electronics (ISIE), pp 2508–2513. https://doi.org/10.1109/ISIE.2014.6865014

7. Ait Cheikh M-S, Larbes C, Tchoketch Kebir G-F, Zerguerras A (2007) Maximum power point tracking using a fuzzy logic control scheme. Rev Energ Renouvelables 10:387–395

8. Rahmani R, Seyedmahmoudian M, Mekhilef S, Yusof R (2013) Implementation of fuzzy logic maximum power point tracking controller for photovoltaic system. Am J Appl Sci 3:209–218. https://doi.org/10.3844/ajassp.2013.209.218

9. Aicha D, Rezaoui M, Teta A, Boudiaf M (2019) The MPPT command for a PV system comparative study: control based on fuzzy logic with "P&O". Renew Energy Smart Sustain Cities 62:346–354. https://doi.org/10.1016/j.rser.2015.04.037

10. Khlifi Y, Aziz A (2018) Efficient energy utilization of stand alone photovoltaic system under rapid change in load. AIP Conf Proc 2056:SP-020026. https://doi.org/10.1063/1.5084999

11. Yilmaz U, Kircay A, Borekci S (2018) PV system fuzzy logic MPPT method and PI control as a charge controller. Renew Sustain Energy Rev 81:994–1001. https://doi.org/10.1016/j.rser.2017.08.048

12. Othmani H, Chaouali H, Mezghani D, Mami A (2015) Optimisation de la technique de perturbation et observation par la logique floue. In: 3ème confèrence internationale des ènergies renouvelables. CIER-2015. Int J Sci Res Eng Technol (IJSET) 4:140–144

13. Bounechba H, Bouzid A, Nabti K, Benalla H (2014) Comparison of perturb and observe and fuzzy logic in maximum power point tracker for PV systems. J Energy Proc 50:667–684. https://doi.org/10.1016/j.egypro.2014.06.083

14. Kiswantono A, Prasetyo E, Amirullah A (2019) Comparative performance of mitigation voltage sag/swell and harmonics using DVR-BES-PV system with MPPT-fuzzy Mamdani/MPPT-fuzzy Sugeno. J Int J Intell Eng Syst 12:222–235. https://doi.org/10.22266/ijies2019.0430.22

15. Bendib B, Belmili H, Krim F (2015) A survey of the most used MPPT methods: conventional and advanced algorithms applied for photovoltaic systems. J Renew Sustain Energy Rev 45:637–648. https://doi.org/10.1016/j.rser.2015.02.009

16. Khlifi Y, Hajji B, Messaoudi A (2019) A new maximum power point tracking PV control for rapid changes in irradiation level. In: Proceedings of the 1st international conference on electronic engineering and renewable energy, vol 512, pp 384–391. https://doi.org/10.1007/978-981-13-1405-6-46

A Nonlinear Control of Energy Storage System-PV-Based Stand-Alone DC-Microgrid

B. K. Oubbati , **Abdelhamid Rabhi, S. Benzaouia, M. Boutoubat, M. Belkheiri, and Y. Oubbati**

Abstract   In this paper, a nonlinear integral sliding mode control (ISMC) for Hybrid Energy Storage System (HESS)-based stand-alone DC-microgrid has been proposed. This hybrid system consists of a PV system, supercapacitor, and battery. A classical PI-based linear control strategy has been considered to control battery and supercapacitor systems, based on the decoupling of high- and low-frequency component to estimate reference current. Due to frequent discharge during operation, supercapacitor power can reach the lowest value, which affects controller performances and even makes the system unstable. Thereby, a nonlinear ISMC is performed as an inner loop controller to regulate the battery and supercapacitor powers. Also, the PI controller is implemented as an outer loop controller to regulate the DC-link voltage. As a result, the proposed control strategy has achieved high dynamic system performances.

Keywords Integral sliding mode control (ISMC) · Hybrid energy storage system (HESS) · PV · DC-microgrid

1 Introduction

Currently, renewable energy sources are becoming a very attractive solution to guarantee our daily load needs. Photovoltaic systems are widely used recently to produce

B. K. Oubbati (✉) · M. Belkheiri
LTSS Laboratory, Laghouat University, Laghouat, Algeria
e-mail: i.oubbati@lagh-univ.dz

B. K. Oubbati · A. Rabhi
MIS Laboratory, Picardie Jules Verne University, Amiens, France

S. Benzaouia
SASV/LIS, Aix Marseille University, Marseille, France

M. Boutoubat · Y. Oubbati
LACoSERE Laboratory, Laghouat University, Laghouat, Algeria

electricity. Their use is increased for many reasons such as the decreasing of the classical energy sources (gas and oil …), their abundant availability, and their eco-friendly aspect.

The production of power from the PV systems depends strongly on the quantity of solar irradiations. Also, due to the nonlinearity behavior of the PV system, the maximum power point tracking controller is necessary to push the PV system to produce maximum power [1–4]. However, since the load demand and PV system have variable behavior, the PV system is not able alone to satiate load demand. Hence, the battery storage system seems to offer a promising opportunity to alleviate the issues of load demand.

Since the load demand and PV system have variable behavior, the PV system is not able alone to satisfy load demand. Hence, the battery storage system seems to offer a promising opportunity to alleviate the issues of load demand. Batteries are commonly implemented in a stand-alone PV power system to fulfill the power mismatch between the load and PV system production. Moreover, due to the changing output of the PV system and the intermittent high power demand of the load, batteries are becoming encounter frequent deep cycles and irregular charging [5]. These drawbacks would increase the replacement cost of the battery and shorten the battery life span. Supercapacitor-battery Hybrid Energy Storage System (HESS) is thus a practical solution to reduce the capital cost of the battery, battery stress, and sizing. The advantage of using the HESS with a stand-alone PV power system would be able to provide both power capacities and high energy to solve situations such as variation of the load and weather changing conditions (irradiation and temperature). So, a control strategy is necessary for HESS to manage energy sustainability to the maximum extent as it is the control structure that manages the power flow of the battery and supercapacitor (SC) and the energy utilization [6, 7].

In the literature, many control strategies have been proposed for this kind of HESS for remote area power system (RAPS) consisting of supercapacitor-battery and PV power systems. In [8], the authors proposed an optimal control structure for HESS with a PV system. This control strategy is fully achieved using a low-pass filter and a Fuzzy Logic Controller (FLC). In addition, the membership functions of the FLC are optimized by Particle Swarm Optimization (PSO) to achieve optimal battery peak current reduction. But, the implementation complexity of this method is quite high.

In [9], a comprehensive study of HESS for stand-alone PV system has been presented to review the state of the art of HESS and discusses potential topologies that are suitable for improving the life span of the battery. Moreover, a control strategy for battery-supercapacitor with a stand-alone PV power system is proposed [10]. This control strategy consisting of PI-based linear control has been proposed to enhance the battery life span and battery stress minimization. The obtained results have shown low performances. Another nonlinear control strategy has been used for HESS with PV power system-based DC-microgrid [11], in order to overcome the limitation of a linear PI controller.

In this paper, an integral sliding mode control (ISMC)-based nonlinear control structure is proposed for battery-supercapacitor with PV power system-based stand-alone DC-microgrid. This control structure is based on decoupling of low and high

frequency. The performances of this ISMC-based nonlinear control strategy have been tested under fast change variations of the weather conditions in terms of reducing the battery stress and improving the battery life span. Moreover, the main advantage of this proposed control is the robustness and system stability.

2 Control Strategy Based on Integral Sliding Mode Controller

Figure 1 shows the global control system. The main objective of the ISMC is to minimize the stress and increase the lifetime of battery simultaneously. As can be seen from Fig. 1, the DC-bus voltage (V_{dc}) is regulated to its reference value (V_{ref}), through a proportional integral (PI) controller. This controller generates the necessary total current reference (I_{t-ref}) to control the DC-link voltage. Then, $I_{tot-ref}$ is separated into low-frequency component (I_{LFC_ref}) and high-frequency component (I_{HFC_ref}) as writing hereafter:

$$I_{LFC\_ref} = \text{lowpass filter } (I_{t-ref}) \tag{1}$$

$$I_{HFC\_ref} = I_{t-ref} - I_{LFC\_ref} \tag{2}$$

The low-frequency component gives the reference current of battery as given below:

$$I_{bat\_ref} = I_{LFC\_ref} \tag{3}$$
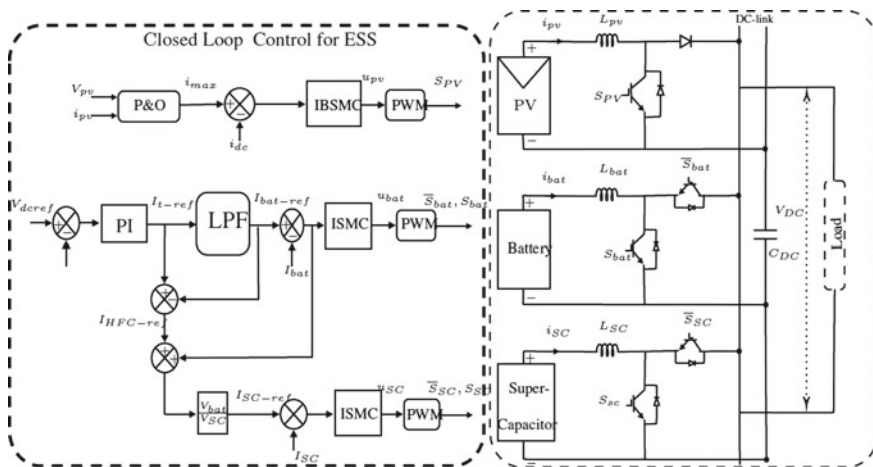
Fig. 1 Structure of stand-alone DC-microgrid with proposed control strategy

The battery current (I_{bat}) is controlled to follow its reference value by using the integral sliding mode controller. Then, the ISMC controller generates the control signal (u) for the bidirectional DC-DC converter. This command is given to the PWM generator to generate switch control corresponding to DC-DC converter interfaced with battery (see Fig. 1). The uncompensated battery power is written as follows:

$$P_{\text{bat\_uc}} = \left(I_{\text{HFC\_ref}} + I_{\text{bat\_err}}\right) V_{\text{bat}} \tag{4}$$

This uncompensated battery power is to be compensated by SC. Consequently, the reference current of SC is given as:

$$I_{\text{SC\_ref}} = \frac{P_{\text{bat\_uc}}}{V_{\text{SC}}} = \left(I_{\text{HFC\_ref}} + I_{\text{bat\_err}}\right) \frac{V_{\text{bat}}}{V_{\text{SC}}} \tag{5}$$

According to Fig. 1, I_{sc} ref is compared with the actual SC current (I_{sc}), and the error is given to the ISMC controller. The ISMC controller generates the command (u). These duty ratios are given to the PWM generator to generate switching pulses corresponding to SC switches. Finally, to push the PV system to produce maximum power under fast changing irradiation, we used nonlinear controller (integral backstepping sliding mode) developed previously [1].

2.1 Integral Sliding Mode Controller

For battery control, the average model of the bidirectional DC-DC converter is written as follows:

$$\begin{cases} \frac{d}{dt} i_L = \frac{1}{L}[v_{\text{bat}} - (1-u)v_{\text{bus}}] \\ \frac{d}{dt} v_{\text{bus}} = \frac{1}{C_{\text{bus}}}[(1-u)i_L - i_b] \end{cases} \tag{6}$$

To develop the ISMC, we consider the sliding surface as follows:

$$S = e + k_s \int e \, dt \tag{7}$$

where e is the error between the actual I current and the reference value I_{ref} and is given as:

$$e = I_{\text{ref}} - I \tag{8}$$

In order to force the sliding surface to be zero. So, the derivation of sliding surface is given as:

$$\dot{S} = \dot{e} + k_s e = \dot{I}_{\text{ref}} - \frac{1}{L}(V_{\text{bat}} - (1 - u)V_{\text{dc}}) + k_s(I_{\text{ref}} - I) \tag{9}$$

The equivalent control of integral sliding mode is given as:

$$u_{\text{eq}} = 1 - \frac{L}{V_{\text{dc}}}\left(V_{\text{bat}} - \dot{I}_{\text{ref}} - k_s(I_{\text{ref}} - I)\right) \tag{10}$$

The classical control law of integral sliding mode controller is given as:

$$u = u_{\text{eq}} + u_{\text{dis}} \tag{11}$$

Hence, the duty cycle for controlling the bidirectional DC/DC converter for battery system is given as:

$$u = 1 - \frac{L}{V_{\text{dc}}}\left(V_{\text{bat}} - \dot{I}_{\text{ref}} - k_s(I_{\text{ref}} - I)\right) - k_1 \text{sign}(S) \tag{12}$$

Note that, the same steps are used to design the ISMC for the supercapacitor current control and for supercapacitor system:

$$u = 1 - \frac{L}{V_{\text{dc}}}\left(V_{\text{SC}} - \dot{I}_{\text{ref}} - k_{s-\text{SC}}(I_{\text{ref}} - I)\right) - k_{1-\text{SC}} \text{sign}(S) \tag{13}$$

Stability System Proof

The Lyapunov criteria have been considered in this work, to prove the stability of the hybrid system. So, the candidate of the Lyapunov function is taken as follows:

$$V = \frac{1}{2}S^2 \tag{14}$$

By deriving this equation, we obtain:

$$\dot{V} = S\dot{S} < 0 \tag{15}$$

The system stability is fully guaranteed if the derivative Lyapunov function is always negative. From Eqs. 9 and 15, one can write:

$$\dot{V} = S\dot{S}$$
$$= \left(e + k_s \int e.\mathrm{d}t\right)\left(\dot{I}_{\text{ref}} - \frac{1}{L}(V_{\text{bat}} - (1 - u)V_{\text{dc}})\ldots \atop +k_s(I_{\text{ref}} - I)\right) \tag{16}$$

replacing Eq. 12 into Eq. 16

$$\dot{V} = S\left(\begin{array}{c} \dot{I}_{\text{ref}} - \frac{1}{L}\left(\left(\frac{L}{V_{\text{dc}}}\left(V_{\text{bat}} - \dot{I}_{ref} - k_s(e)\right)\right)V_{\text{dc}}\right) \\ +k_s(e) - \frac{V_{\text{dc}}}{L}k_1\text{sign}(S) - \frac{V_{\text{bat}}}{L} \end{array}\right) \tag{17}$$

The simplification of last equation leads to:

$$\dot{V} = S\left(-\frac{k_1\text{sign}(S)}{L}\right) = -\frac{k_1}{L}|S| < 0 \tag{15}$$

where k_1 is a positive value. So, the system is always stable. The same steps are used to proof the stability of the supercapacitor system.

3   Simulation Results

The nonlinear controller (ISMC) has been evaluated and tested under changing weather conditions (irradiations). The aim is to satisfy the load demand. The HESS parameters are presented in Table 1. Moreover, the initial state of charge (SoC) of the battery is taken equal to 60% (Table 2).

From Figs. 2, 3, 4, 5, 6 and 7, one can remark that:

- For [0 s < t < 1 and 1.5 s < t < 2 s], the produced MPPT PV power is greater than the load demand, so the battery intervenes to store the power surplus. In case of constant irradiation, the PV produces a power of 980 W.
- For [0 s < t < 1 s], the PV-produced power is lower than the load demand, and the battery operates as generator and supplies the pour deficit to the load (discharge mode). At the beginning (t = 0), the battery has absorbed an important peak of current and power which can cause its damage (see Figs. 3a and 4). In order to

Table 1 Parameters of DC-link

Energy sources and converters	Parameters
PV system one module Energies WU-120	MPP = 120.7 W, V_{oc} = 21, I_{sc} = 8 A, V_{max} = 17 V, I_{max} = 7.1A, array data (parallel = 4 and series = 2)
Battery	Lithium-Ion; V_n = 24 V, I_c = 14A, initial SOC = 60%
SC	V = 40 V, C = 29f
Converters and DC-link	L_{pv} = 0.352 mH, L_{bat} = 0.3mH, L_{sc} = 0.36 mh C = 320 uF, Vdcref = 50 V

Table 2 Parameters of controllers

Controllers	Gains
ISMC	$k_{\text{s\_bat}}$ = 400, k_{1_bat} = 0.01, $k_{\text{s\_SC}}$ = 2430, k_{1_SC} = 0.01
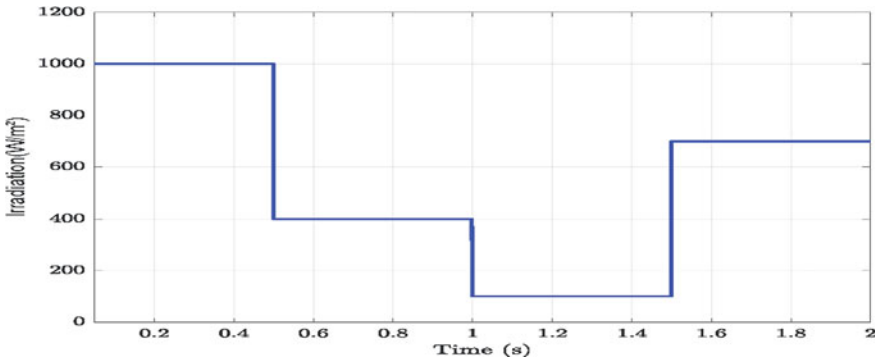
Fig. 2 Irradiation curve (W/m^2)
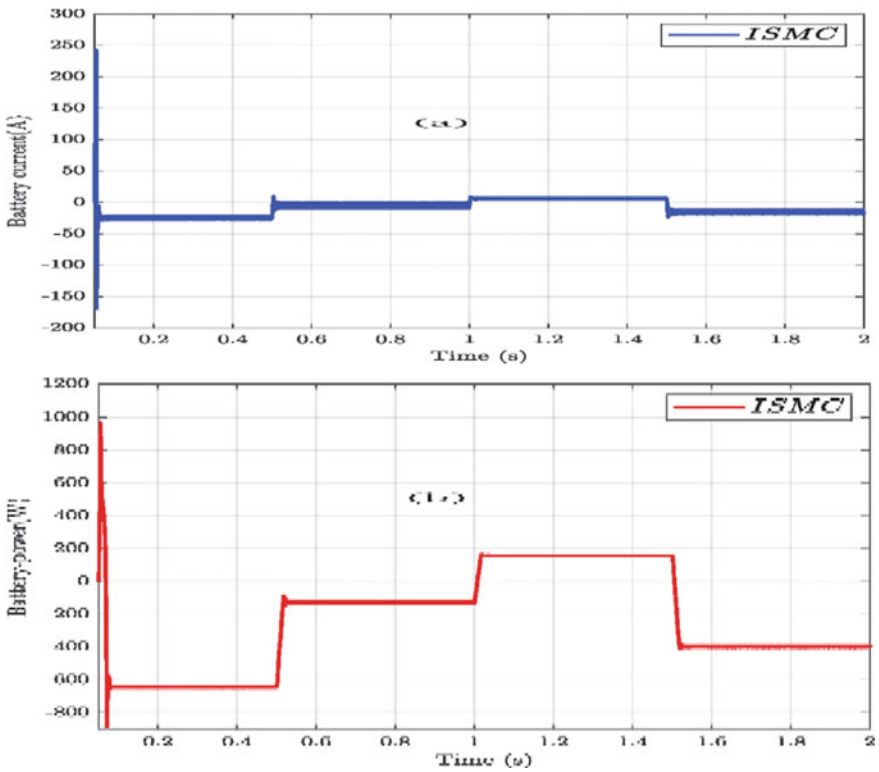
Fig. 3 Battery current and power response in case without adding supercapacitor
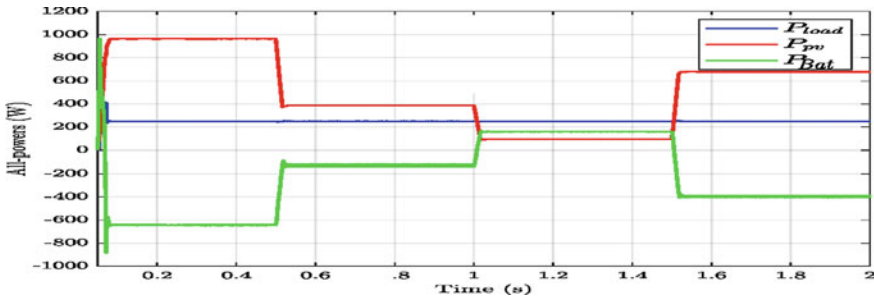
Fig. 4 All-power responses of the hybrid system without adding the supercapacitor

extend life span of the battery, a supercapacitor system is added to the hybrid system whose principle role is to absorb the peak current caused by the load and increase consequently the life span of the battery. SC absorbs a high-frequency component of excess power supply temporarily (see Fig. 5a, b)), owing to slow dynamics of battery, and battery current converges to the steady state slowly. Also, Fig. 5c shows a slow incrementation in the % SoC of the battery.

In order to evaluate the performances of these controllers (ISMC, IBSMC), the output powers are shown in Figs. 3 and 7. They show that the ISM controller is able to ensure the required power to load and IBSMC achieved the MPPT for PV system. Figure 5c shows increase or decrease in %SoC of battery according to the functionality mode of battery (discharging or charging mode). Also, using of the ISM controller has led to reduce the responses overshoot and the adding of SC has permitted the battery to operate with a slow dynamic current, and consequently, its life span is extended.

4   Conclusions

As we know, the ESS has nonlinear dynamics. So, it is more suitable to use a nonlinear control strategy to improve the life span of the battery, minimize the stress on the battery, and enhance the stability of the hybrid system. In this paper, an integral sliding mode ISMC-based nonlinear control strategy is proposed for a remote area power system consisting of ESS that is associated with a PV system. This control structure is based on high-frequency power and low decoupling, and it can regulate the DC-link at a constant value. The performances of this strategy control have been tested under the fast change irradiation. The obtained simulation results demonstrate an improved battery life span and battery stress minimization, using the proposed strategy control.
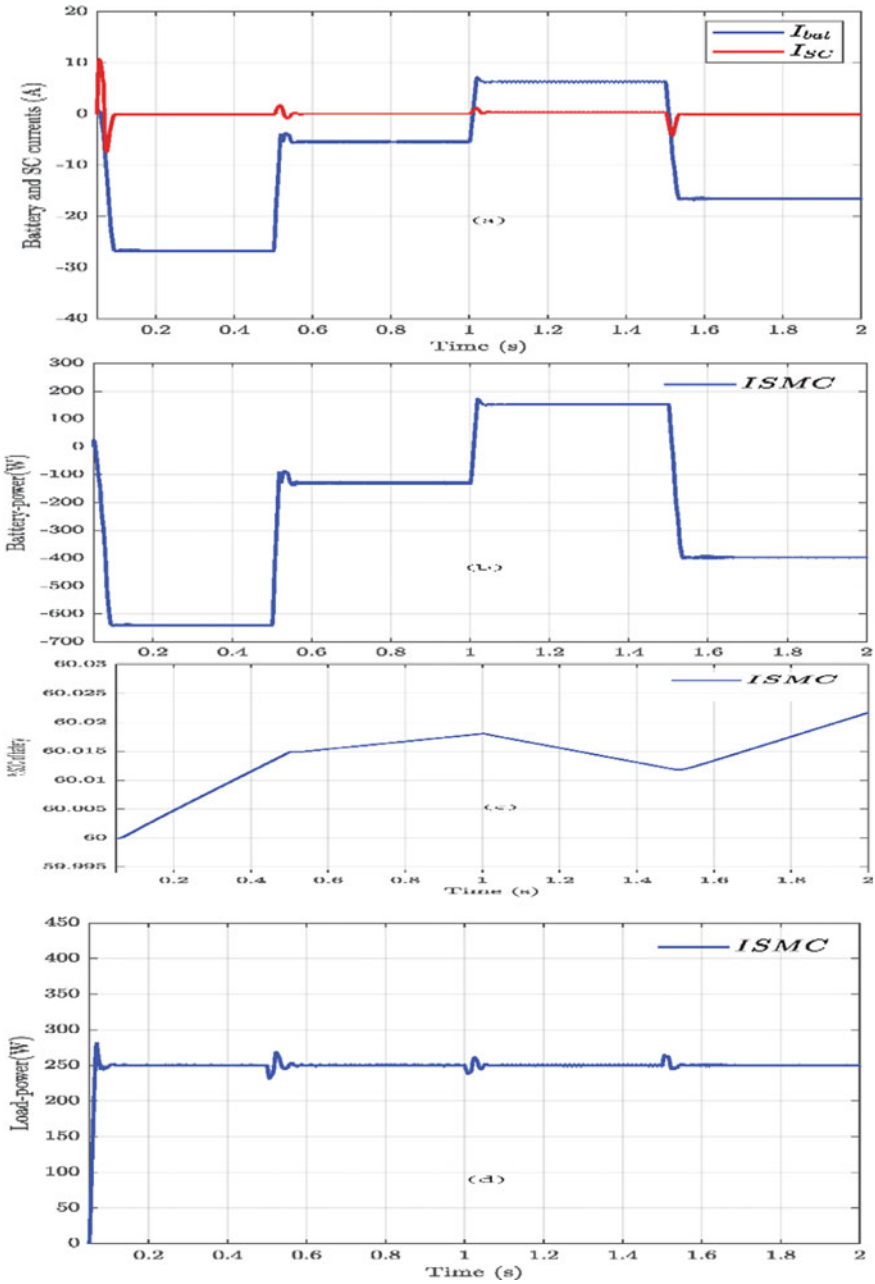
Fig. 5 Different responses of the hybrid system in the case of adding the SC: **a** battery and SC current (A), **b** battery power (W), **c** SoC of battery (%), and **d** load power (W)
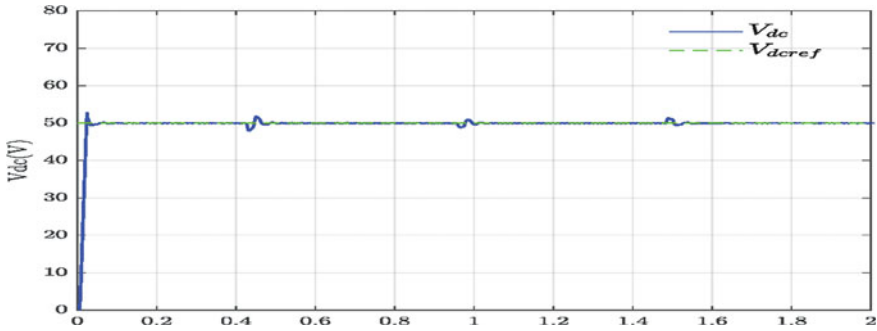
Fig. 6 DC voltage (V) of the hybrid system
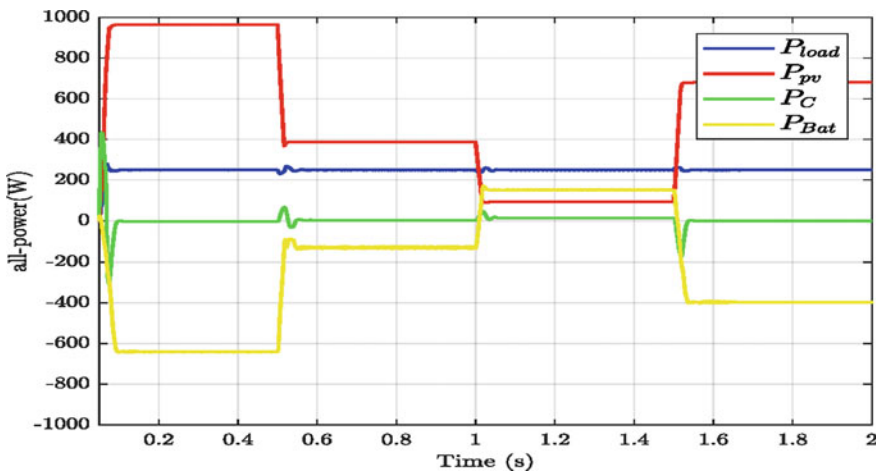
Fig. 7 All-power responses of the hybrid system controller in the case of adding the SC

References

1. Oubbati BK, Boutoubat M, Rabhi A, Belkheiri M (2020) Experiential integral backstepping sliding mode controller to achieve the maximum power point of a PV system. Control Eng Pract 102:104570. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S09670661 20301581
2. Oubbati BK, Boutoubat M, Belkheiri M, Rabhi A (2018) Global maximum power point tracking of a PV system MPPT control under partial shading. In: 2018 International conference on electrical sciences and technologies in Maghreb (CISTEM). IEEE, pp 1–6
3. Kihal A, Krim F, Laib A, Talbi B, Afghoul H (2019) An improved MPPT scheme employing adaptive integral derivative sliding mode control for photovoltaic systems under fast irradiation changes. ISA Trans 87:297–306. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0019057818304555
4. Yang B, Yu T, Shu H, Zhu D, An N, Sang Y, Jiang L (2018) Perturbation observer based fractional-order slidingmode controller for MPPT of grid-connected PV inverters: design and

real-time implementation. Control Eng Pract 79:105–125. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S096706611830279X

5. Jaszczur M, Hassan Q (2000) An optimisation and sizing of photovoltaic system with supercapacitor for improving self-consumption. 279:115776. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306261920312617

6. Kotra S, Mishra MK (2017) A supervisory power management system for a hybrid microgrid with HESS. 64(5):3640–3649. [Online]. Available: http://ieeeplore.ieee.org/document/7815391/

7. Manandhar U, Ukil A, Gooi HB, Tummuru NR, Kollimalla SK, Wang B, Chaudhari K (2017) Energy management and control for grid connected hybrid energy storage system under different operating modes. 10(2):1626–1636. [Online]. Available: https://ieeexplore.ieee.org/document/8110707/

8. Chong LW, Wong YW, Rajkumar RK, Isa D (2016) An optimal control strategy for standalone PV system with battery-supercapacitor hybrid energy storage system. 331:553–565. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S037877531631223X

9. Jing W, Lai CH, Wong WS, Wong MD (2018) A comprehensive study of battery-supercapacitor hybrid energy storage system for standalone PV power system in rural electrification. 224:340–356. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S030626191830672X

10. Kollimalla SK, Mishra MK et al (2014) A new control strategy for interfacing battery supercapacitor storage systems for PV system. In: 2014 IEEE Students' conference on electrical, electronics and computer science. IEEE, pp 1–6

11. Ravada BR, Tummuru NR (2020) Control of a supercapacitor-battery-PV based stand-alone DC-microgrid. 35(3):1268–1277. [Online]. Available: https://ieeexplore.ieee.org/document/9044297/

Real-Time Control for PMSG System Without Mechanical Sensor

Badreddine Lahfaoui

Abstract The demand for energy is increasing rapidly due to the population growth and development that the state experienced in their various sectors. Environmental concerns and the need for new sources of electricity generation have arisen in recent years. So, it is necessary to innovate renewable energy sources to reduce the environmental impacts on the one hand and to meet the demand for electrical energy on the other. In order to extract the maximum power from our wind turbine, the MPPT algorithm is put on a dSPACE card. The input of our system is a small horizontal axis wind turbine while the output is a resistive load. The objective of this work is to track the maximum power of wind turbine system for a constant value of wind speed and transfer this optimal power to the DC load. We are focused on the use of a large value of the load to validate the algorithm. The data acquisition of all our experimental results is carried out using the Controldesk software.

Keywords Systems and applications · Power electronics engineering · Command and control systems for renewable energy · Wind turbine · Control card · Electric power generation

1 Introduction

In isolated sites, people often use small power wind turbines (from 100 W to a few kW) coupled to a permanent magnet synchronous generator without the need for a mechanical multiplier or gearbox required in the conversion chains.

The role of the electronic power interface between the load and the generator is often to control the latter in order to extract the maximum power from the wind. These converters make it possible to place the working point of the wind turbine at the MPP. The maximum power is regulated either mechanically by acting on the orientation of the blades or electrically by matching the impedance of the load with

B. Lahfaoui (✉)

Mohammed First University Oujda, EST Oujda, LGEM, 60000 Oujda, Morocco

e-mail: b.lahfaoui@ump.ac.ma

that of the system. Various research studies are published on the control of PMSG for wind turbine [1–6] or DFIG wind turbine [7–10].

The objective of this paper is to study the power control of a small wind turbine permanent magnet synchronous generator (PMSG) with an internal rectifier connected to a DC-DC converter (boost) and a resistive load for using its in isolated site.

This paper is structured as follows: first of all, a presentation of the wind turbine studied as well as its characteristics in the second section. The third section summarizes in detail the different blocks of our system (the algorithm, DC-DC converter, all the equipment used). Then, the fourth section will focus on the experimental results developed during this research work. And, finally we end with a conclusion.

2 Small Wind Turbine System: The Specification

2.1 Small Wind Turbine

It is clear that the output of power supplied by small wind turbines is lower than that of large ones. They are generally used to power homes, farms and isolated sites. The physical and technological laws are the same as large wind turbines. Small wind turbines often have direct drive generators, direct current output without the need for a parallel device such as the mechanical multipliers used for large wind turbines.

The wind turbine used for optimization in our research study is a small horizontal axis wind turbine with six blades (see Fig. 1), rated power equal to 300 W [1]. The Sect. 2 presents all the measurements captured practically such as the power, voltage and current; all these data are modeled on two graphs, the first with a power abscissa axis and the other with a current axis, knowing that the ordinate axis is relative to voltage.
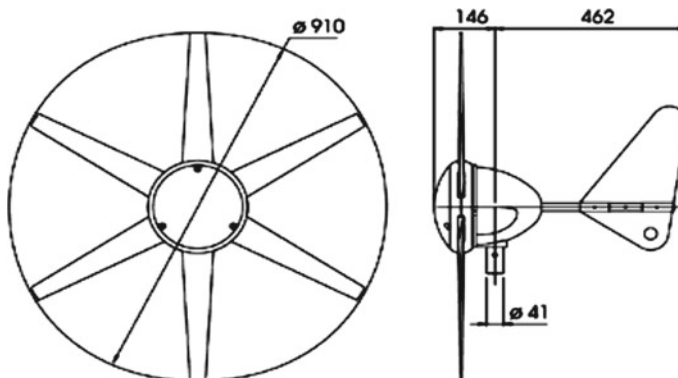
Fig. 1 Engineering drawing of wind turbine
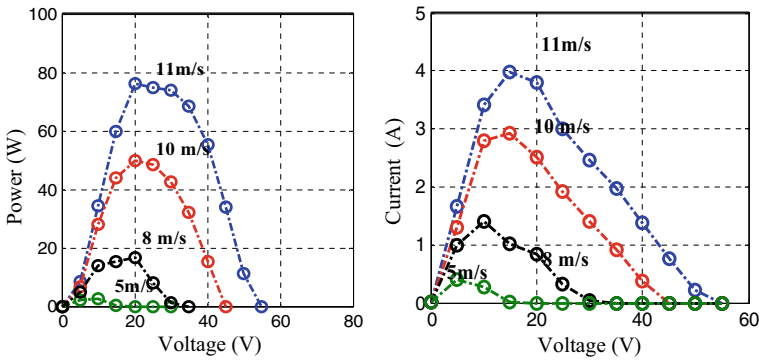
Fig. 2 Experimental electrical characteristics of the wind turbine

2.2 Experimental Data Measurements: Power, Voltage, Current

To draw and trace the technical characteristics of our wind turbine in the laboratory, we put this wind turbine inside a tunnel with a variable airflow thanks to a bellows tunnel, a connection of the machine with a variable resistance and carried out while measuring the voltage and the current at the terminals of the load (V_{WT}, I_{WT}). These last electrical parameters are measured for each wind speed value and each resistance value. Figure 2 illustrates all these measured data.

3 Digital Control of the Wind Turbine System

3.1 The Digital Algorithm: Applications in Control Card

The perturb and observe numerical control have been implemented on a dSPACE 1104 card to perform power regulation.

The power regulation of our horizontal axis wind turbine was made with the use of the dSPACE 1104 board. The algorithm that follows the MPP point is called P&O.

The command starts at the beginning by taking measurements of current and voltage $V_{WT(k)}$, $I_{WT(k)}$ in order to calculate the power at instant k and also recorded the power calculated at the previous instant $k - 1$. After that, a parameter of voltage $\Delta V_{WT} = V_{WT}(k) - V_{WT}(k - 1)$ is calculated following the steps of the algorithm which will control the sign of the value of ΔP_{WT} based on the sign of the value of ΔV_{WT}. In the end of the control process, an increment and decrement of the duty cycle D_{WT} by a value of ΔD_{WT} is made.

For a real-time implementation, this algorithm is translated into MATLAB language via the Simulink tool and then sent to the dSPACE card to be manipulated by the ControlDesk software [11, 12].

3.2 The Digital Experimental Control Wind System

Looking to transfer the optimal energy of the small wind turbine, we have implemented the algorithm shown in Fig. 5 in the dSPACE hardware. Figure 3 represents the equipment used in the test, a schematic block (see Fig. 4) formed by:

- Wind turbine source 300 W,
- Artificial wind speed source (m/s),
- One step-up converter,
- Resistive load,
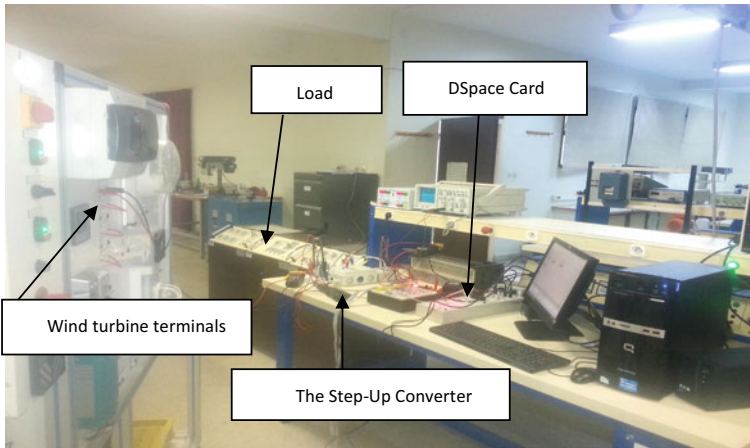
Fig. 3 The equipment used in the experiment
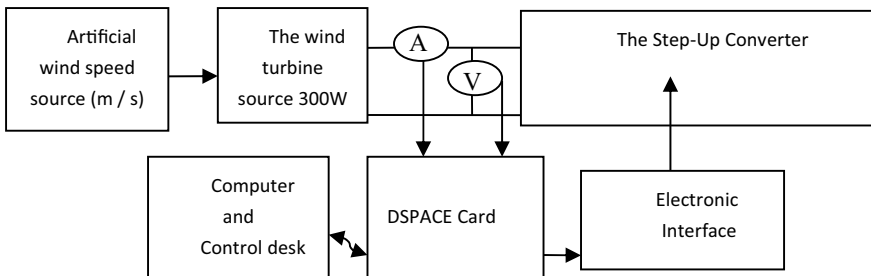
Fig. 4 Schematic block used in the experiment
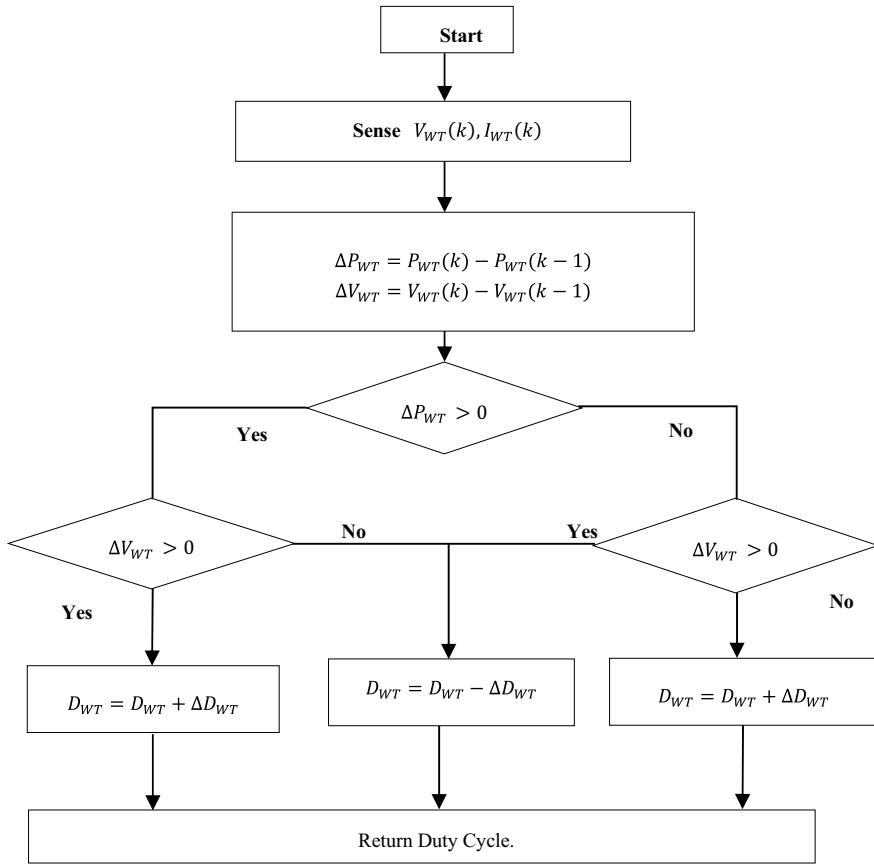
Fig. 5 Digital control implemented in dSPACE 1104

- Current sensor,
- Voltage sensor,
- dSPACE hardware,
- Desktop computer.

3.3 Specification of Step-Up Converter

In this part of this paper, we present the type of converter used in the actuator part of our system, and all the electronic components with these values are cited in Table 2.

Table 1 contains the voltage and current sensors used in the experiment.

Table 1 Measuring equipments

Equipment	Parameters
Current sensor	10 A to 1 V and 100 A to 1 V
Voltage sensor	ST1000

Table 2 DC-DC converter components

Elements	Values
Inductances	L'1 = 100 μH
internal resistance	r'1 = 0.2 Ω
Input capacities	C1 = 1000 μF/250 V
Output capacities	C2 = 1000 μF/250 V
R (load)	2 kΩ
Switch	IRF3710
Diode	MUR820G

4 Experimental Results of Numerical Control System

The results mentioned in this paper is carried out under conditions without disturbance, i.e., with a fixed load equal to 2 kΩ and a constant speed equal to 10 m/s.

In this work, we are focused on the use of a large value of the load to validate the algorithm.

We can clearly notice in Fig. 6 that the electrical power is regulated at a maximum value due to the P&O algorithm implemented in the dSPACE card:

- The regulated power at the terminals of the wind turbine is approximately equal to 50 W.
- The duty cycle is approximately equal to 0.8.

We used the ControlDesk software as a single work environment, from the start of experimentation right to the end. It performs all the necessary tasks of calculation, display. The interval of visualization along this test is 200 s.

5 Conclusion

The curve representing the characteristics of a wind system (power as a function of rotational speed) has a point of maximum power that can be extracted from the wind. Each wind turbine has what is called a power coefficient (C_p), the value of the latter depends on the mechanical design form of the blades of the wind turbine, the speed of the wind, the speed of the rotor of the turbine. Therefore, the main purpose of the
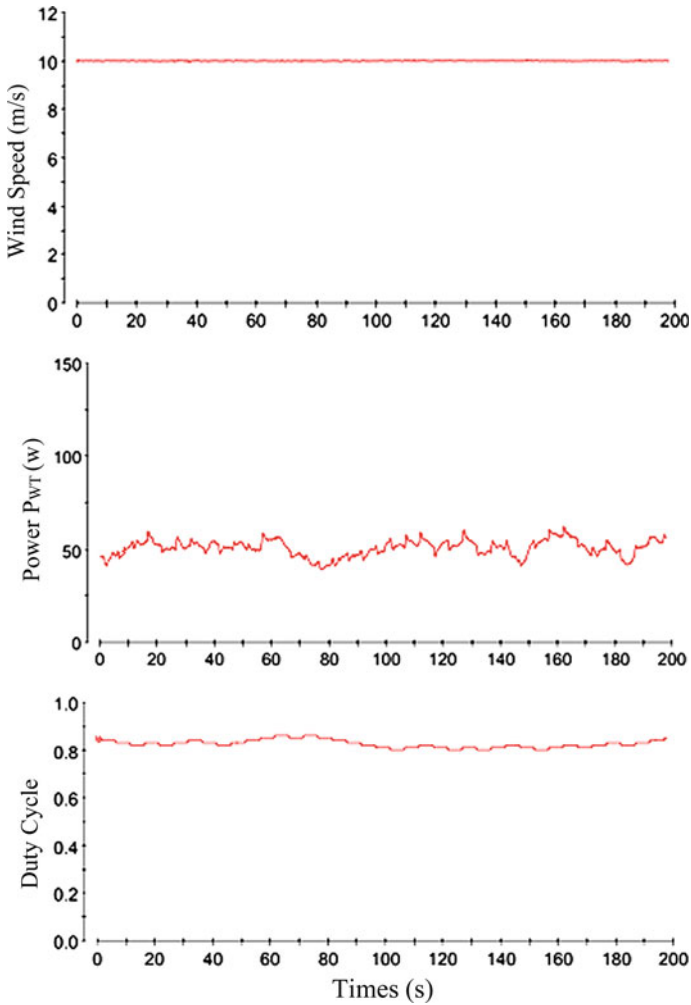
Fig. 6 Tracking the maximum power using dSPACE 1104 card (in ControlDesk)

monitoring process MPPT is to regulate the wind turbine to operate very close to this maximum point.

The work developed in this paper is focused on the maximization of power of a wind turbine with horizontal axis by a numerical control called P&O carried out in real time within our laboratory by the use of the dSPACE card and also a DC-DC converter connected to a resistive load. The results found show the validity of our technical proposal even for a large resistive load.

References

1. Lahfaoui B, Zouggar S, Elhafyani ML, Rabhi AH (2019) Implementation of a real-time MPPT of hybrid renewable energy system composed of wind turbine and solar PV cells. In: Hajji B, Tina G, Ghoumid K, Rabhi A, Mellit A (eds) Proceedings of the 1st international conference on electronic engineering and renewable energy. ICEERE 2018. Lecture notes in electrical engineering, vol 519. Springer, Singapore. https://doi.org/10.1007/978-981-13-1405-6_58

2. Lahfaoui B, Zouggar S, Mohammed B, Elhafyani ML (2017) Real time study of P&O MPPT control for small wind PMSG turbine systems using arduino microcontroller. Energy Procedia 111:1000–1009

3. Lahfaoui B, Zouggar S, Elhafyani ML, Seddik M (2015) Experimental study of P&O MPPT control for wind PMSG turbine. In: 2015 3rd International renewable and sustainable energy conference (IRSEC), Marrakech, Morocco, 2015, pp 1–6. https://doi.org/10.1109/IRSEC.2015.7455020

4. Lahfaoui B, Zouggar S, Elhafyani ML, Benslimane A (2015) Modeling validation and MPPT technique of small wind PMSG turbines using DSPACE hardware. In: 2015 3rd International renewable and sustainable energy conference (IRSEC), Marrakech, Morocco, 2015, pp 1–6. https://doi.org/10.1109/IRSEC.2015.7454923

5. Lahfaoui B, Zouggar S, Elhafyani ML, Seddik M (2016) An experimental study of P&O MPPT control for photovoltaic systems. Int J Power Electron Drive Syst (IJPEDS) 7

6. Badreddine L, Zouggar S, Elhafyani ML, Kadda FZ (2014) Experimental modeling and control of a small wind PMSG turbine. In: 2014 International renewable and sustainable energy conference (IRSEC), Ouarzazate, Morocco, 2014, pp 802–807. https://doi.org/10.1109/IRSEC.2014.7059880

7. Bossoufi B, Karim M, Lagrioui A, Taoussi M, Derouich A (2015) Observer Backstepping control of DFIG-Generators for wind turbines variable-speed: FPGA-based implementation. Renew Energy J 81:903–917.

8. Bossoufi B et al (2020) DSPACE-based implementation for observer backstepping power control of DFIG wind turbine. IET Electr Power Appl 14(12):2395–2403. https://doi.org/10.1049/iet-epa.2020.0364

9. Bossoufi B, Karim M, Taoussi M, Alami Aroussi H, Bouderbala M, Deblecker O (2021) Rooted tree optimization for backstepping power control of DFIG wind turbine: dSPACE Implementation. IEEE Access 9(1):26512–26522

10. Bouderbala M, Bossoufi B, Deblecker O, Aroussi HA, Taoussi M, Lagrioui A, Mtahhir S, Masud M, Alraddady FA (2021) Experimental validation of predictive current control for DFIG: FPGA implementation. Electronics 10(21):2670. https://doi.org/10.3390/electronics10212670

11. Al-Gizi A, Al-Saadi M, Al-Chlaihawi S, Craciunescu A, Fadel MA (2018) Experimental installation of photovoltaic MPPT controller using arduino board. In: 2018 International conference on applied and theoretical electricity (ICATE), Craiova, Romania, 2018, pp 1–6. https://doi.org/10.1109/ICATE.2018.8551397

12. Priyadarshi N, Padmanaban S, Holm-Nielsen JB, Blaabjerg F, Bhaskar MS (2020) An experimental estimation of hybrid ANFIS–PSO-Based MPPT for PV grid integration under fluctuating sun irradiance. IEEE Syst J 14(1):1218–1229. https://doi.org/10.1109/JSYST.2019.2949083

New Strategy for Unbalance Compensation Based on PWM AC-Chopper for Railway Applications

Mir Ismail, Benslimane Anas, Bouchnaif Jamal, Lahfaoui Badreddine, Ayat Yassine, and Yandouzi Mimoun

Abstract This paper focuses on the problem of voltage unbalance for single-phase railway substations using the AC-chopper controlled impedance concept (CCI). This topology is based on the use of an active Steinmetz circuit. The sizing and control method of the compensator has been studied based on real data provided by the Moroccan railway office (ONCF). A comparison was realized between the CCI compensator, voltage source inverter-based STATCOM VSI, and current source inverter-based STATCOM CSI in terms of power losses in the semiconductors. The three topologies have been implemented in the MATLAB/Simulink environment. The obtained results prove the efficiency of the CCI topology compared to the VSI and CSI compensators.

Keywords Unbalance compensation · STATCOM · VSI · CSI · CCI · AC-chopper · Railway

1 Introduction

The high-speed railway lines are supplied by a 2 * 25 kV voltage between the catenary and a complementary cable called NEGATIVE FEEDER. A 25 kV voltage between the catenary and the rail is obtained due to an autotransformer with a midpoint [1].

Railway substations are two-phase loads, as shown in Fig. 1, which are dynamic loads with high power, giving the unbalance problem. It will influence the quality of the power supply by increasing the power generator losses, reducing the transformer output capacity, disturbing protection relay devices, and consequently operating errors. These negative effects seriously affect the safe operation of the power system.

M. Ismail (✉) · B. Anas · A. Yassine · Y. Mimoun
Renewable Energy, Embedded System and Data Processing Laboratory, National School of Applied Sciences, Mohamed First University, Oujda, Morocco
e-mail: ismail.mir@ump.ac.ma

B. Jamal · L. Badreddine
Laboratory of Electrical Engineering and Maintenance, Higher School of Technology, University Mohammed I Oujda, Oujda, Morocco
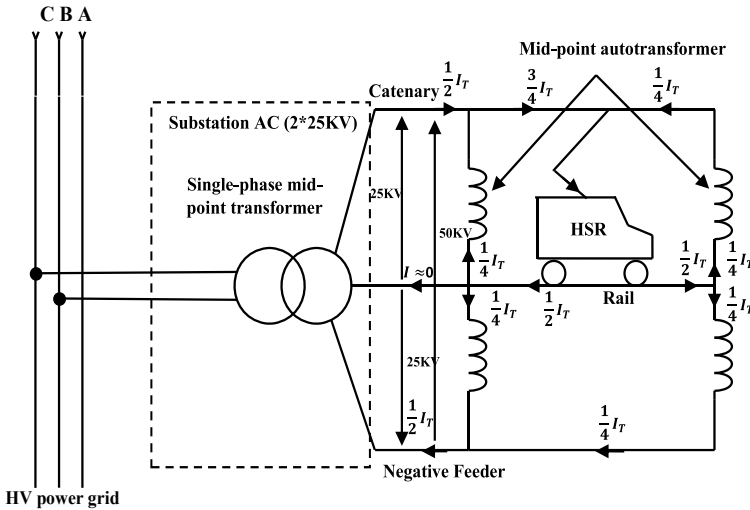
Fig. 1 Power circuit of the high-speed railway substation's supply AC system (2 * 25 kV–50 Hz)

The unbalance problem is an electrical disturbance that degrades the power quality of the Moroccan energy supplier (ONEE); for this reason, the use of compensators is mandatory [2, 3].

In order to overcome the unbalance problem, different strategies have been developed. The first compensator is the STATCOM based on the voltage source inverter (VSI) topology; it is composed of a number of compensation cells coupled in parallel to the secondary of a three-phase transformer connected through Yy0 each cell contains an AC-side filter inductor (L), a voltage source inverter with sinus-triangle PWM control, and a DC-side voltage storage source constituted by a capacitor (C_{dc}). The second is the STATCOM compensator based on the current source inverter (CSI); it is composed of several cells coupled in parallel with the secondary of the three-phase transformer with a Y_{y0} connection. Each cell contains a DC source realized by an inductor L_{dc}, a CSI converter composed of six unidirectional power electronic switches (only IGBT), and an LC filter connected between the AC side of the inverter and the secondary of the transformer; the IGBT switches are controlled by PWM signals [4, 5]. The efficiency of the CSI compensator is high compared to the VSI in terms of power losses, but the disadvantage is the complexity of the design of the protection devices due to the current source. For this reason, a new compensation topology based on voltage sources with AC-chopper controlled impedance concept (CCI) is proposed. Generally, CCI is controlled by current loops; in [6, 7], authors proposed a control loop based on measuring the inner current provided by railway substation in order to estimate the reference current and after obtain the duty cycle for controlling the AC-chopper.

This paper presents an unbalance compensation topology for substations used in high-speed railways using the CCI compensator for Tangier (Pk10) and Kenitra

(Pk185). The compensator structure, sizing, control, and a comparison of power losses with the topologies VSI and CSI are presented.

The rest of this paper is organized as follows: In the second section, the technical characteristics provided by ONCF and the results obtained without compensation are presented. The third section deals with the sizing and control of the CCI compensator. Finally, the calculation of power losses in the CCI semiconductors and comparison with VSI and CSI topologies.

2 Unbalance Analysis

In order to simulate and analyze the unbalance caused by HSR substations, ONCF has provided the characteristics of both substations (Tangier and Kenitra) for the 2030 horizon as well as the information on the high voltage (HV) power supply line. Table 1 presents the relative characteristics of the substations, including the technical specifications of the power supply line [8–11].

The powers consumed by both substations for 9 h from 7 am to 4 pm are shown in Fig. 2a for Tangier (pk10) and Fig. 3a for Kenitra (185). Figures 2b and 3b present the unbalance factor without compensator for the substations under study (Tangier and Kenitra, respectively); note that the unbalance factor is more than 2% and the Moroccan energies provider (ONEE) imposes an unbalance factor limited to 1%; therefore, the ONCF will be penalized. For this reason, the railway operator has engaged to install compensators based on the FACTS system to respect the standards and avoid penalties.

Table 1 Characteristics of the Kenitra and Tangier substations

Characteristics	Tangier-side substation (PK10)	Kenitra-side substation (PK185)
Rated power of the single-phase transformer	40 MVA	40 MVA
Rated voltage of the line HT	225 kV	225 kV
Short-circuit power at the PCC	2800 MVA	3000 MVA
Line parameters	$R_{\text{line}}\ (\Omega/\text{Km}) = 0.07$	$R_{\text{line}}\ (\Omega/\text{Km}) = 0.07$
	$L_{\text{line}}\ (\text{mH/Km}) = 1.26$	$L_{\text{line}}\ (\text{mH/Km}) = 1.26$
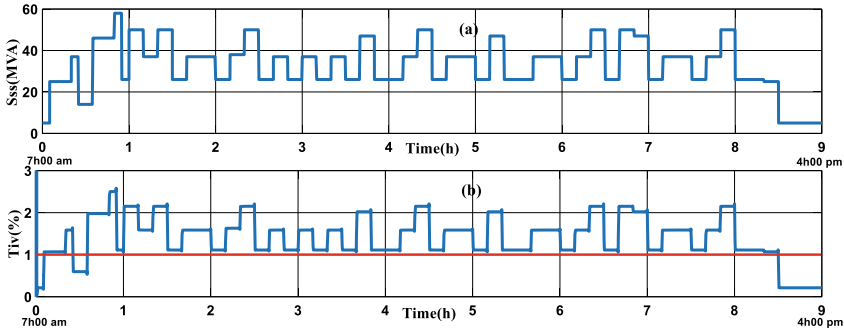
Fig. 2 **a** Daily rail traffic for the Tangier substation (pk10) in 2030. **b** Predicted unbalance factor for the Tangier substation
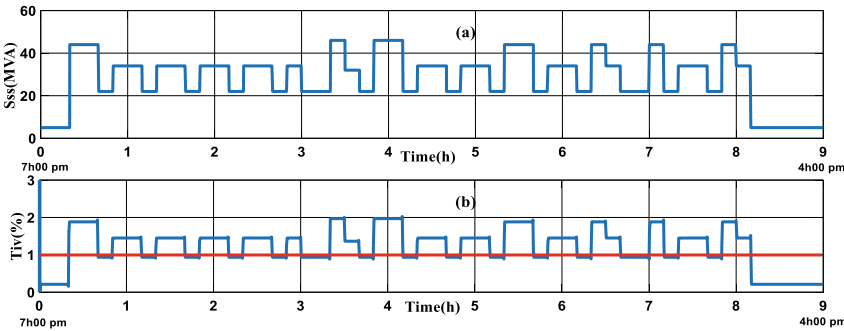
Fig. 3 **a** Daily rail traffic for the Kenitra substation(pk185). **b** Predicted unbalance factor for the Kenitra substation

3 Sizing and Control Strategy of the Active Steinmetz Compensator (CCI)

The classical Steinmetz circuit is used to compensate the unbalance of fixed loads by coupling fixed capacitive and inductive impedances. For a variable load (railway substations), the classical Steinmetz circuit is not able to compensate the unbalance.

The CCI compensator is inspired by the concept of the Steinmetz circuit instead of injecting fixed impedances; the injection will be by variable capacitive and inductive impedances.

Figure 5 shows the general structure of the CCI compensator; this structure is divided into two main parts: the power part and the control part. The setpoint impedance ($z_{in,c}$) is calculated from the power consumption S_{SS} using Eqs. (1) and (2). After the calculation of $Z_{in\,c}$, a nonlinear control method will be used to generate the duty cycle α and transform it to PWM signals to control the AC-choppers of the compensator, an LC filter is installed between the input part of the compensator and the transformer secondary in order to reduce THD.

$$R_{ss} = \frac{U(V)^2}{S_{ss}(VA)} \tag{1}$$

where: U is the voltage of the high voltage line (225 kV), R_{SS} is the equivalent resistance of the substation, S_{SS} is the power consumed by the substation.

$$Z_C = Z_L = \sqrt{3}R_{ss} \tag{2}$$

where: Z_C is the capacitive impedance and Z_L is the inductive impedance.

3.1 Sizing of the CCI Compensator

The IGBTs have been chosen from the manufacturer ABB (model 5SNA 1200G450300) to satisfy the technical constraints in the high voltage application, the Table 2 shows the technical characteristics of the IGBTs used.

Transformer 225/3.5 kV ensures the voltage adaptation for the IGBTs as well as the galvanic isolation between the high voltage line and the compensator; Eq. (3) shows the transformation ratio of the transformer.

Table 2 Technical characteristics of the IGBT

Component	Parameter	Value
IGBT	V_{0T}: Threshold voltage of the transistor	3.2 v
	r_{dT}: Dynamic resistance of the transistor	2.94 mΩ
	V_{ref}: Rated voltage	4500 v
	$a_{\text{on}} + a_{\text{out}}$	1.2×10^{-6}
	$b_{\text{on}} + b_{\text{off}}$	6.1×10^{-3}
	$c_{\text{on}} + c_{\text{off}}$	1.08
Antiparallel diode integrated with IGBT	V_{0D}: Threshold voltage of the diode	3.25 v
	r_{dD}: Dynamic resistance of the diode	2.98 mΩ
	V_{ref}: Rated voltage of the diode	4500 v
	a_{rec}	-5.25×10^{-7}
	b_{rec}	2.61×10^{-3}
	c_{rec}	375×10^{-3}

Table 3 Parameters of the CCI compensator

	CCI Capacitive	CCI Inductive
Single-phase coupling transformer	225 kv/3.5 kv 34 MVA	225 kv/3.5 kv 34 MVA
Numbers of AC-chopper cells	$N1 = 12$	$N2 = 8$
LC filter	$L_{F1} = 92\,\mu H$, $C_{F1} =$ 65.51 $\mu F/Q_{CF} = 250$ KVAR	LF2 = 92 μH, $C_{F1} =$ 65.51 $\mu F/Q_{CF} = 250$ KVAR
AC-chopper output impedance Z_{out}	$L_{CV} = 0.362$ mH, $C_v =$ 831 μF	$L_V = 8$ mH, $r = 31.2$ mΩ

A step-down AC-chopper is used, based on the interval of the setpoint impedance variation to define the duty cycle α, if $Z_{in,c} = Z_{in,min}$ the duty cycle takes the max value ($\alpha = 1$) and $Z_{in,min} = Z_{out}$, Eq. (4) shows the relation between alpha and the input/output impedance of the AC-chopper.

Equation (5) is used to calculate the reactive power produced by the compensator to determine the capacitive/inductive parameters. The parameters of the CCI are presented in Table 3.

$$m_{T1} = m_{T2} = \frac{3.5}{225} = 15.55 \times 10^{-3} \tag{3}$$

$$Z_{in} \approx \frac{Z_{out}}{\alpha^2} \tag{4}$$

where: Z_{in} is the input impedance and Z_{out} is the output impedance.

$$Q_C = \frac{U^2}{Z_{comp}} \tag{5}$$

where: Q_C is the reactive power of the compensator, and Z_{com} is the impedance of the compensator.

3.2 Control Strategy of the CCI Compensator

The look-up table method with a PID corrector has been used to control the system due to the nonlinearity between alpha, Z_{in} and Z_{out} as in Eq. (4). Look-up table requires an increasing function for the order of the data, to satisfy this condition we use the admittance $Y = \frac{1}{Z}Y = \frac{1}{Z}$. Tables 4 and 5 presents an extract of the data inserted in Look-Up Table. Figure 6 shows the direct chain of the system, applying a step input to determine the open-loop transfer function (Eq. 6).

To improve the system's performance (speed, precision and stability), a PI corrector is introduced in the direct chain (Fig. 7). Table 6 shows the parameters

Table 4 Extract capacitive data from look-up table

α	Z_{in} (kΩ)	Y_{in}
0.28	15.6846613	0.063756557
0.285	15.2070623	0.065758921
0.29	14.7208141	0.067931026
0.295	14.281267	0.070021798
0.3	13.942297	0.071724193

Table 5 Extract inductive data from look-up table

α	Z_{in} (kΩ)	Y_{in}
0.27	20.87892	0.047895198
0.275	20.21945	0.049457329
0.28	19.31513	0.051772885
0.285	18.44628	0.054211472
0.29	17.89582	0.055878971

Table 6 Parameters of the PI corrector and the performance of the system in closed loop

PI	CCI Inductive	CCI Capacitive
k_p	14.4×10^{-3}	14.4×10^{-3}
τ_i (ms)	0.4	0.4
T_r (ms)	59	59
Static error (%)	0	0
Overflow in closed loop (%)	0	0

of the PI corrector obtained by the application of the Ziegler–Nichols method as well as the response time (T_r), static error and overflow in closed loop. The system's response in the open and closed loops is shown in Figs. 8 and 9.

$$H(p) = \frac{1.003}{1 + 3.87 \times 10^{-3} p + \frac{p^2}{314.74^2}} \tag{6}$$

3.3 Result of the Unbalance Compensation by CCI

The results obtained are simulated by MATLAB Simulink environment which are presented in Figs. 10 and 11. The unbalance factor (T_{iv}) for both substations has been respected the limits imposed by ONEE ($T_{iv} \leq 1$), and also, the harmonic distortion factor of the voltage (THD_L) and of the current (THD_C) do no exceed the standard (8% imposed by NF EN 50160); the CCI compensator approves its efficiency to resolve the unbalance problem.

4 Power Losses for the CCI Compensator

Two types of power losses are considered for the present system: conduction and switching losses. To determine the conduction losses due to the internal construction of the IGBTs (IGBTs are based on semiconductors modeled by an internal resistor and a voltage source), and for a PWM control, the output current is defined by Eq. 7.

The power losses at the level of the transistors ($T1$ and $T1C$) and antiparallel diodes ($D1$ and $D1C$) are determined by the Eqs. 8, 9, 10 and 11 for the switching cell 1 (Fig. 4).

Still, due to the symmetry of the switching cells, the expressions of the losses for $T2$, $D2$, $T2C$, and $D2C$ are respectively equal to those obtained for $T1$, $D1$, $T1C$, and $D1C$.
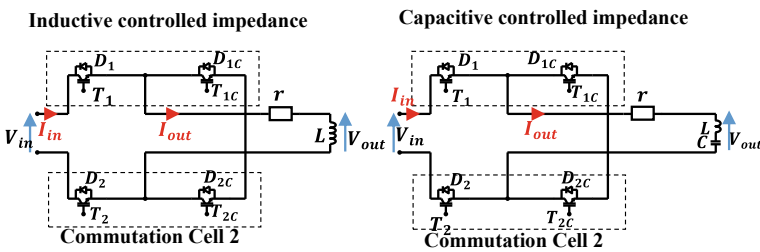
Fig. 4 Inductive and capacitive variable impedances for CCI compensator
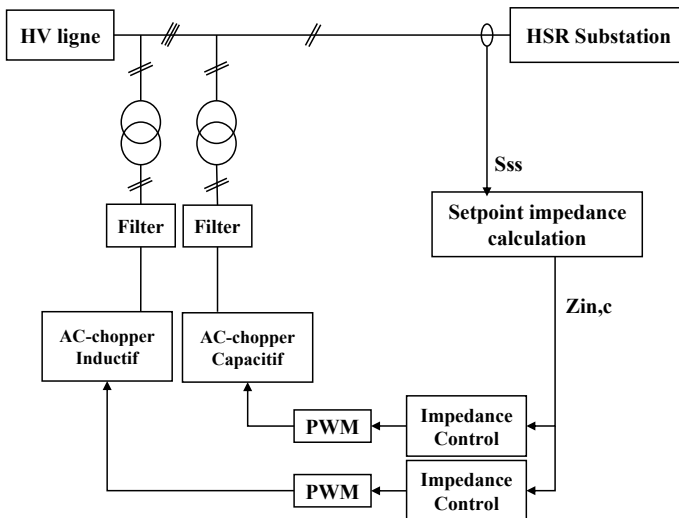
Fig. 5 General structure for CCI compensator

Fig. 6 Open loop of the CCI compensator
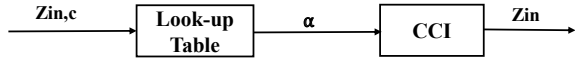
Fig. 7 Closed loop of the system
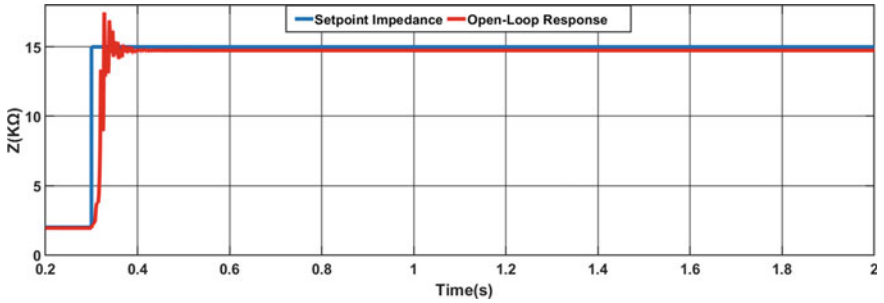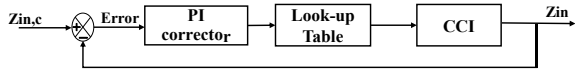
Fig. 8 Response of the system in the open loop
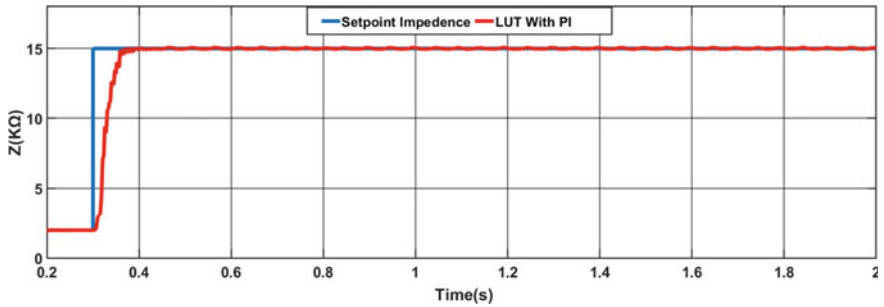
Fig. 9 Response of the system in the closed loop
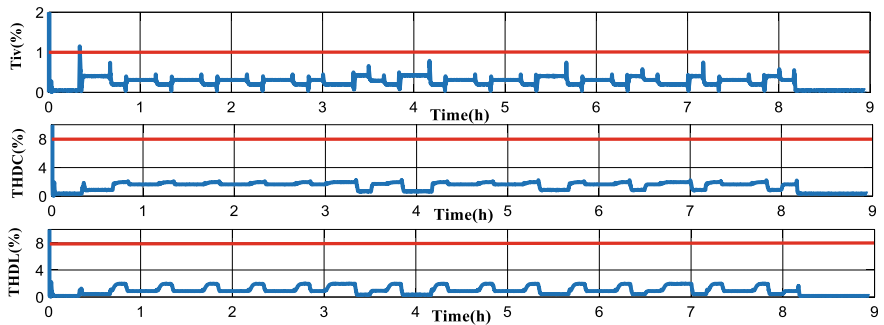
Fig. 10 T_{iv}, THD_C, and THD_L according to daily railway traffic Tangier 225 kV with unbalance compensation by CCI
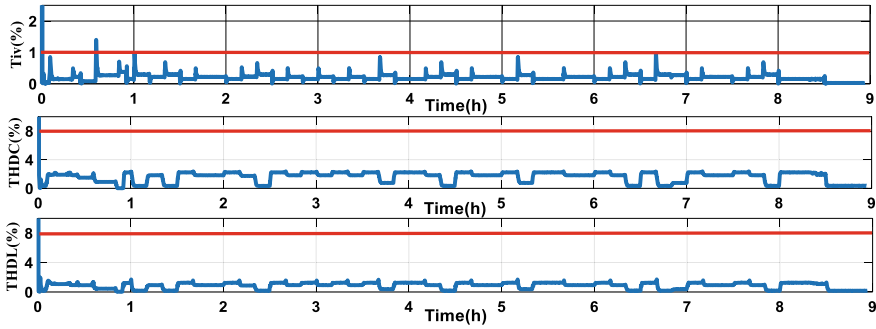
Fig. 11 T_{iv}, THD_C, and THD_L according to daily railway traffic Kenitra 225 kV with unbalance compensation by CCI

For switching losses, the voltage applied during the change of state of the semiconductor varies according to Eq. 12.

The switching phase of the semiconductors in a cell is one quarter of a grid period. Switching cell 1 switches the voltage Vin when it is positive. The semiconductors are charged according to the sign of the current, which changes every quarter. If the voltage Vin is negative, the components are forced to close. The calculation of the losses is done from Eq. (13 and 14) [8, 12].

$$i_{\text{out}}(t) = I_{\text{out}} \cdot \sqrt{2} \cdot \sin\left(\omega_{\text{net}} \cdot t + \frac{\pi}{2}\right) \tag{7}$$

$$P_{\text{Cond},T1} = V_{0T} \cdot \alpha \cdot \frac{I_{\text{out}} \cdot \sqrt{2}}{\pi} + r_{dT} \cdot \frac{\left(I_{\text{out}} \cdot \sqrt{2}\right)^2}{4} \cdot \alpha \tag{8}$$

$$P_{\text{Cond},T1C} = V_{0T} \cdot (1 - \alpha) \cdot \frac{I_{\text{out}} \cdot \sqrt{2}}{\pi} + r_{dT} \cdot \frac{\left(I_{\text{out}} \cdot \sqrt{2}\right)^2}{4} \cdot (1 - \alpha) \tag{9}$$

$$P_{\text{Cond},D1} = V_{0D} \cdot \alpha \cdot \frac{I_{\text{out}} \cdot \sqrt{2}}{\pi} + r_{dD} \frac{\left(I_{\text{out}} \cdot \sqrt{2}\right)^2}{4} \cdot \alpha \tag{10}$$

$$P_{\text{Cond},D1C} = V_{0D} \cdot (1 - \alpha) \cdot \frac{I_{\text{out}} \cdot \sqrt{2}}{\pi} + r_{dD} \frac{\left(I_{\text{out}} \cdot \sqrt{2}\right)^2}{4} \cdot (1 - \alpha) \tag{11}$$

$$V_{\text{in}}(t) = V_{\text{in}} \cdot \sqrt{2} \cdot \sin(\omega_{\text{net}} \cdot t) \tag{12}$$

$$P_{\text{Com}T1} = P_{\text{Com}1C}$$

$$= \frac{f_{\text{dec}}}{2\pi} \frac{V_{\text{in}} \cdot \sqrt{2}}{V_{\text{ref}}} \left[\frac{(I_{\text{out}} \cdot \sqrt{2})^2}{3} \cdot (a_{\text{on}} + a_{\text{off}}) + \frac{I_{\text{out}} \cdot \sqrt{2}}{2} (b_{\text{on}} + b_{\text{off}}) + (c_{\text{on}} + c_{\text{off}}) \right] \tag{13}$$

$$P_{\text{Com}D1} = P_{\text{Com}D1C}$$

$$= \frac{f_{\text{dec}}}{2\pi} \cdot \frac{V_{\text{in}} \cdot \sqrt{2}}{V_{ref}} \left[\frac{(I_{\text{out}} \cdot \sqrt{2})^2}{3} \cdot a_{\text{rec}} + \frac{I_{\text{out}} \cdot \sqrt{2}}{2} \cdot b_{\text{rec}} + \cdot c_{\text{rec}} \right] \tag{14}$$

The results of the modeling and implementation of the different equations in the MATLAB/Simulink environment of the total losses in each inductive and capacitive controlled impedance in the two substations of Tangier and Kenitra and the efficiency are represented by Figures 12 and 13, respectively.

For the Tangier substation, the total power losses in the two AC-choppers (capacitive and inductive) reach 0.42 MW; the results obtained show a high efficiency as a function of consumption variation.

For the Kenitra substation, the losses are extremely low in both impedances, resulting in a high efficiency. Table 7 shows the different values of power losses for the three different topologies (CSI, VSI, and CCI). It is clear that the losses in the
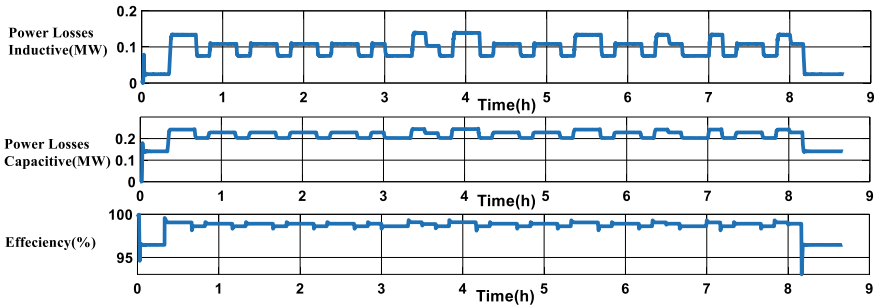
Fig. 12 Inductive and capacitive power losses for the Tangier substation as well as the efficiency
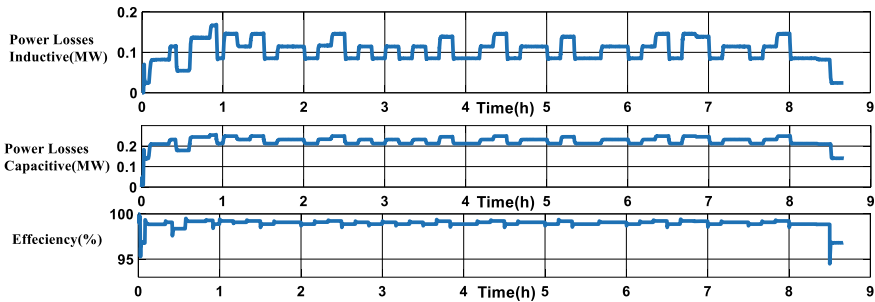
Fig. 13 Inductive and capacitive power losses for the Kenitra substation as well as the efficiency

Table 7 Comparison of the semiconductor energy power losses between the three topologies CCI, VSI, and CSI STATCOM

Unbalance compensator topology	Energy lost daily (MWh)	Energy lost annually (MWh)
CCI	2.8901	1054.89
VSI_STATCOM [8]	44.46	16,227.9
CSI_STATCOM [8]	38.7	14,125.5

CCI topology are very low compared to the VSI and CSI structures, according to [8], which has a huge advantage in minimizing power losses.

5   Conclusion

The electrical power supplied to the substations of high-speed railway lines necessitates permanent control to ensure the power's quality and avoid penalties to the ONCF. Therefore, the railway operator is obliged to install compensators to optimize the power consumption respecting the norms required by the power supplier ONEE. This paper focuses on the study on a new unbalance compensation topology based on the concept of controlled impedance CCI starting from the dimensioning of the components (transformers, IGBT, filter, …) then the method of control of the compensator until the implementation and the modeling at the level of MATLAB Simulink. The results constated that the standards were respected (T_{iv}, THD_L, and THD_C). A comparative study was established between the CCI compensator and the compensators (VSI and CSI) proposed in [8] concerning the power losses in the semiconductors; the results approved a high performance of the CCI compensator compared to the VSI and CSI compensators.

References

1. Grunbaum R (2007) FACTS for dynamic load balancing and voltage support in rail traction. In: European conference on power electronics and applications. IEEE, pp 1–10
2. Ismail M, Anas B, Yassine A, Mohammed B (2022) Improved control technique based on neural network for AC-Chopper of railway substations. In: 2022 2nd International conference on innovative research in applied science, engineering and technology (IRASET). IEEE, pp 1–6
3. Župan A, Teklić AT, Filipović-Grčić B (2013) Modeling of 25 kV electric railway system for power quality studies. In Eurocon 2013. IEEE, pp 844–849
4. Uzuka T, Ikedo S, Ueda K (2004) A static voltage fluctuation compensator for AC electric railway. In: 2004 IEEE 35th annual power electronics specialists conference (IEEE Cat. No. 04CH37551), vol 3. IEEE, pp 1869–1873

5. Tanta M, Monteiro V, Sousa TJ, Martins AP, Carvalho AS, Afonso JL (2018) Power quality phenomena in electrified railways: Conventional and new trends in power quality improvement toward public power systems. In: International young engineers forum (YEF-ECE). IEEE, pp 25–30
6. Ladoux P, Fabre J, Caron H (2014) Power quality improvement in ac railway substations: the concept of chopper-controlled impedance. IEEE Electrification Mag 2(3):6–15
7. Raimondo G (2012) Power quality improvements in 25 kV 50 Hz railways substation based on chopper-controlled impedances. In: Doctoral dissertation, Institut National Polytechnique de Toulouse-INPT; Seconda Universita degli Studi di Napoli
8. Benslimane A, Bouchnaif J, Essoufi M, Hajji B, El Idrissi L (2020) Comparative study of semiconductor power losses between CSI-based STATCOM and VSI-based STATCOM, both used for unbalance compensation. Prot Control Mod Power Syst 5(1):1–14
9. Benslimane A, Ouariachi ME, Bouchnaif J, Grari K (2018) Analysis of the electrical unbalance caused by the moroccan high-speed railway in the high voltage power grid for the starting horizon (2018) and the horizon (2030). In: International conference on electronic engineering and renewable energy. Springer, Singapore, pp 336–343
10. Benslimane A, Bouchnaif J, Azizi M, Grari K (2016) An experimental study of the unbalance compensation by voltage source inverter based STATCOM. Int J Power Electron Drive Syst 7(1):45
11. Benslimane A, Bouchnaif J, Azizi M, Grari K (2013) Study of a STATCOM used for unbalanced current compensation caused by a high-speed railway (HSR) sub-station. In: International renewable and sustainable energy conference (IRSEC), pp 441–446. IEEE
12. Kantar E, Usluer SN, Hava AM (2013) Design and performance analysis of a grid connected PWM-VSI system. In: 8th International conference on electrical and electronics engineering (ELECO). IEEE

Evaluation of an MPPT DC–DC Boost Controller using a PV Emulator-Based Test System

Mohammed Chaker, Driss Yousfi, Mohammed Essoufi, and Amine El Houre

Abstract This paper proposes a PV Emulator with the objective of testing MPPT algorithms under controllable weather conditions. The emulator is based on a Phase-Shifted Full-Bridge converter with a current regulation loop. The reference current of this loop is generated using a mathematical PV model. In this work, the main focus is on testing the P&O algorithm, which is implemented on a DC–DC boost converter. First, simulation work under Matlab/Simulink environment is realized to check the robustness of the implemented algorithm for different loads and irradiance levels. Then, the P&O-controlled boost converter is tested experimentally using the developed PV Emulator. The control part is ensured by using the EZDSP F28335 board. Both simulation and experimental results are presented at the end of this paper.

1 Introduction

In recent years, solar PV (photovoltaic) energy has become one of the most important renewable energy sources [1–3], owing to technological breakthroughs and their impact on climate change.

The MPPT (Maximum Power Point Tracker) controller is a fundamental part of PV systems. That is why test systems for MPPT algorithms have been developed [4, 5]. The purpose of MPPT is to extract and conserve the maximum power from the PV modules under any environmental condition by outputting their optimal current–voltage operating point according to the load [6].

Several MPPT algorithms have been presented throughout the years to track the maximum power of a PV panel. Although these algorithms share the same main objective, they differ significantly in terms of efficiency, complexity, steady-state

M. Chaker (✉) · D. Yousfi · M. Essoufi · A. El Houre
ESETI Laboratory, National School of Applied Sciences, Mohammed First University, Oujda, Morocco
e-mail: chaker.mohammed01@gmail.com

oscillations, tracking speed, hardware implementation, and whether or not they track global MPP [7, 8]. Among these algorithms, the P&O (perturb and observe) method is the most popular, thanks to its simplicity and ease of use [8–10].

The biggest obstacles for the experimental validation of MPPT algorithms are the PV panel cost, their large space requirements, and, of course, the need for controlled weather conditions that affect the solar cell performance [11]. In order to overcome these difficulties, a PV Emulator has been designed to test MPPT algorithms. The main power circuit of this PV Emulator is an isolated DC–DC power converter.

In this paper, the developed PV Emulator is used to test an MPPT converter with a built-in perturb and observe technique to verify the emulator's effectiveness in terms of testing MPPT algorithms.

2 System Description

As illustrated in Fig. 1, the overall system is composed of two power subsystems: The PV Emulator is the first, and the MPPT-controlled DC–DC converter with the load is the second.

2.1 Description of the Developed PV Emulator

The developed PV Emulator can reproduce the behavior of real PV panels at different climatic conditions. It is built around a PSFB (Phase-Shifted Full-Bridge) DC–DC converter with a current controller.
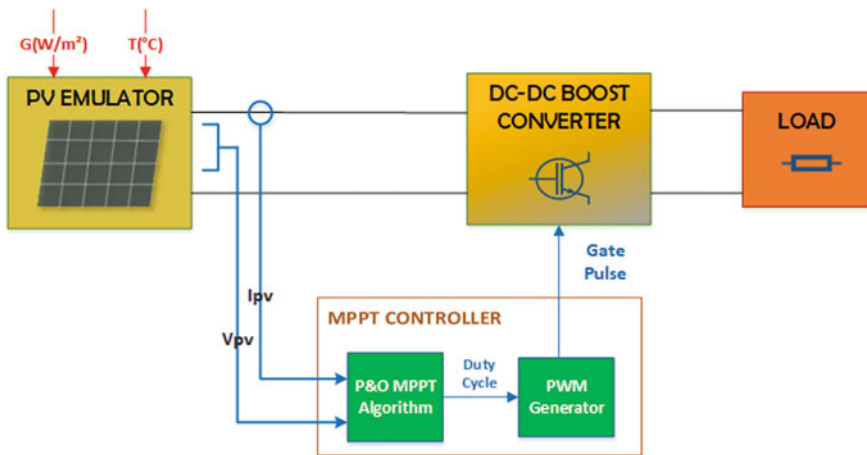
Fig. 1 Block diagram of testing P&O algorithm using the PV Emulator

Figures 2 and 3 show the synoptic diagram and the power circuit of the developed PV Emulator, respectively. The power circuit is mainly composed of an LF (Low Frequency) rectifier feeding a Full-Bridge inverter, an HF (High Frequency) transformer, and an HF rectifier with an LC filter.

The control circuit takes in charge the PV characteristic generation using the mathematical model of the PV module described in Eq. (1). Furthermore, the PI controller uses a Phase Shift PWM control technique to ensure PV current tracking.

Under STC (Standard Test Conditions), i.e., $G_n = 1000$ W/m^2 and $T_n = 25$ °C, the current of a PV module can be expressed as a function of voltage by the simplified exponential expression shown in Eq. (1) [12]:

$$I_{\text{PVn}} = I_{\text{scn}} \cdot \left(1 - K_1 \cdot \left(\exp\left(\frac{V_{\text{pvn}}}{K_2 \cdot V_{\text{ocn}}} \right) - 1 \right) \right) \tag{1}$$

where

I_{pvn}, V_{pvn}: current and voltage supplied by the PV module at STC.

$$K_1 = \left(1 - \frac{I_{\text{mppn}}}{I_{\text{scn}}} \right) \cdot \left(\exp\left(\frac{-V_{\text{mppn}}}{K_2 \cdot V_{\text{ocn}}} \right) \right) \tag{2}$$
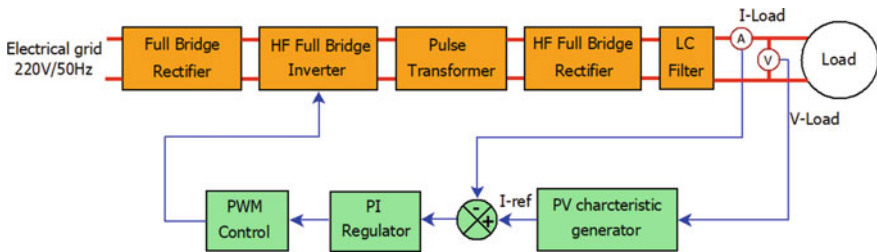
Fig. 2 Synoptic diagram of the designed PV Emulator
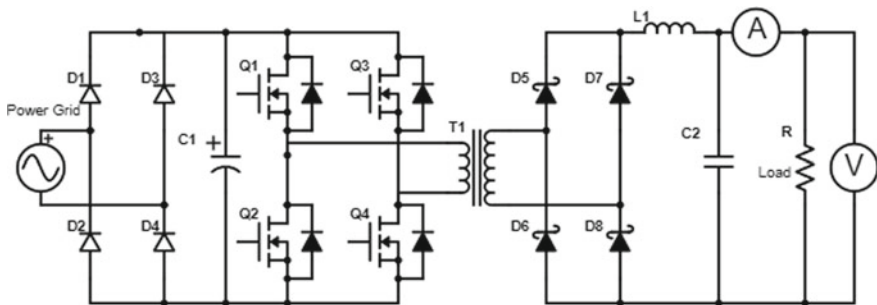
Fig. 3 Power circuit scheme of the designed PV Emulator

$$K_2 = \frac{\left(\frac{V_{\text{mppn}}}{V_{\text{ocn}}} - 1\right)}{\ln\left(1 - \frac{I_{\text{mppn}}}{I_{\text{scn}}}\right)} \tag{3}$$

K_1 and K_2 depend on the following parameters at STC.

The PV module short-circuit current I_{scn}, the PV module open circuit voltage V_{ocn}, the PV module maximum power point current I_{mppn}, and the PV module maximum power point voltage V_{mppn}.

Equation (1) allows defining the electrical characteristic I_{pvn} as a function of V_{pvn} of the PV module at STC. Replacing Eqs. (4–7) in Eq. (1) allows generating the current–voltage characteristic of the PV module corresponding to the new temperature and solar irradiance values.

$$I_{\text{sc}}(G, T) = I_{\text{scn}} \frac{G}{G_n}(1 + \alpha(T - T_n)) \tag{4}$$

$$V_{\text{oc}}(T) = V_{\text{ocn}} + \beta(T - T_n) \tag{5}$$

$$I_{\text{mpp}}(G, T) = I_{\text{mppn}} \frac{G}{G_n}(1 + \alpha(T - T_n)) \tag{6}$$

$$V_{\text{mpp}}(T) = V_{\text{mppn}} + \beta(T - T_n) \tag{7}$$

where α and β are the current and the voltage temperature coefficient, respectively.

2.2 Description of the MPPT-Controlled DC–DC Boost Converter

To extract the maximum power available at the terminals of the PV Emulator at any time and transfer it to the load, a boost DC/DC power converter (Fig. 4) with an implemented P&O mechanism is integrated into the system. The power components used for the realization of the boost converter are the IRF530 power MOSFET and the ESM765-100 diode, whereas the passive components used have the following values: $C_{\text{in}} = 60\,\mu\text{F}$, $C_{\text{out}} = 100\,\mu\text{F}$, and $L = 2\,\text{mH}$.

The P&O method works by perturbing the system and observing its impact on the output power of the PV generator. The perturbation is provided by changing the duty cycle of the PWM signal sent to the boost converter transistor.

According to Fig. 5, if the operating voltage is perturbed in a given direction and the power increases, then the perturbation has shifted the operating point to the MPP (maximum power point). The P&O algorithm will continue to perturb the voltage in the same direction. On the other hand, if the power decreases, then the perturbation has moved the operating point away from the MPP. The algorithm will reverse the
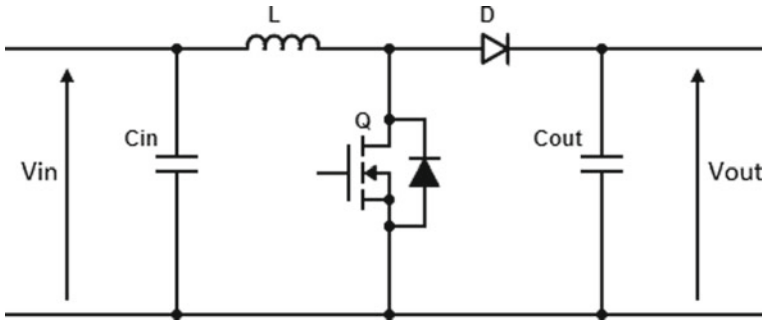
Fig. 4 Power circuit scheme of the DC–DC boost converter

direction of the next perturbation. This process is repeated periodically until the MPP is reached.

Any change in weather conditions results in a different MPP, which triggers the algorithm to look again for this point.
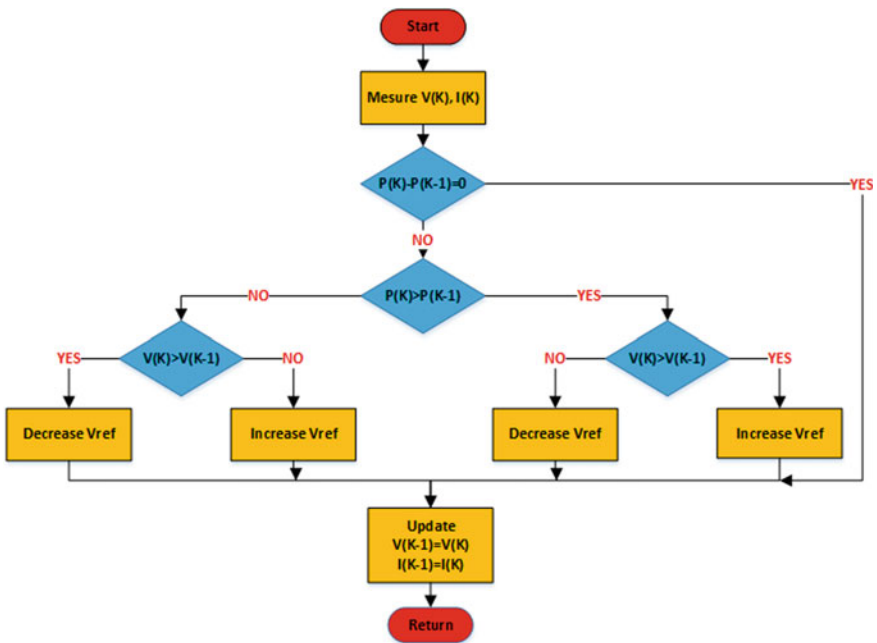
Fig. 5 Flowchart of the P&O algorithm

3   Simulation Results

For an extensive study before starting the experimental part, the general system (PV Emulator + boost converter with P&O control + load) is simulated under Matlab/Simulink environment.

First, to verify the robustness of the implemented P&O algorithm, the global system is tested by implementing the reference PV characteristic (Fig. 6) with two different loads $R = 100\ \Omega$ and $R = 150\ \Omega$. It is noticed that the maximum power point of the reference PV characteristic is equal to 143.1 W.

The obtained results presented in Fig. 7 show that the steady-state power supplied to the first load ($R = 100\ \Omega$) oscillates around an average power of 141.3 W, whereas the power of the second load ($R = 150\ \Omega$) oscillates around 140.6 W. Hence, the P&O algorithm is working properly. The time of the dynamic regime of the generated power is 0.18 s and 0.23 s for R $= 100\ \Omega$ and $R = 150\ \Omega$, respectively.

In a second test, the PV characteristic is generated for a constant temperature equal to 25 °C and with varying irradiance. The power delivered, in these conditions, to 100 Ω load is presented in Fig. 8.

Table 1 shows that for each irradiance value, the overall system delivers an average power, which is very close to the reference MPP. The largest value of the difference between the measured average power and the reference MPP is 1.7 W for 1000 W/m^2. This means a relative error of only 1.25% to the reference.

Consequently, the simulation results show that the implemented P&O algorithm is able to extract the maximum power generated by the PV Emulator for both loads
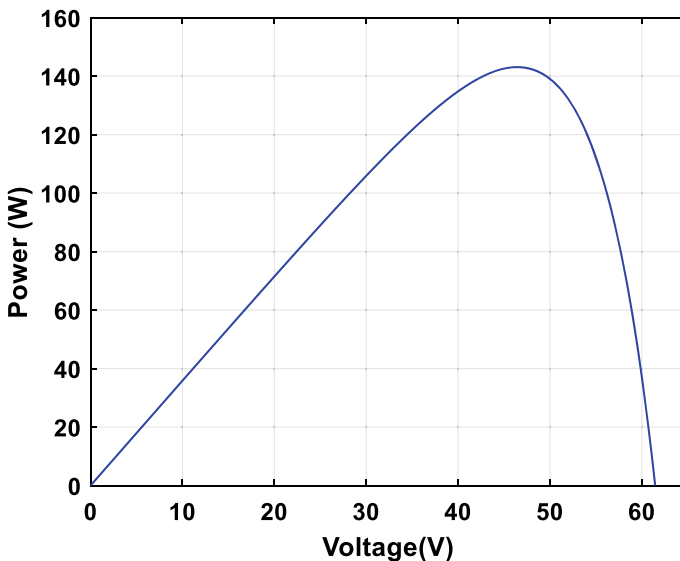
Fig. 6   Reference P–V characteristic implemented in simulation
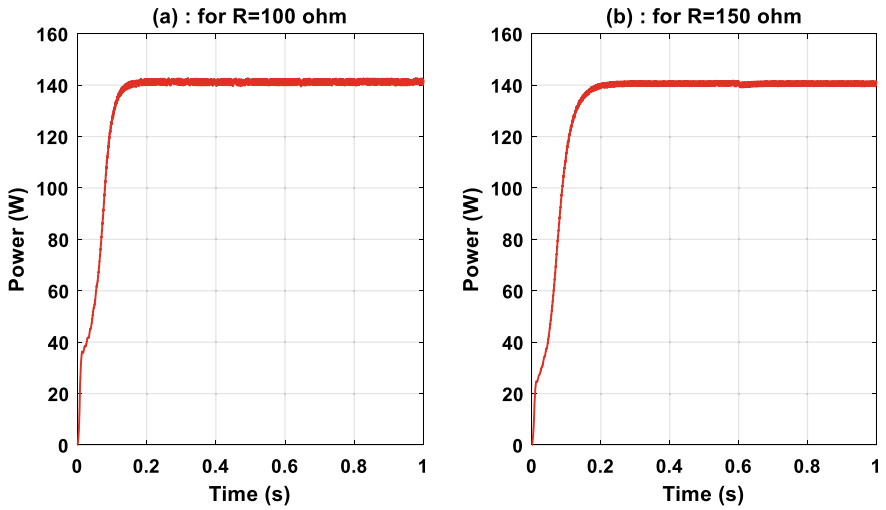
Fig. 7 Power delivered to the load for $R = 100\ \Omega$ and $R = 150\ \Omega$
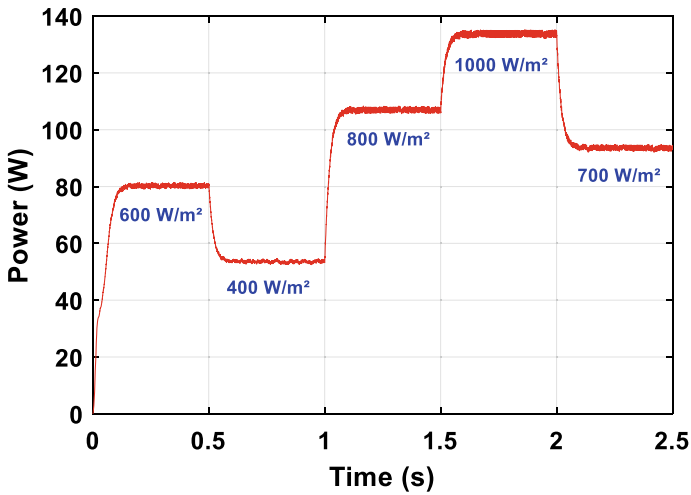
Fig. 8 Power delivered to $R = 100\ \Omega$ for different irradiance levels

$100\ \Omega$ and $150\ \Omega$. It is also clear that the MPPT controller can detect the maximum power following any change of irradiance.

Table 1 Reference MPP and average power generated for different irradiance levels

Irradiance (W/m^2)	MPP of the reference PV characteristic (W)	Mean value of the generated power (W)
400	54.21	53.61
600	81.32	80.36
700	94.87	93.7
800	108.4	107
1000	135.5	133.8

4 Experimental Results

For validation purposes, a PV Emulator feeding an MPPT-controlled DC–DC boost converter (Fig. 9) has been experimentally designed.

The PV Emulator is built around MOSFETs Full-Bridge converter and an HF transformer (Fig. 3), resulting in a power converter with a high efficiency that exceeds 94%. The boost converter (Fig. 4) is designed using the IRF530 power MOSFET and ferrite core-based inductor.

The real-time control part, including the PV characteristic generator and the current loop, is implemented using eZdspF28335 controller board.

Making use of the mathematical model described in Sect. 2.1, the reference P–V characteristic shown in Fig. 10 is deployed into the PV Emulator. Table 2 shows the main electrical properties of this reference PV module.
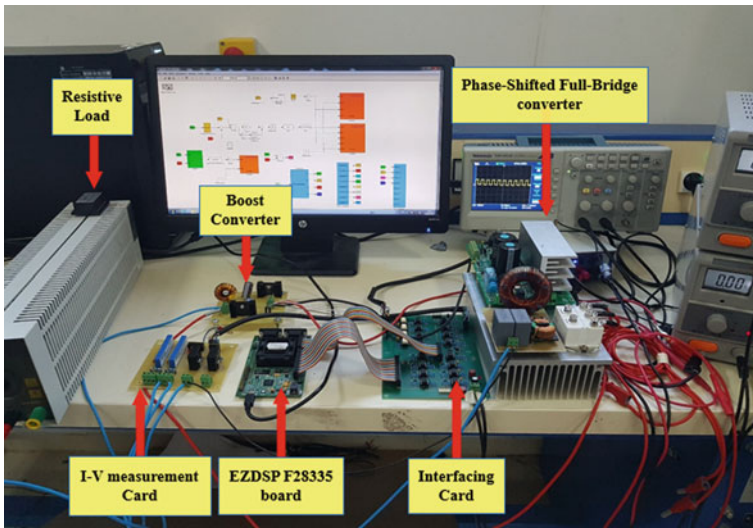
Fig. 9 Photograph of the experimental test bench

Fig. 10 Reference P–V characteristic implemented in real-time
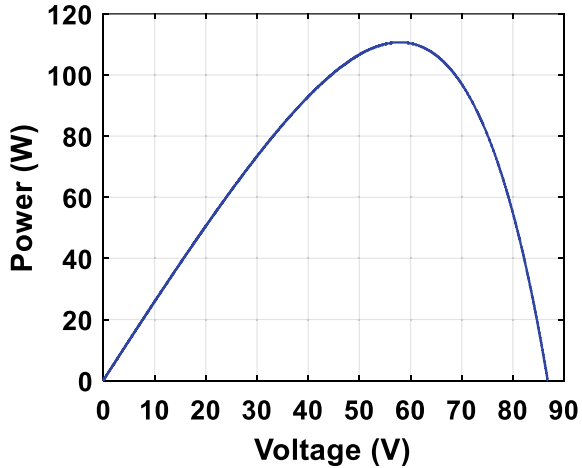
Table 2 Main electrical properties of the reference PV module

Electrical property	Value
P_{MPP}	110.6 W
V_{MPP}	57.97 V
I_{MPP}	1.909 A
V_{oc}	86.78 V
I_{sc}	2.67 A

From Fig. 11, it is clear that after activation of the two power converters at $t1$, the P&O power controller starts looking for the MPP available at the terminals of the PV source represented here by the PV Emulator. The power delivered to the load oscillates around the MPP after a short transient time equal to 0.27 s.

As a result, these findings demonstrate that the built PV Emulator can be used to evaluate MPPT controllers such as P&O algorithm-based DC–DC boost converter.

5 Conclusion

In this paper, the capabilities of the Photovoltaic Emulator in terms of testing MPPT algorithms with significant power level are demonstrated. The test system consists of a PV Emulator feeding a P&O-controlled DC–DC boost converter.

First, the simulated test system with implemented P&O algorithm under Matlab/Simulink shows that for two different loads and varying irradiance, the delivered average power stabilizes in less than 0.23 s. Furthermore, the steady-state power error, relatively to the reference MPP, does not exceed $\pm 1.25\%$. The simulation study was very helpful for PV generator test and current controller design.

Fig. 11 Power delivered to
the resistive load in real-time
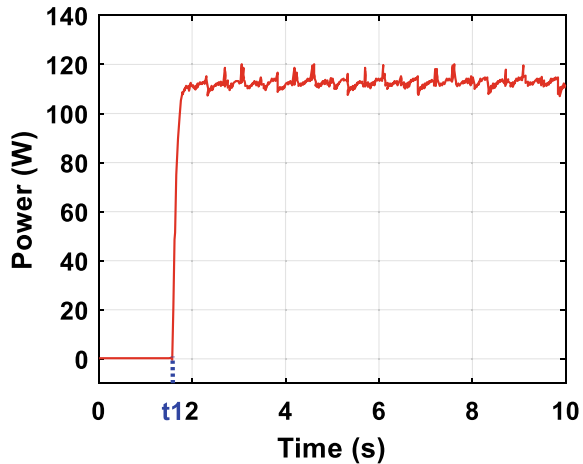
Afterward, the first experimental results of the test system show the high performances of the implemented P&O algorithm which is able to track the MPP in 0.27 s. These findings demonstrated the effectiveness of the designed PV Emulator for testing MPPT-controlled converters.

In perspective, more experimental tests will be conducted on the MPPT controller for different loads and environmental conditions.

References

1. Strasser T, Andren F, Kathan J, Cecati C, Buccella C, Siano P, Leitao P, Zhabelova G, Vyatkin V, Vrba P, Marik V (2015) A review of architectures and concepts for intelligence in future electric energy systems. IEEE Trans Ind Electron 62:2424–2438. https://doi.org/10.1109/TIE.2014.2361486
2. Jayawardana IDG, Ho CNM, Pokharel M, Valderrama GE (2021) A fast-dynamic control scheme for a power-electronics-based PV emulator. IEEE J. Photovoltaics 11:485–495. https://doi.org/10.1109/JPHOTOV.2020.3041188
3. de Paulo AF, Porto GS (2018) Evolution of collaborative networks of solar energy applied technologies. J Clean Prod 204:310–320. https://doi.org/10.1016/j.jclepro.2018.08.344
4. Farcas C, Ciocan I, Tulbure A (2018) Solar emulator for a photovoltaic module. In: 2018 IEEE 24th international symposium for design and technology in electronic packaging (SIITME). IEEE, Iasi, Romania, pp 314–318
5. Chalh A, Motahhir S, El Hammoumi A, El Ghzizal A, Derouich A (2018) Study of a low-cost PV emulator for testing MPPT algorithm under fast irradiation and temperature change. Technol Econ Smart Grids Sustain Energy 3:11. https://doi.org/10.1007/s40866-018-0047-8
6. Kumar N, Hussain I, Singh B, Panigrahi BK (2018) Framework of maximum power extraction from solar PV panel using self predictive perturb and observe algorithm. IEEE Trans. Sustain. Energy. 9:895–903. https://doi.org/10.1109/TSTE.2017.2764266

7. Verma D, Nema S, Shandilya AM, Dash SK (2016) Maximum power point tracking (MPPT) techniques: recapitulation in solar photovoltaic systems. Renew Sustain Energy Rev 54:1018–1034. https://doi.org/10.1016/j.rser.2015.10.068
8. Motahhir S, El Hammoumi A, El Ghzizal A (2020) The most used MPPT algorithms: review and the suitable low-cost embedded board for each algorithm. J Clean Prod 246:118983. https://doi.org/10.1016/j.jclepro.2019.118983
9. Ishaque K, Salam Z, Lauss G (2014) The performance of perturb and observe and incremental conductance maximum power point tracking method under dynamic weather conditions. Appl Energy 119:228–236. https://doi.org/10.1016/j.apenergy.2013.12.054
10. Femia N, Petrone G, Spagnuolo G, Vitelli M (2005) Optimization of perturb and observe maximum power point tracking method. IEEE Trans Power Electron 20:963–973. https://doi.org/10.1109/TPEL.2005.850975
11. Das C, Mandal K, Roy M (2020) Design of PV emulator fed MPPT controlled DC-DC boost converter for battery charging. In: 2020 IEEE first international conference on smart technologies for power, energy and control (STPEC). IEEE, Nagpur, India, pp 1–6
12. Chaker M, Yousfi D, Hajji B, Kourchi M, Ajaamoum M, Belarabi A, Abd Rahim N, Selvaraj J (2020) Design and implementation of a photovoltaic emulator using an insulated full bridge converter based switch mode power supply. In: Proceedings of the 2nd international conference on electronic engineering and renewable energy systems. Springer Singapore, Singapore, pp 531–541

Introduction of a Correction Factor for Predicting Real Stirling Engine Efficiency Based on Ideal Adiabatic Modelling

Kaoutar Laazaar and Noureddine Boutammachte

Abstract Stirling engine presents an excellent opportunity for power production based on renewable energies thanks to its multi-fuel capacity, high efficiency, silent operation, low maintenance requirement and ecological suitability. Previous studies focused on developing mathematical models for predicting real Stirling engine performance which is based on complex equations. However, there is no simple method that enables the estimation of real engine outputs without the necessity to add many equations into modelling. In the present paper, we introduce a new correction factor for predicting the real Stirling engine efficiency for its three types (Alpha, Beta and Gamma) based on ideal modelling only. This will help Stirling engine researchers to estimate the practical engine performances quickly based only on the correction factor without the necessity to add complex equations. To do this, we have used the non-ideal adiabatic model because it has been regarded as one of the high accurate models among second order approaches since it can provide accurate results with a reasonable time. The results demonstrate that Stirling engines can produce high work output and efficiency close theoretically to that of Carnot if the engine is well designed. In addition, it was found that the consideration of pressure drop loss reduces the output power by 18.92% and decreases the thermal efficiency by 11.51% for Alpha type Stirling engine. It was also concluded that the non-ideal recovery of regenerator loss has the second effect on Stirling engine performances; its account reduces the efficiency by 10.69%. Moreover, a correction factor of 0.66, 0.52 and 0.53 has been found for predicting the real efficiency of Alpha, Beta and Gamma types, respectively.

Keywords Stirling engine · Renewable energies · Correction factor · Adiabatic modelling · Efficiency prediction

K. Laazaar (✉) · N. Boutammachte
2ER Team, Energy Department, ENSAM, Moulay Ismail University, Meknes, Morocco
e-mail: Kaoutarlaazaar1992@gmail.com

1   Introduction

The excessive demand for energy, the decrease of fossil fuels stocks and the greenhouse gases emissions rise are the main factors for looking for other alternative solutions. One of the best promising solutions is the use of renewable energies with thermo-mechanical conversion systems such as Stirling engines technology. Stirling engine (SE) is one of the performant technologies that can convert any source of heat input into mechanical or electrical power. It is a type of heat engine in which the principal of operation is based on a closed regenerative thermodynamic cycle [1]. The particularity of Stirling engines is their ability to operate with an external combustion unlike other thermal motors [2]. This advantage allows the use of several heat sources including renewable energies. In fact, Stirling engines are constructed with different mechanical configurations namely Alpha, Beta and Gamma. The three configurations have the same thermodynamic cycle but are different in their mechanical construction [1].

A number of researchers in previous studies have sought to determine the real performance of Stirling engines including its output power and efficiency based on mathematical modelling. Udeh et al. [3] have developed governing differential equations of Stirling engines by coupling some additional losses into the traditional model. They have found an optimum mean effective pressure for maximum energetic efficiency of the engine. Cheng et al. [4] have proposed an energy method for solving the relationship between the crank angle of the main shaft, and the work generated by the working gas for an Alpha type four-cylinder double-acting Stirling engine. Their findings have revealed that the maximum shaft power of the proposed engine is 1103 W at 878 rpm at a heating temperature of 1200 K with a loading torque of 12 N m.

Most studies in the existing literature have only focused on modelling ideal model and integrating the different losses occurring inside Stirling engine. These approaches require solving complex mathematical equations and take a long time in numerical simulations. Although extensive research has been carried out on evaluating Stirling engine output power and efficiency, far too little attention has been paid to predicting the real Stirling engine performances without the necessity to add complex equations into the models. In the present paper, we propose a correction factor for estimating the real Stirling engine performance for Alpha, Beta and Gamma types based on ideal adiabatic modelling only. This will help the researchers in the field to predict practical engine performances in a quick manner while providing accurate results.

2   Principle of Operation and Mathematical Model

Theoretically, the cycle of Stirling engine can be the most efficient device converting heat into mechanical work since its efficiency can be equal to that of Carnot cycle which is the most efficient thermodynamic cycle. Stirling engine cycle consists of four

Table 1 Set of mathematical equations

$dT_e =$ $T_e\left(\frac{dP}{P} + \frac{dV_e}{V_e} - \frac{dm_e}{m_e}\right); dT_c =$ $T_c\left(\frac{dP}{P} + \frac{dV_c}{V_c} - \frac{dm_c}{m_c}\right)$ (1)	Temperatures in compression and expansion spaces
$dQ_r =$ $\frac{V_r dP c_v}{R} - c_p(T_{kr}m_{kr} - T_{rh}m_{rh})$ (2)	Heat transfer in regenerator
$dW = dW_e + dW_c$ (3) $\eta_{adi} = \frac{W}{Q_h}$ (4)	Network output and thermal efficiency
$dp_i = \frac{2f_i\mu V_i GL_i}{m_i d_i^2}$ (5)	Pressure drop in heat exchangers
$Q_{r\text{loss}} =$ $(1 - \varepsilon)(Q_{r\max} - Q_{r\min}).f_r$ (6)	Non-ideal recovery of the regenerator

thermodynamic processes (isothermal compression, isochoric heating, isothermal expansion and isochoric cooling) [5].

The mathematical model used in this research is the non-ideal adiabatic model. It consists of the division of the engine into five control volumes including the compression space, expansion space, heater, cooler and regenerator [6]. The set of mathematical equations is summarized in Table 1. More details can be found in [1].

3 Results and Discussions

3.1 Model Validation

In order to validate the non-ideal adiabatic model used in the present paper, the geometrical and physical specifications of the GPU-3 Stirling engine manufactured by General Motors Corporation were chosen for the model validation. Figure 1 shows the comparison between our model, the experimental results of NASA Lewis Research Center and previous numerical model. The current model presents better performance in comparison with the simple analysis [7] because in the later, less mechanical losses are considered. The output power of our model is closer to the experimental results because in the present model, several mechanical losses are taken into consideration. The slight difference between the experimental findings and the current analysis could be attributed to the frictional losses of the engine crank angle mechanism. Furthermore, the pistons movement in the adiabatic approach is simplified to sinusoidal volume variations. Nevertheless, it is very difficult to achieve this assumption in practical engines since the piston's movement is quasi sinusoidal and depends on the crank shaft mechanism.
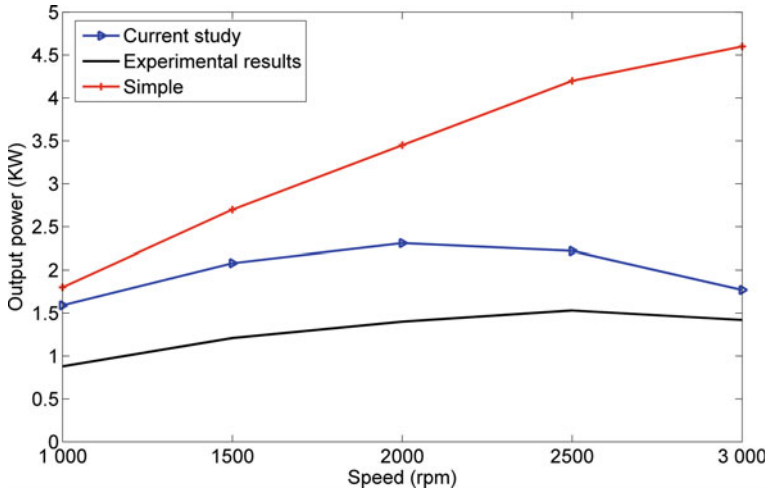
Fig. 1 The effect of speed on GPU-3 stirling engine power

Table 2 Comparison between current model, previous thermal models and experiment [7]

Model	Power (KW)	Efficiency (%)	Power error (%)	Efficiency error (%)
Ideal adiabatic model	8.30	62.30	176.70	41.00
Finite speed thermodynamic model	4.80	29.27	60.00	7.97
Timoumi model	4.27	38.50	42.30	17.20
Developed analytical isothermal model	6.09	52.90	103.00	31.60
Third order analysis	4.26	42.00	42.00	20.70
Current study	3.90	24.19	30.00	2.89
Experimental results	3.00	21.30	–	–

Another comparative study is given in Table 2. The present model demonstrates high accuracy in comparison with previous models thanks to the consideration of different mechanical and thermal losses.

3.2 Analysis of Ideal Adiabatic Model and Losses Assessment

Figure 2 presents the variations of pressure and volume inside the expansion and compression spaces of Alpha type SE during an engine cycle at 1500 rpm rotational speed.
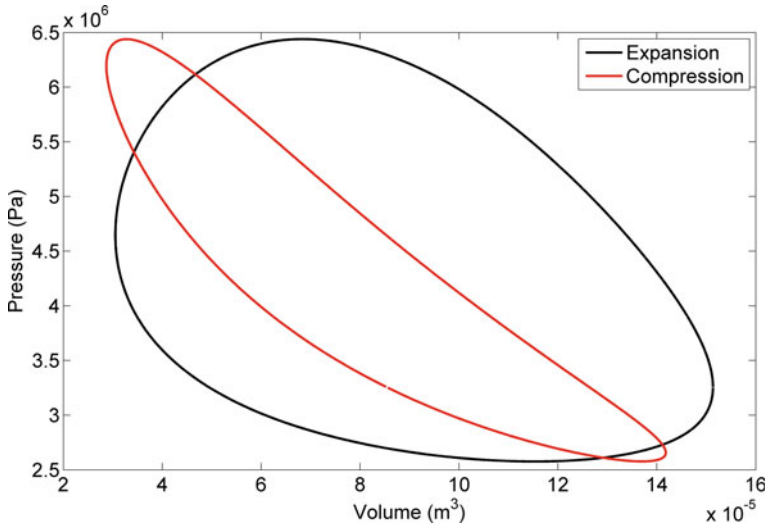
Fig. 2 Pressure–volume diagram of ideal model

The integration of each curve allows obtaining the work generated and consumed by the expansion and compression processes, respectively. A large area of work output can be obtained in ideal model as no loss is considered. In ideal operating conditions, the expansion and compression works are being equal to 336.25 J and 131.59 J, respectively, which results in a network output of 204.66 J with an ideal thermal efficiency of 60.86%.

Figure 3 depicts the variation of heat flow inside Stirling engine heat exchangers. In the ideal adiabatic model, the regenerator is assumed to be perfect (effectiveness = 1), which means that the heat absorbed by the matrix is equal to the heat released to the working gas and no thermal or mechanical loss occur during the engine cycle. It must also be noted that the heat transferred to the heater during an engine cycle is equal to 336.26 J which is the same as the work done by the expansion process; whereas the heat rejected to the cooler is equal to 131.74 J which is the same as the work consumed by the compression process.

The pressure drop inside Stirling engine heat exchangers is one of the mechanical losses which takes place in practical engines. Its effect on output power for Alpha type Stirling engine is studied in Fig. 4. The non-ideal model deviates significantly from the ideal one while integrating the pressure drop effect especially in higher speed values in which the frictional losses undergo an important increase. At 3500 rpm rotational speed, the output power produced in ideal case is equal to 11,938.73 W while this value does not exceed 5022.88 W when the pressure drop loss is considered. The large difference between the two models is due to the decrease of engine pressure knowing that this parameter is very important in Stirling engine operation since the work output is defined as the product of pressure and volume. Consequently, the
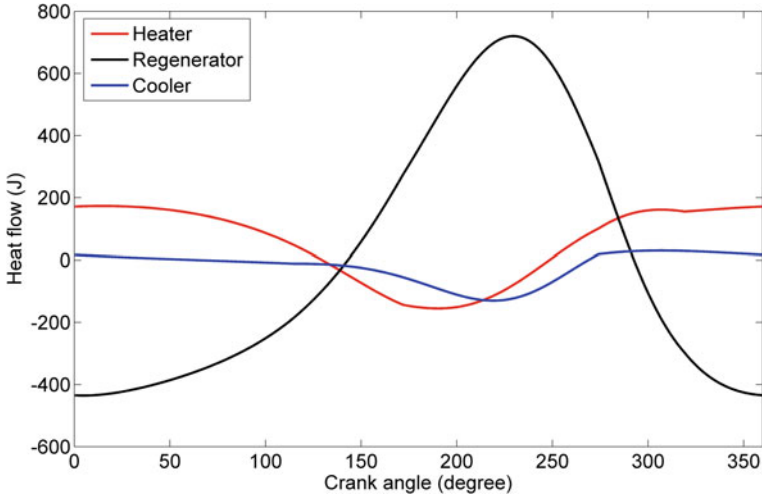
Fig. 3 Heat flow rate inside stirling engine heat exchangers in the ideal model

decrease in pressure leads to the work output decrease, hence the reduction of output power.

The effect of regenerator non-ideal recovery loss on ideal model for different hot temperature values of Alpha configuration Stirling motor is shown in Fig. 5. The consideration of non-ideal recovery of the regenerator significantly reduces the engine thermal efficiency.

For instance, the thermal efficiency decreases by more than 13% at 1200 K heater temperature. This is because the regenerator, in real case, is non-perfect hence its
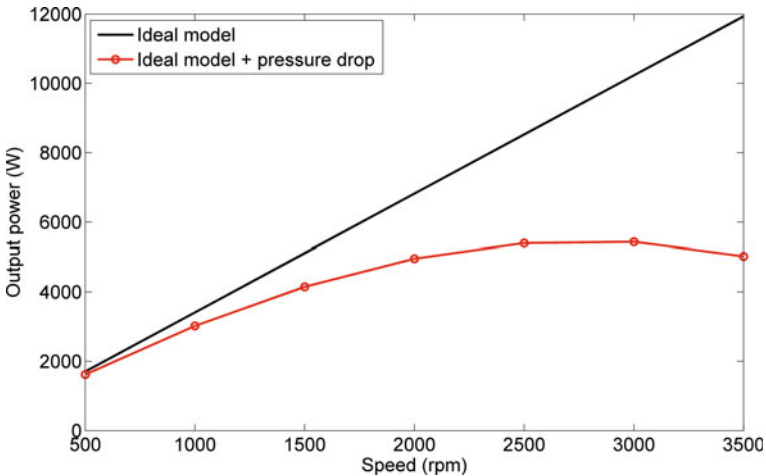
Fig. 4 The effect of pressure drop loss on ideal adiabatic model
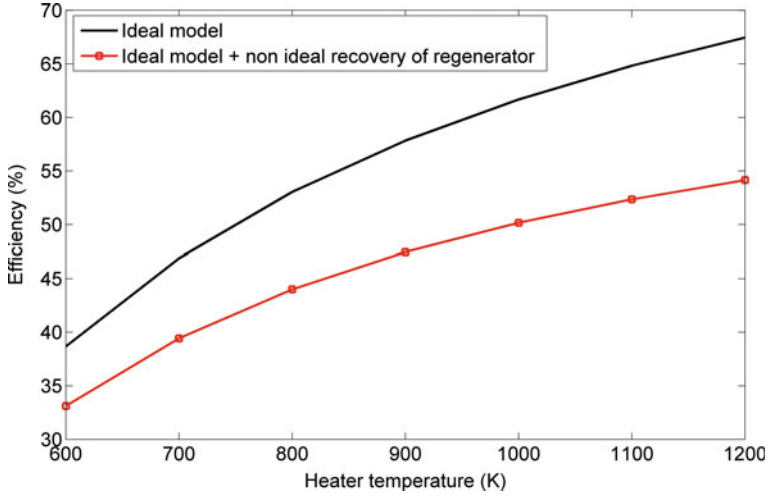
Fig. 5 The effect of non-ideal recovery of regenerator on ideal model

effectiveness is less than 1. As a result, a huge amount of heat transferred to the heater is not absorbed by the regenerator leading to the decrease of the engine efficiency.

The shuttle loss contributes to the decrease of the expansion space temperature and the increase of the compression one. Consequently, more heat input must be added in the heater which increases the energy cost of Beta and Gamma Stirling engines. The requirement of more heat input due to shuttle loss is depicted in Fig. 6.
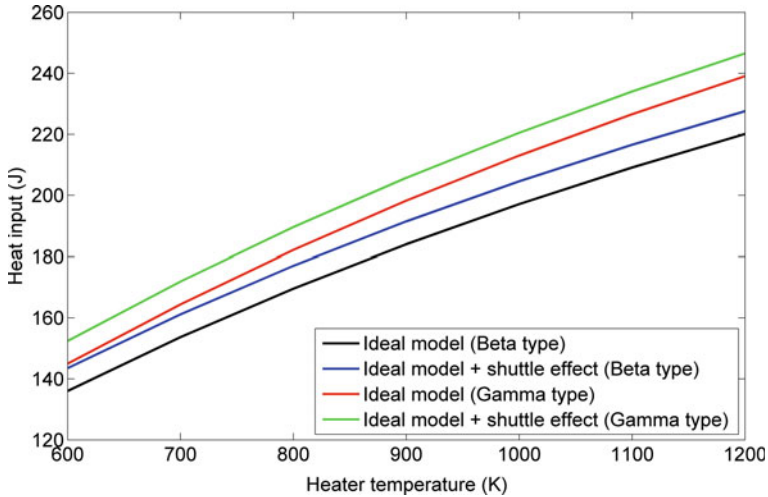
Fig. 6 Effect of shuttle loss on heat input of Beta and Gamma types stirling engine

It must be noted that the prediction of real output power and thermal efficiency of Stirling engine types from the ideal adiabatic model must consider a correction factor which takes all the thermal and mechanical losses into account. The correction factor resulted from our analysis can be calculated as $K =$ real efficiency/ideal efficiency.

The value of the correction factor for Alpha, Beta and Gamma types is 0.66, 0.52 and 0.53, respectively.

4 Conclusions

This paper presents the ideal adiabatic model which has been used for estimating the ideal performances of different Stirling engines configurations namely Alpha, Beta and Gamma. The results of the ideal analysis show that Stirling engines can provide high performances in terms of output power and efficiency. Afterwards, the thermal and mechanical losses have been integrated into the ideal model to predict the real outputs of Stirling engine types. The losses have been added separately into the model to evaluate their effect on Stirling engine performances. It has been found that the pressure drop has the important effect on Stirling engine output power and efficiency. The non-ideal recovery of regenerator loss has the second effect on Stirling engine performances; its account reduces the efficiency by 10.69%. To help Stirling engine researchers to predict the output performances, we have proposed a correction factor for each of Stirling engine types namely Alpha, Beta and Gamma.

The introduction of this factor will help on the assessment of real Stirling engine efficiency by calculating only the ideal one without the necessity to add complex equations into the analysis.

References

1. Laazaar K, Boutammachte N (2020) New approach of decision support method for Stirling engine type choice towards a better exploitation of renewable energies. Energy Convers Manage 223:113326. https://doi.org/10.1016/j.enconman.2020.113326
2. Ipci D, Karabulut H (2018) Thermodynamic and dynamic analysis of an alpha type Stirling engine and numerical treatment. Energy Convers Manage 169:34–44. https://doi.org/10.1016/j.enconman.2018.05.044
3. Udeh GT, Michailos S, Ingham D, Hughes KJ, Ma L, Pourkashanian M (2020) A new non-ideal second order thermal model with additional loss effects for simulating beta Stirling engines. Energy Convers Manage 206:112493. https://doi.org/10.1016/j.enconman.2020.112493
4. Cheng C, Yang H, Tan Y (2022) Theoretical model of a a-type four-cylinder double-acting Stirling engine based on energy method. Energy 238:121730. https://doi.org/10.1016/j.energy.2021.121730
5. Laazaar K, Boutammachte N (2019) The determination of adequate type of Stirling engine for cogeneration in industrial sector. Int J Mech Prod Eng Res Dev 9:617–26. https://doi.org/10.24247/ijmperdfeb201960

6. Laazaar K, Boutammachte N (2020) Modelling and optimization of stirling engine for waste heat recovery from cement plant based on adiabatic model and genetics algorithms. In: International conference on artificial intelligence & industrial applications, Springer, pp 287–96
7. Laazaar K, Boutammachte N (2022) Development of a new technique of waste heat recovery in cement plants based on Stirling engine technology. Appl Therm Eng 210:118316. https://doi.org/10.1016/j.applthermaleng.2022.118316

Enhancement of Standalone PMSG Wind Turbine System Utilizing Nonlinear Proportional-Integral Control Technique

Ahmed Omar Elgharib, Soufyane Benzaouia, and Aziz Naamane

Abstract This paper studies the control of a wind turbine connected system using permanent magnet synchronous generator (PMSG) driven by a wind turbine and how to achieve the maximum power point tracking (MPPT) using nonlinear proportional-integral controller (NPIC), and also, such a controller permits the PMSG to operate at an optimal speed and high efficiency, especially by utilizing PMSG. This wind turbine is connected to the load by means of controlled rectifier. The controlled rectifier is used for transferring power to the load. A variable wind speed profile is used in such a system. This paper shows the dynamic performance of the complete system using Matlab Simulink.

Keywords Wind turbine stand alone system · Wind turbine modelling · Nonlinear control

1 Introduction

The last two decades have seen a few vital improvements within the plan of feedback controllers for solid shape behaviour settlement (set-point control) and move following targets. [1, 2] In any case, there are certain open issues still stay in this subject that are of extraordinary hypothetical and viable interest. In specific, from the point of view of outside unsettling influence torque rejection, no bound together system for planning basic control structures available right now that is motivated from linear control hypothesis such as proportional-integral (PI), proportional-integral-derivative (PID), and their variations.

This outcome employs the Euler parameter (quaternion) kinematics and a lack of involvement channel whose situation empowers development of position stabi-

A. O. Elgharib (✉) · S. Benzaouia · A. Naamane
Aix Marseille University, Université de Toulon CNRS LIS, Marseille, France
e-mail: ahmedomarahmed@aast.edu; ahmedomr_eg@hotmail.com

A. O. Elgharib
Arab Academy for Science, Technology, and Maritime Transport, Giza, Egypt

lizing controller. Structure of channel vector state is appeared to be administered across steady first-order linear differential condition carried out by vector portion of quaternion (position estimation). Furthermore, lack of involvement channel does not have high-pass features and in hypothesis concedes subjectively moderate transmission capacity. Moreover, for adequately less frequencies within filter's input flag, channel yield approximates a pseudo-velocity-like state. In any case, this false for input signals with higher size of frequencies. Ensuing expansions to this velocity-free controller system were displayed by [3]. Energize generalizations to the case of position tracking along endorsed reference directions were presented by [4, 5]. In a later advancement, [5] is drawn closer to the position stabilization issue from nonlinear PID control point of view in which the full-state vector counting precise speed was utilized for criticism purposes. In spite of the fact that control plot requires full-state criticism when compared to the lack of involvement channel-based developments, this result brings valuable arrangement for necessarily input. In terms of strength with regard to idleness parameter instabilities, PID development of [5] is nearly autonomous of dormancy network within sense that it certainly needs earlier information on the biggest eigen value of idleness network. Position control with universally steady closed-loop elements utilizing Euler parameters is moreover examined in [6]. This work combines viewpoints of details for passivity filter in conjunction with the choice of a Lyapunov work containing cross/mixed terms including the conditions. Nearby unsettling influence dismissal is outlined through a linearization approximately the ostensible harmony point. Wind turbines can generate electric force by using the wind intensity in order to manage electric generator [7]. Generator can originate power and shifts from crest to attainable transformer and changes from output voltage to individual utilization or an across the nationwide grid [8]. Effectiveness and capability for wind farms have huge influence on wind power that can alter with large scope towards variable geographical areas [9]. The coherence of the ability to produce energy is the key distinction between the onshore and offshore wind farms. Offshore wind farms have the capacity to deliver more electricity at a steadier rate than their onshore partners [10], and the power circuit shown in Fig. 1 is studied firstly in [11].

This paper is organized as follows: Sect. 1 is the introduction; Sect. 2 discusses the modelling fo the WECS; Sect. 3 proposes the control of the PMSG side converter; Sect. 4 shows the results and its discussion, and Sect. 5 is the conclusion.

1.1 *Nomenclature*

PMSG: Permanent Magnet Synchronous Generator, NPIC: Nonlinear Proportional-Integral Controller, MPPT: Maximum Power Point Tracking, WECS: Wind Energy Conversion System.
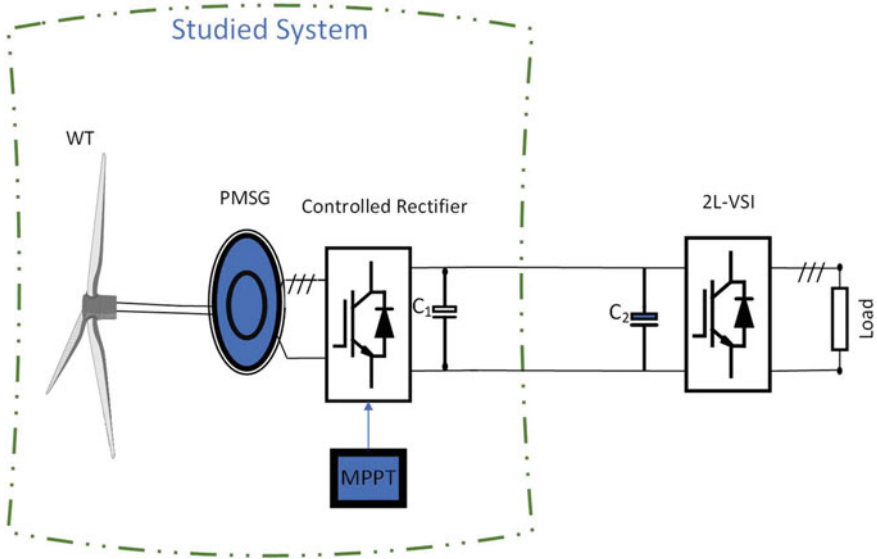
Fig. 1 Schematic diagram for WECS

2 Wind Energy Conversion System Modelling

2.1 Model of the Turbine

The wind power is described by:

$$p_v = \frac{\left(\rho \cdot A \cdot v_w^3\right)}{2} \tag{1}$$

where v_w is the wind speed, A is the circular area, and ρ determines density of air. The mechanical power that can be evaluated by the turbine is as follows:

$$p_m = \frac{1}{2} C_p(\lambda, \beta) \cdot \rho \cdot A \cdot v_w^3 \tag{2}$$

$$\lambda = \frac{\Omega_m R}{v_w} \tag{3}$$

where C_p indicates to the power coefficient, it is concerned with the pitch angle β and the tip-speed ratio λ. R indicates to turbine radius and Ω_m is the rotor speed for turbine. The coefficient of power C_p is expressed by the following two equations:

$$c_p(\lambda, \beta) = c_1 \left(\frac{C_2}{\gamma} - C_3 \cdot \beta - C_4 \right) e^{\frac{-C_5}{\gamma}} + c_6 \cdot \lambda \tag{4}$$

$$\frac{1}{\gamma} = \frac{1}{\lambda + 0.08\beta} - \frac{0.035}{\beta^3 + 1} \tag{5}$$

where $C_1 = 0.5176, C_2 = 116, C_3 = 0.4, C_4 = 5, C_5 = 21, C_6 = 0.0068$. The main model shows a changeable pitch model. However, for this situation, the model is varied to clarify the firmed pitch turbine by reaching to zero according to pitch angle. Therefore, the coefficient of power is only utilized as lambda function. The mechanical torque generated by the turbine is determined by this equation:

$$T_m = \frac{p_m}{\Omega_m} = \frac{1}{2 \cdot \Omega_m} C_p(\lambda, \beta) \cdot \rho \cdot A \cdot v_w{}^3 \tag{6}$$

As the wind turbine is linked directly with permanent magnet synchronous generator (PMSG), the shaft system of the WECS can be clarified by a one-mass model. The motion equation is represented by the following:

$$T_m = J \frac{d\Omega_m}{dt} + f \cdot \Omega_m + T_{em} \tag{7}$$

$$J = J_{\text{turbine}} + J_a \tag{8}$$

where J is the overall inertia moment for rotating parts, f is the friction coefficient, and T_m is the electromagnetic torque obtained by generator.

2.2 Model of PMSG

The d–q voltages of stator for the generator are introduced by the next equations:

$$V_{ds} = R_s I_{ds} + L_d \frac{dI_{ds}}{dt} - \omega_r \psi_{qs} \tag{9}$$

$$V_{qs} = R_s I_{qs} + L_q \frac{dI_{qs}}{dt} + \omega_r \psi_{ds} \tag{10}$$

$$\psi_{ds} = L_d I_{ds} + \psi_0 \tag{11}$$

$$\psi_{qs} = L_q I_{qs} \tag{12}$$

where L_d and L_q are the inductance of the generator on the d and q axis respectively, R_s is resistance of the stator, ω_r is the electrical rotating speed of the PMSG that is introduced by
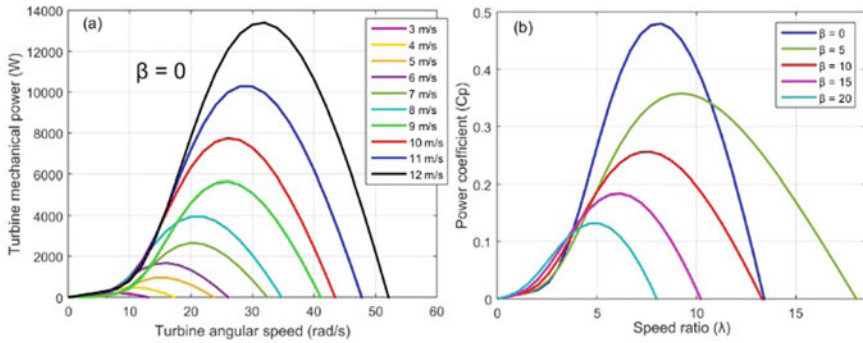
Fig. 2 a Wind curves at different wind speeds, **b** features C_p versus λ for various ratings for pitch angle β

$$\omega_r = p \cdot \Omega_m, \tag{13}$$

and ψ_0 is the permanent magnetic flux. By substituting Eqs. (11) and (12) into Eqs. (9) and (10), the differential equations of PMSG are presented by the following:

$$
\begin{cases}
L_d \dfrac{dI_{ds}}{dt} = V_{ds} - R_s I_{ds} + \omega_r L_q I_{qs} & (13) \\[2mm]
L_q \dfrac{dI_{qs}}{dt} = V_{qs} - R_s I_{qs} - \omega_r L_d I_{ds} - \psi_0 \omega_r & (14)
\end{cases}
$$

The electromagnetic torque is explained by:

$$T_{em} = \frac{3}{2} p \left[\left(L_d - L_q \right) I_{ds} I_{qs} + \psi_0 I_{qs} \right] \tag{15}$$

where p is the number of pole pairs. The PMSG is supposed to be wound-rotor, then $L_d = L_q$, and the term of the electromagnetic torque in the rotor is presented as shown below (Fig. 2):

$$T_{em} = \frac{3}{2} p \psi_0 I_{qs} \tag{16}$$

3 PMSG Side Converter Control

A vector control technique is used in the AC/DC converter so as to control the speed of the generator Fig. 3.

The speed of the generator (ω_m) can be constrained by changing the electromagnetic torque (T_m) to its reference T_{em}^*. This is happened by acting on the q-axis current I_{qs} utilizing the equation:
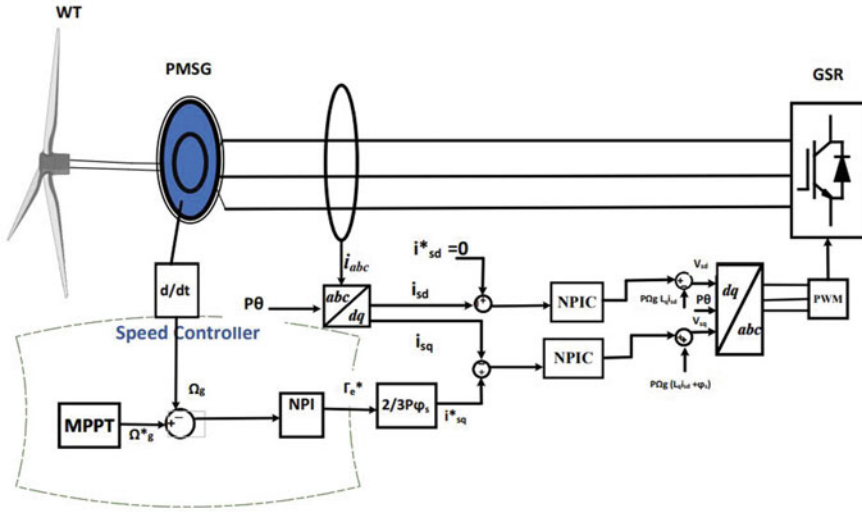
Fig. 3 Control system strategy based on NPIC

$$I_{qs}^* = \frac{2}{3p\psi_0} T_{em}^*$$
(17)

The stator current of the d-axis I_{ds} part is limited to zero to accomplish the optimized torque of the generator. The rotational velocity for the maximum reference can be determined utilizing

$$\Omega_{\text{mopt}} = \frac{\lambda^* \cdot v_W}{R}$$
(18)

λ^* addresses the optimum tip-speed ratio. The power circuit shown in Fig. 3 discusses the control technique for such a system across the MPPT utilizing three online NPIC controllers on such a system.

3.1 Nonlinear Control Law

It is apparent from nonlinear control law that the portion of exponential one head for moving close to zero, in case either of the two factors "ε" and "r" is significant, making the expression tends to "1," and there will be no serious variance will happen within the integral gain. This can be required as well in the event that either "ε" or "r" is significant then the integral gain required to be constrained and ought to be adjusted as it is. Taking into consideration a case where, "ε" is near to zero but "r" is significant. It is didactic to seem that how alpha changes with respect to 'ε' and 'r'.

3.2 Inner Loop

For regulating the components of current I_{ds} and I_{qs} to the reference belong to them (i_{sd}^* and i_{sq}^*). The error expression can be illustrated as follows:

$$\begin{cases} \varepsilon_{I_{ds}}(t) = I_{ds}{}^*(t) - I_{ds}(t) & (19) \\ \varepsilon_{I_{qs}}(t) = I_{qs}{}^*(t) - I_{qs}(t) & (20) \end{cases}$$

The control voltages of q and d axis are planned utilizing NPIC; they can be introduced as follows:

$$\begin{cases} V_{ds}^*(t) = k_{p_{Ids}}\varepsilon_{I_{ds}}(t) + \alpha_{I_{ds}}(t)k_{i_{I_{ds}}}\int \varepsilon_{I_{ds}}(t)\mathrm{d}t - p\Omega_m L_q I_{qs} & (21) \\[2mm] V_{qs}^*(t) = k_{p_{Iqs}}\varepsilon_{I_{qs}}(t) + \alpha_{I_{qs}}(t)k_{i_{I_{qs}}}\int \varepsilon_{I_{qs}}(t)\mathrm{d}t + p\Omega_m (L_d I_{ds} + \psi_0) & (22) \end{cases}$$

with

$$\begin{cases} \alpha_{I_{ds}}(t) = \left[1 - \exp\left\{-\left(a\bar{\varepsilon}_{I_{ds}}^2(t) + br_{I_{ds}}^2(t)\right)\right\}\right]^2 & (23) \\ \alpha_{I_{qs}}(t) = \left[1 - \exp\left\{-\left(a\varepsilon_{I_{qs}}^2(t) + br_{I_{qs}}{}^2(t)\right)\right\}\right]^2 & (24) \end{cases}$$

3.3 Outer Loop

The error expression for the speed controller is given by:

$$\varepsilon_{\Omega_m}(t) = \Omega_{\text{mopt}}(t) - \Omega_m(t) \tag{25}$$

The outer control is utilized to regulate rotor speed to its reference and provides reference current of q-axis to the inner current loop. Speed controller is designed using nonlinear proportional-integral controller version to allow dealing with highly nonlinear behaviour of wind turbines. The NPIC speed controller is given as follows

$$\begin{cases} I_{qs}{}^* = \dfrac{2}{3p\psi_0}\left(k_{p_o}\varepsilon_{\Omega_m}(t) + \alpha_o(t)k_{i_o}\int \varepsilon_{\Omega_m}(t)\mathrm{d}t\right) & (26) \\[2mm] \alpha_o(t) = \left[1 - \exp\left\{-\left(a\varepsilon_{\Omega_m}{}^2(t) + br_{\Omega_m}{}^2(t)\right)\right\}\right]^2 & (27) \end{cases}$$

where Kp_0 and Ki_0 represent, respectively, proportional and integral gains of NPIC, alpha (node) is the modification factor and $I_{qs}{}^*$ is the reference q-axis current provided to inner current loop.

4   Results and Discussion

This section shows the results of such a system utilizing nonlinear proportional-integral controller (NPIC) on variable wind speed profile, power coefficient according to wind turbine and its efficiency, and the theoretical maximum mechanical power.

Figure 4 shows the variable wind speed profile with the time given in the system, in order to show the range of working on such a system.

Figure 5 indicates the maximum power coefficient for such a system which is around 0.48.
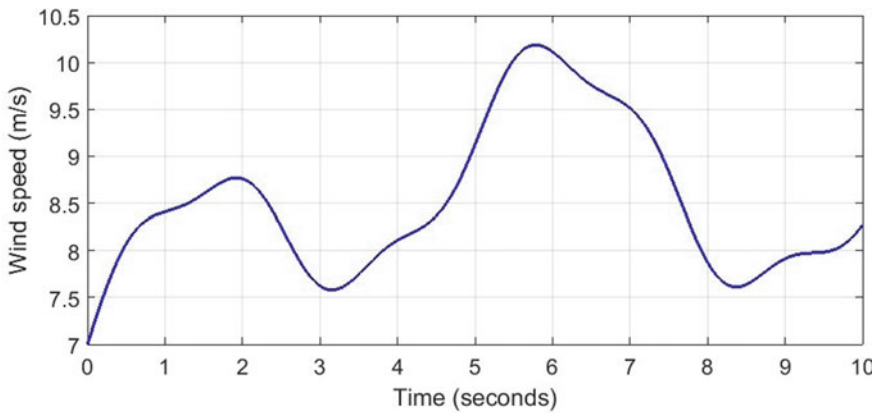
Fig. 4 Wind speed variation
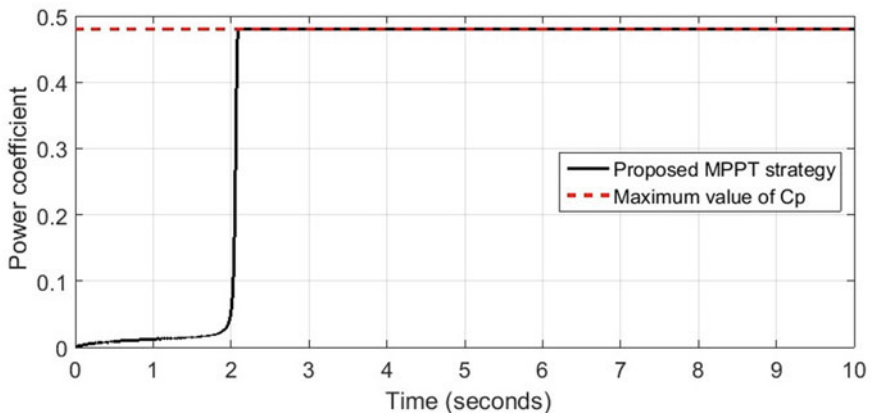
Fig. 5 Power coefficient: C_p
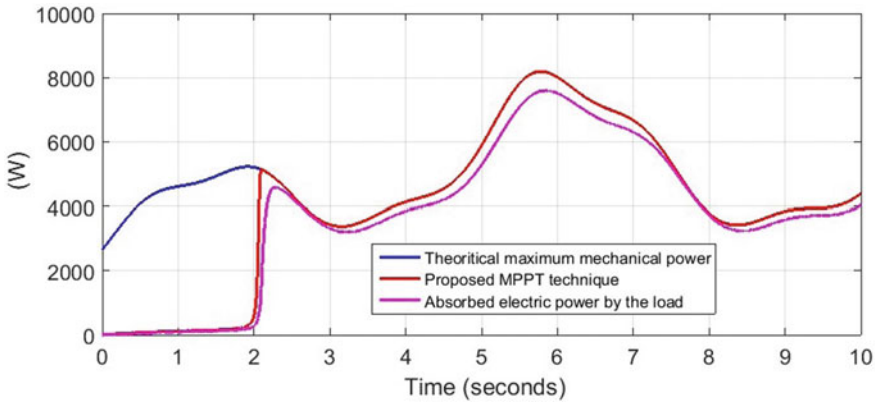
Fig. 6 Ratings of power
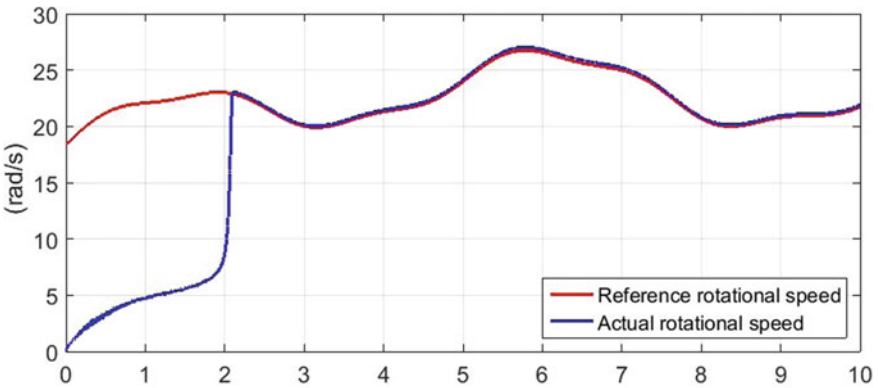
Fig. 7 Rotor speed with its reference

Figure 6 discusses three main parts of the system which are as follows: (1) Theoretical maximum mechanical power, (2) Proposed MPPT technique according to the nonlinear proportional-integral controller(NPIC) and (3) Absorbed electric power by load that indicates losses for both mechanical and converter ones.

Figure 7 shows speed rotor tracking with the reference in order to be followed in the best way.

Figure 8 proves that the q axis and d axis current have the same tracking error; moreover, it shows the speed tracking error of the wind turbine.

Figure 9 identifies the performance of the balanced three phase currents (abc).
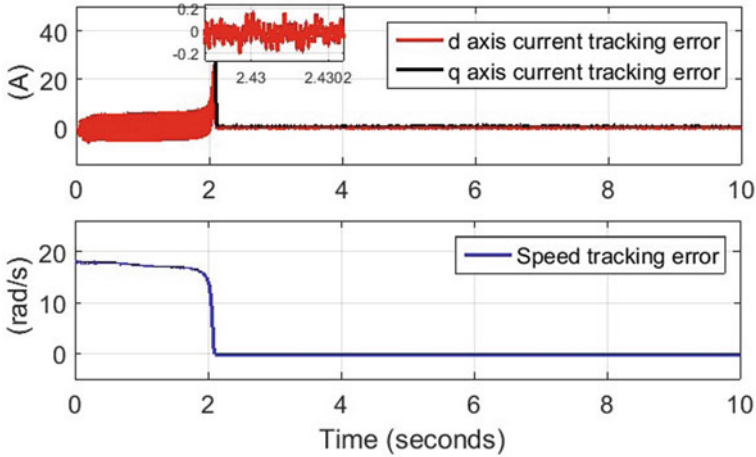
Fig. 8 Tracking errors for overall system
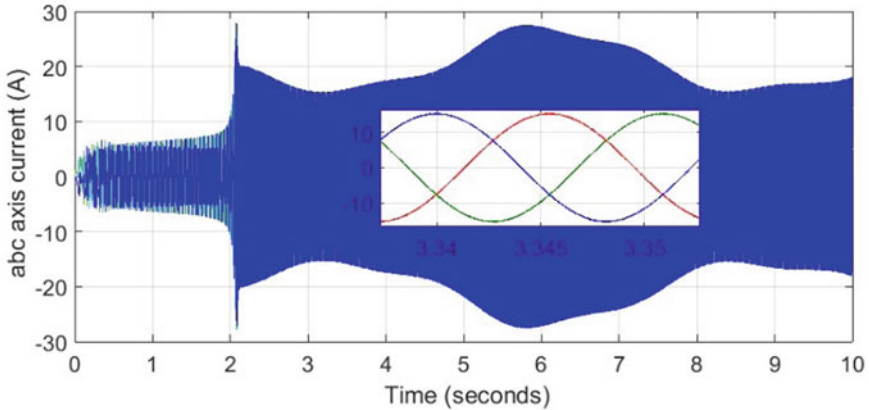
Fig. 9 Generator three phase current

5 Conclusion

A new control technique (NPIC) is utilized in this paper that has good performance by utilizing MPPT technique in a wind turbine connected system for increasing power from the wind, as it is common and well control method with favourable dynamic performance. The PMSG was controlled by the new controller (NPIC) as it assigns a very well dynamic performance and improves the influence and validate research methods on wind turbine system. Moreover, the utilized control methods can take advantage of high efficiency, particularly by using the PMSG. The standalone wind turbine system has been validated by simulation results using Matlab Simulink environment.

References

1. Junkins JL, Turner JD (2012) Optimal spacecraft rotational maneuvers. Elsevier
2. Schaub H, Junkins JL (2003) Analytical mechanics of space systems, AIAA education series. Reston: AIAA
3. Tsiotras P (1998) Further passivity results for the attitude control problem. IEEE Trans Autom Control 43(11):1597–1600
4. Akella MR (2001) Rigid body attitude tracking without angular velocity feedback. Syst Control Lett 42(4):321–326
5. Subbarao K, Akella MR (2004) Differentiator-free nonlinear proportional-integral controllers for rigid-body attitude stabilization. J Guid Control Dynam 27(6):1092–1096
6. Paielli RA, Bach RE (1993) Attitude control with realization of linear error dynamics. J Guid Control Dynam 16(1):182–189
7. Wenehenubun F, Saputra A, Sutanto H (2015) An experimental study on the performance of savonius wind turbines related with the number of blades. Energy Proc 68:297–304
8. Zheng CW, Li CY, Pan J, Liu MY, Xia LL (2016) An overview of global ocean wind energy resource evaluations. Renew Sustain Energy Rev 53:1240–1251
9. Possner A, Caldeira K (2017) Geophysical potential for wind energy over the open oceans. Proc Natl Acad Sci 114(43):11338–11343
10. Kaldellis JK, Kapsali M (2013) Shifting towards offshore wind energy-recent activity and future development. Energy Policy 53:136–148
11. Elgharib AO, Alhasheem M, Swief RA, Naamane A (2021) Wind turbine performance assessment boost converter based applying pi controller integrating genetic algorithm. In: 2021 International conference on microelectronics (ICM), IEEE, pp 236–241

Electric Vehicle

NARX Black-Box Modeling of a Lithium-Ion Battery Cell Based on Automotive Drive Cycles Measurements

Jaouad Khalfi [ORCID]**, Najib Boumaaz, Abdallah Soulmani, Sara Laafar, Asmaa Maali, and El Mehdi Laadissi**

Abstract In this work, we developed an accurate voltage estimation solution for Li-ion batteries using the NARX model. The proposed estimator is based on experimental data that represents actual battery usage in electric vehicles. Three versions of the NARX model with different configurations are presented. The prediction performance of the proposed model is evaluated in term of the mean squared error criteria (MSE) between the measured battery cell output and the NARX model output for different driving cycles of an electric vehicle. One NARX model was selected and was compared with other model types including a linear model, a nonlinear model and an equivalent electrical circuit model using the same datasets. The NARX model developed in this work is more accurate for the estimation of the battery voltage compared to the other models. It can be considered as an interesting candidate for integration in a battery management system for an electric vehicle.

Keywords Electric vehicle · NARX model · Lithium-Ion battery cell · Measurements

J. Khalfi (✉) · N. Boumaaz · A. Soulmani · S. Laafar · A. Maali
Laboratory of Electrical Systems, Energy Efficiency and Telecommunications (LSEEET),
Department of Physics, Faculty of Sciences and Technology, Cadi Ayyad University, Marrakech,
Morocco
e-mail: jaouad.khalfi@ced.uca.ma

N. Boumaaz
e-mail: Najib.BOUMAAZ@emines.um6p.ma

A. Soulmani
e-mail: a.soulmani@uca.ma

S. Laafar
e-mail: sara.laafar@ced.uca.ma

A. Maali
e-mail: asmaa.maali@ced.uca.ma

E. M. Laadissi
Laboratory of Engineering Sciences for Energy (LabSIPE), National School of Applied Sciences,
Chouaib Doukkali University, El Jadida, Morocco
e-mail: laadissi.e@ucd.ac.ma

1   Introduction

Energy production is heavily based on fossil resources, such as coal, oil, and gas. However, these resources are not renewable and there will come a day when they will run out, whether in a few tens or even hundreds of years depending on the resource. In addition, the excessive use of these fossil fuels pollutes the planet and causes climate change, the consequences of which we are already beginning to feel, particularly in terms of weather conditions [1].

In order to contain this global warming within the limits provided by the global climate agreement many governments wish to engage in an energy transition to end these fossil fuels [2].

In addition, global warming and forecasts of a shortage of fossil resources are warning the governments of developed countries, who are introducing policies to reduce greenhouse gas emissions. The electrification of the means of transport is the best solution to reduce these emissions [3, 4]. In this process, the use of electric vehicles (EVs) is encouraged and subsidized.

In the global economic context, the sale of electric vehicles has exploded. A forecast for the size of the global fleet in 2030 is around 220 million electric and hybrid vehicles. This boom is due to the considerable reduction in the cost of manufacturing batteries, since batteries are very important and expensive components in an electric vehicle [5, 6].

There are a wide variety of batteries with different properties. Inexpensive and capable of delivering high currents, lead-acid batteries are widely used in industry and in the automotive industry for starting heat engines [7–9]. Nickel-cadmium batteries are very reliable, robust and require little maintenance. They are appreciated in areas where security is critical, such as aeronautics, offshore platforms or the military [10].

In this work, we focus on another technology. Developed thirty years ago, lithium-ion batteries have literally invaded our daily lives. More powerful, more efficient, more durable, lighter, and less bulky than their predecessors. We rely heavily on this technology to sweeten our future [11–13].

Lithium-ion batteries are systematically equipped with a management system to overcome the problem of instability. This is the "Battery Management System". This is a small electronic card whose main function is to ensure that the battery remains within its safe operating range. This system protects the battery against overcurrent, overvoltage, undervoltage, overheating, overload, over-discharge, and imbalance [14–16].

The development of mathematical models is essential to estimate the best battery parameters such as voltage, state of charge etc. which are used by the battery management system of electric vehicles [9]. Several types of models in the literature have been developed to estimate these parameters. These models can be divided in two large families. The first family is formed by the models which describe the electrical behavior of the battery. These are the equivalent electrical circuit models [17–19]. A second family includes black box models type. These models use various theoretical approaches, such as linear models like the Box-Jenkins polynomial model [20], or

a nonlinear models based on structured blocks like the Hammerstein-Wiener model [21–23] or models based on neuron networks like the NARX model (Nonlinear Auto-Regressive with exogenous input) [24] developed in this work. These models succeeded in modeling the li-ion battery of electric vehicles and they gave interesting results in terms of precision.

2 NARX Model Structure

The problem of identifying linear dynamical systems based on black box models is currently well known and a large number of techniques have been developed to estimate the parameters of linear, discrete, or continuous models. The best alternative to these approaches is to model the system using a nonlinear model and estimate these parameters. In this type of approach, several models have been proposed, starting from the classic work of wiener until recent development in the field of neural networks.

The identification of nonlinear systems by neural networks has been the subject of much research for a long time because of the capacity for learning, approximation, and generalization that these networks possess [24]. Indeed, this new approach provides an efficient solution through which large classes of nonlinear systems can be modeled without a precise mathematical description. In this work, we present the resolution of identification problems by neural techniques using the NARX model, and show the ability of this approach to identify batteries with great success, since it has the advantage of being able to approximate the nonlinear behavior of the battery cell during the driving cycles of electric vehicles with interesting accuracy.

In Fig. 1, we have represented a block diagram of the NARX model. It is characterized by the nonlinear relations between the predicted process output, the past outputs and the past inputs. The NARX model output is defined by Eq. (1):

$$
\begin{aligned}
y(t) = f\big(y(t-1), y(t-2), \ldots, y\big(t-n_y\big), \\
u(t-1), u(t-2), \ldots, u(t-n_u)\big)
\end{aligned}
\tag{1}
$$

where

- $y(t-1), y(t-2), \ldots, y\big(t-n_y\big)$ are lagged terms of model output;
- $u(t-1), u(t-2), \ldots, u(t-n_u)$ are terms of lagged input;
- n_y and n_u are the orders of output and input for the dynamical model providing that f is a nonlinear function.
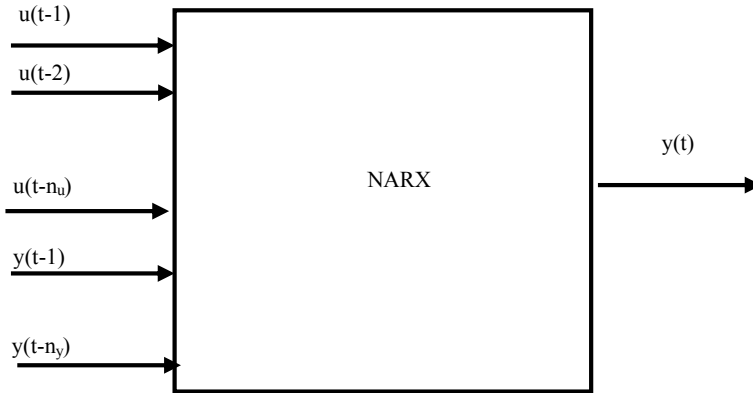
Fig. 1 NARX structural model

3 Experimental Data

In order to develop this model, it is a question of the offline analysis of the data accumulated during the various missions carried out by electric vehicles to broaden the field of application of the proposed model. Given the large amount of data available, as well as the diversity of uses, the data will be processed beforehand. The main objective in this work is to obtain indicators of the battery condition using this data. We specify that the proposed solutions are in a context of offline use.

The University of Wisconsin-Madison maintains a database for new Panasonic 18650 battery cell with a lithium nickel cobalt aluminum oxide (LiNiCoAlO$_2$ or NCA) chemistry [25, 26]. These measured data represent the dynamics of the battery during the driving cycles of electric vehicles on urban, rural and highways roads and it consists of a random mix of the following driving cycles [27]:

- **LA-92**: the unified cycle driving schedule (UCDS), also known as The California Unified Cycle. LA-92 it was developed in 1992 by the California air resources board (CARB) is a chassis-dynamometer driving schedule for light-duty vehicles.
- **US06**: It addresses the need for aggressive, high-speed driving behavior and/or high-acceleration, rapid speed fluctuations, and driving behavior following startup.
- **UDDS**: (UDDS: The urban dynamo driving schedule) is the standard driving cycle for the certification of light-duty trucks and passenger vehicles.
- **HWFET**: it simulates interstate rural and highway conditions of driving, known as the highway fuel economy test.
- **Neural network**: it consists of a portions combination of the two driving cycles LA-92 and US06, and it was designed to may be useful for training neural networks, since it provide additional dynamics.

To identify our NARX model for the battery cell, the measurements used in the driving cycle process are shown in Fig. 2. These measurements consist on the
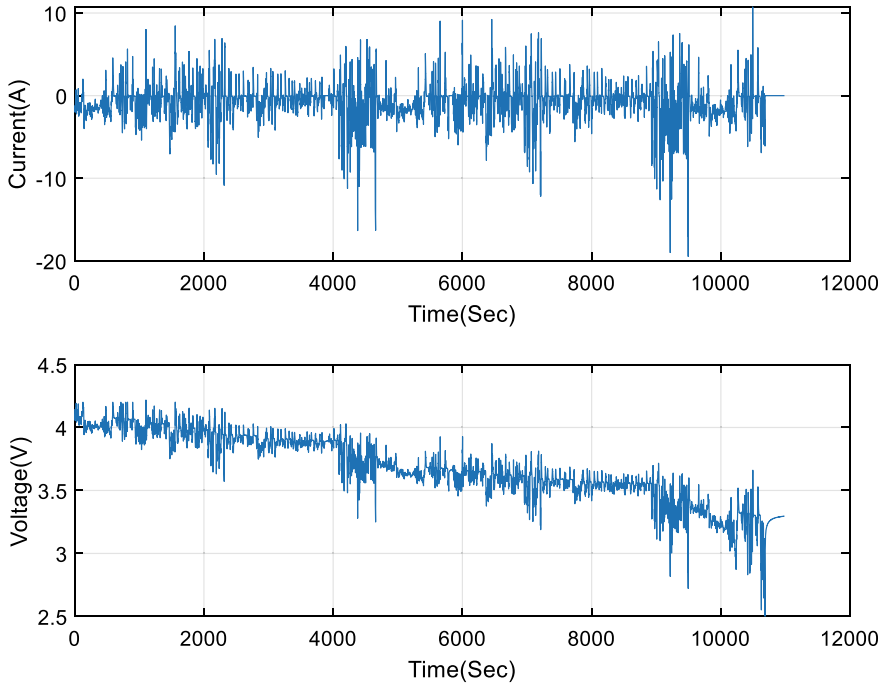
Fig. 2 Experimental data for NARX model estimation [28]

measured voltage at the battery cell terminal state in (V) and the measured current applied to the battery cell in (A). The time of test, measured in (s) which starts from 0 s at the beginning of each data set.

4 Results and Discussion

4.1 NARX Model Estimation

In order to compare the various identification configuration of the NARX model, three simulation examples are setup. The first version of these examples (NARX1) corresponds to the identification of the output of the battery cell whose number of hidden neurons have been set at 10 and number of delays at 2 for NARX neural network. In the second example (NARX2), the number of hidden neurons is set to 15. And for the third scenario (NARX3), we have considered 20 hidden neurons. For each configuration, the identification of the NARX model is carried out with the same experimental data that represents the driving cycles of electric vehicles using

Table 1 Performance of the NARX model for different configurations using MSE criteria

Model version	Mean squared error		
	Training	Validation	Test
NARX1	$1.61371e-4$	$5.60150e-4$	$1.28311e-4$
NARX2	$7.36306e-4$	$1.44954e-3$	$1.07881e-3$
NARX3	$2.35234e-3$	$2.02340e-3$	$1.94927e-3$

the levenberg-marquardt algorithm. These experimental data are divided in a random way in three groups:

- A group for training using 70% of the dataset.
- A group for validation using 15% of the dataset.
- And a group for testing that uses 15% of the dataset.

To determine the accuracy of the identified model, three indicators are observed as shown in Table 1.

- The mean squared error in the learning phase
- The mean squared error in the validation phase
- The mean squared error in the test phase.

Table 1 presents the three proposed models with different configurations (NARX1, NARX2, and NARX3) and the performances of each model using the MSE criteria. We note that the model NARX1 presents the best accuracy in term of MSE. In this model, the selected number of hidden neurons is 10 and the number of delays is 2. The structure of this model is presented in Figure 3, where $x(t)$ represents the input current vector at time instant "t" and $y(t)$ is the corresponding output voltage at the terminal of battery cell, "W" are the weight vectors in the output layer and the hidden layer and b the biases vectors of the NARX1 model. The training algorithm of Levenberg–Marquardt (LM) is used for training the NARX1 model. LM method has the combined advantage of gradient descent method, which is assured to converge and Newton's method, which converges rapidly. The LM method is an efficient method to train neural networks with high accuracy. Figure 4 presents the simulation results obtained in training phase of NARX1 model, in term, of the error between measured and simulated responses. We can conclude that these two responses are similar, which confirms the precision of the NARX1 model.

4.2 Comparison Between the NARX Model (NARX1) and Other Models

The performances of the NARX model (NARX1) using the mean squared error (MSE) between simulated and measured data are compared to the following three other models:
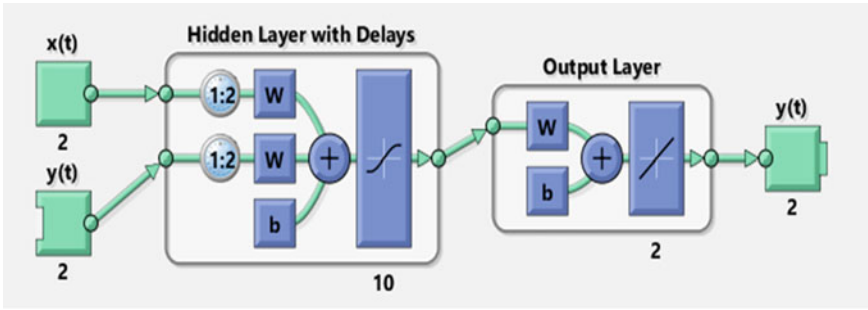
Fig. 3 NARX1 model structure
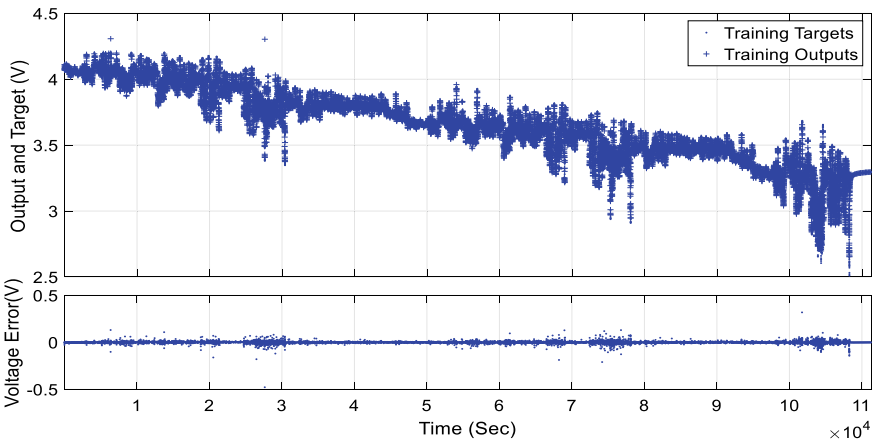
Fig. 4 NARX1 model results for training phase

- **Second order model**: This model was developed using equivalent electrical circuit model and consists of an internal resistance R_0 in series with an ideal voltage source U_{oc} and two dipoles RC parallel circuit pairs to capture charge transfer resistance effects and double layer capacitance [17].
- **Box–Jenkins linear model**: This is a linear polynomial model of order 7 [20].
- **Nlhw2**: This model was developed based on Nonlinear Modeling using a Hammerstein–Wiener Model [21].

We notice that all these models are based on the same measurement dataset. Table 2 presents the performance of different models. We can conclude that the NARX model (NARX1) have the best accuracy compared to the three other models.

Table 2 The models performance using the MSE criteria

Model type	MSE
Second order model	0.0013
Box–Jenkins model	0.001177
NARX1	0.000161371
Nlhw2	0.0002777

5 Conclusion

In this work, we restricted ourselves to the identification of the li-ion battery using neural networks based on the use of the NARX model. We offer this model for the li-ion battery as part of its use in electric vehicles. We presented three versions of the NARX model with different configurations. After a comparative study between these models, only one model referenced to as NARX1 was selected using quantitative statistical analyzes based on MSE criteria in the phase of learning, validation, and test. Finally the selected model was compared with other model types including a linear model, a nonlinear model and an equivalent electrical circuit model using the same datasets. The NARX model developed in this work achieved interesting results. It can be considered as an excellent model to use in a li-ion battery management system which requires high accuracy.

References

1. Onat NC, Kucukvar M, Tatari O (2015) Conventional, hybrid, plug-in hybrid or electric vehicles? State-based comparative carbon and energy footprint analysis in the United States. Appl Energy 150:36–49
2. Seck GS, Hache E, Barnet C (2022) Potential bottleneck in the energy transition: the case of cobalt in an accelerating electro-mobility world. Resour Policy 75:102516
3. Verma S, Dwivedi G, Verma P (2021) Life cycle assessment of electric vehicles in comparison to combustion engine vehicles: a review. Mater Today Proc
4. Bleviss DL (2021) Transportation is critical to reducing greenhouse gas emissions in the United States. Wiley Interdisc Rev Energy Environ 10(2):e390
5. Yang C (2022) Running battery electric vehicles with extended range: coupling cost and energy analysis. Appl Energy 306:118116
6. Deng J, Bae C, Denlinger A et al (2020) Electric vehicles batteries: requirements and challenges. Joule 4(3):511–515
7. Sadabadi KK, Ramesh P, Tulpule P et al (2021) Model-based state of health estimation of a lead-acid battery using step-response and emulated in-situ vehicle data. J Energy Storage 36:102353
8. El Mehdi Laadissi JK, Belhora F, Ennawaoui C, El Ballouti A (2020) Aging study of a lead-acid storage bank in a multi-source hybrid system. Indones J Electr Eng Comput Sci 20:1109–1117
9. El Mehdi Laadissi AE, Zazi M, El Ballouti A (2018) Comparative study of lead acid battery modelling. ARPN J Eng Appl Sci 13:4448–4452
10. Yazvinskaya NN, Galushkin DN, Isaev IR (2021) Probability of thermal runaway in high-capacity nickel-cadmium batteries with pocket electrodes. In: IOP conference series: materials science and engineering. IOP Publishing, p 012091

11. Hu X, Zou C, Zhang C, Li Y (2017) Technological developments in batteries: a survey of principal roles, types, and management needs. IEEE Power Energy Mag 15:20–31
12. Emadi A (2014) Advanced electric drive vehicles. CRC Press: Boca Raton, FL, USA
13. Chemali E, Preindl M, Malysz P, Emadi A (2016) Electrochemical and electrostatic energy storage and management systems for electric drive vehicles: state-of-the-art review and future trends. IEEE J Emerg Sel Top Power Electron 4:1117–1134
14. Dorn R, Schwartz R, Steurich B (2018) Battery management system. In: Lithium-Ion batteries: basics and applications. Springer, Berlin/Heidelberg, Germany, pp 165–175
15. Slanina Z, Dedek J, Golembiovsky M (2017) Low cost battery management system. J Telecommun Electron Comput Eng (JTEC) 9:87–90
16. Gabbar HA, Othman AM, Abdussami MR (2021) Review of battery management systems (BMS) development and industrial standards. Technologies 9:28
17. Khalfi J, Boumaaz N, Soulmani A, Laadissi EM (2021) An electric circuit model for a lithium-ion battery cell based on automotive drive cycles measurements. Int J Electr Comput Eng 11:2798–2810
18. Laadissi EM, Anas EF, Zazi M, Jaouad K (2019) Parameter identification of a lithium-ion battery model using Levenberg-Marquardt algorithm. J Eng Appl Sci 14:1267–1273
19. Bouzaid S, El Mehdi L, Chouaib E, El Mehdi L Mossaddek M, El Ballouti A (2022) Lithium-ion battery modeling using dynamic models. Mater Today Proc
20. Khalfi J, Boumaaz N, Soulmani A, et al (2021) Box–Jenkins black-box modeling of a lithium-ion battery cell based on automotive drive cycle data. World Electr Veh J 12(3):102
21. Khalfi J, Boumaaz N, Soulmani A et al (2021) Nonlinear modeling of lithium-ion battery cells for electric vehicles using a hammerstein–wiener model. J Electr Eng Technol 16(2):659–669
22. Zazi M (2016) Nonlinear black box modeling of a lead acid battery using Hammerstein-Wiener model. J Theor Appl Inf Technol 89:476
23. Mossaddek M, El Mehdi L, El Mehdi L, Ennawaoui C, Khalfi J, Bouzaid S, Hajjaji A (2022) Nonlinear modeling of lithium-ion battery. Mater Today Proc
24. Laadissi E, El Filali A, Zazi M (2018) A Nonlinear TSNN based model of a lead acid battery. Bull Electr Eng Inf 7(2):169–175
25. Panasonic (2016) Panasonic NCR18650PF Lithium-Ion Battery Datasheet. Available online: https://actec.dk/media/documents/70FC46554038.pdf. Accessed on 1 June 2021
26. Panasonic (2013) Introduction of NCR18650PF, Panasonic. Available online: https://voltaplex.com/media/whitepapers/specification-sheet/Panasonic_PF_Specification_Sheet.pdf. Accessed on 1 June 2021
27. Giakoumis EG (2017) Driving and engine cycles. Springer International Publishing, Cham, Switzerland
28. Phillip K (2018) Panasonic 18650PF Li-ion Battery Data. Mendeley Data. Available online: http://dx.doi.org/10.17632/wykht8y7tg.1. Accessed on 1 June 2021

Design and Demonstrate an Attack Strategy to Control a Vehicle's Computer by Targeting Its Electrical Network

Mohammed Karrouchi, Abdelhafid Messaoudi, Kamal Kassmi, Ismail Nasri, Ilias Atmane, and Jalal Blaacha

Abstract The functionality of most modern automobiles is controlled by electronic control units (ECUs) that interact with one another via the CAN communication protocol (controller area network), as defined by the ISO 11898 standard, which is one of the fundamental features of most modern cars. This physical bus provides excellent data transfer quality by providing wide propagation reaching all areas of a vehicle in a short amount of time. However, because of the absence of secrecy and ease of access (physical or remote), this protocol does not place a high value on security, which causes the CAN bus control system to be vulnerable. Due to this weakness, it is possible to control the car from the outside and place the vehicle's passengers in jeopardy. The objective of this paper is to describe the existing vulnerability of the security system against attacks, and also a demonstration of a hacking technique used to manage the dashboard of a 2014 DACIA Dokker automobile using the automotive network.

Keywords CAN bus vulnerability · Ways of hacking · CAN Bus · OBD2 protocol

1 Introduction

Many electrical units are integrated into modern cars and process the measured data using different sensors installed to perform relevant tasks, such as route planning, collision prevention, automatic parking, and pedestrian detection, as well as current technologies such as Wi-Fi, Bluetooth for powerful operation, efficient data flow, and

M. Karrouchi (✉) · K. Kassmi · I. Nasri · J. Blaacha
Electrical Engineering and Maintenance Laboratory, High School of Technology, Mohammed First University, BP. 473, Oujda, Morocco
e-mail: m.karrouchi@ump.ac.ma

A. Messaoudi
Energy, Embedded Systems and Information Processing Laboratory, National School of Applied Sciences, Mohammed First University, Oujda, Morocco

I. Atmane
Department of Physics, Faculty of Science, Mohammed First University, Oujda, Morocco

to make life easier for the driver. The electronic control units (ECU) communicate with each other via the controller area network (CAN), and this serial protocol uses a pair of twisted wires as a common means of transmission between the nodes. The CAN bus is a local area network and multi-master broadcast serial bus standard for connecting electronic control unit peripherals, or microcontroller devices. Each node can send and receive messages [1]. When a unit fails, it does not affect others. Figure 1 shows us the topology of the CAN module as well as describes the structure of the OBD2 connector. This standard port is a door to access the CAN bus, its a protocol that is implemented in most of the vehicles which are manufactured lately [2]. The OBD2 standard covers only the exchange of diagnostic data [3]. Each message is distributed on the CAN bus containing an identifier (ID) used to identify the content of the frame and define it according to the needs which among the nodes must exploit [4]. This mutation makes cars like computers with wheels, simplifies many things but also opens a possibility of hacking the local networks of these vehicles with the connection [5, 6]. Nowadays, many university researchers have demonstrated the possibility of attacking the vehicle's control system and accessing its physical CAN bus [6, 7]. By design, this bus suffers from a weakness of confidentiality and security, including message authentication and data encryption. This allows hackers to gain access to the network and trigger malicious activity, and this is what Miller and Valasek proved in 2015 by an attack on the JEEP Cherokee [8]. They demonstrated a hacking technique on automotive systems, and they were able to remote control the 2014 Jeep Cherokee and represented how to exploit a vulnerability in the main units of the vehicle to control certain actions including steering and braking. Costantino and Matteucci [9] showed a real example of an attack called CANDY CREAM, and they discovered and targeted the existing vulnerabilities on the Android IVI operating system, connecting to the physical CAN bus; therefore, they were able to take control over the vehicles, on which the android system has been installed. Abbott-McCune and Shay [10] presented current research on local vehicle networks, as well as described techniques for collecting data circulating on the CAN bus, based on a simulation system for further research. Koscher et al. [11] indicated the fragility of the modern vehicle security system and experimentally determined its security frameworks. They have proven that an attacker is capable of targeting virtually any ECU, and bypassing a wide range of critical systems to safety and to control various automotive functions, including stopping the engine and disabling the brakes.

2 Methodelogy

2.1 System Block Diagram

The handling structure is illustrated in Fig. 1, which has consisted of four components: computer, OBD2 port specific to the vehicle, CAN bus control circuit, which

OBD-2 socket

CAN BUS

ECU1

OBD-2 port

ECU2

ECU3

ECU5

ECU4

CAN Bus Shield

Atmega 328

The attacker
with his computer
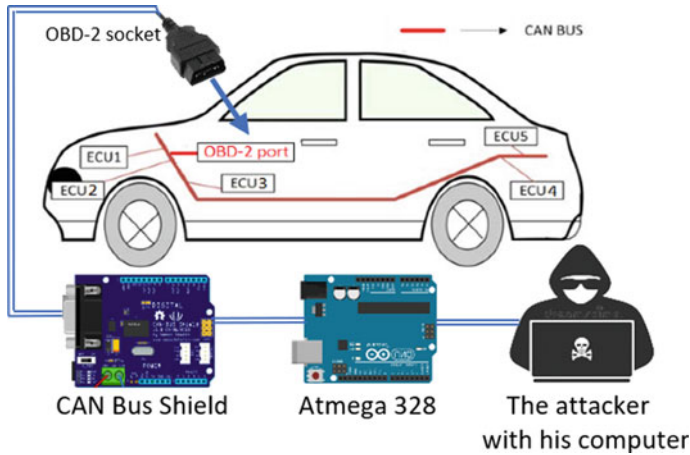
Fig. 1 System block diagram

consists of an ATmega 328 microcontroller to act and supervise the communications between the peripherals, CAN Bus shield consisting of an MCP2515 CAN controller to manage the send and receive rules, and CAN line interface MCP2551 to adapt the voltage levels. The transmission rate on the CAN bus is 500kbit/s, and the frequency value of the quartz connected to the microcontroller used is 16 MHz. The main CAN circuit extracts the data circulating in the vehicle through the OBD2 port, and then broadcasts these messages to the computer using UART serial communication (Universal Asynchronous Receiver-Transmitter). The utility of the computer is to program the ATmega 328 microcontroller board, to read the sequenced data and display it in the serial monitor, also configuring the microcontroller to broadcast unauthorized frames on the physical bus, and create an intrusion on the vehicle. It can be noted that our equipment behaves as an electronic external control unit (ECU), which sends and receives information on the CAN physical line at the same time as the other units.

2.2 The Instructions Followed

Our program implemented in the microcontroller is restructured in three steps. The first contains a program whose role is to collect and read in real-time all the data circulating on the CAN bus and to display them according to their identifiers [12]. Figure 2 summarizes the organization chart of this first internship.

The messages received are filtered according to their identifier thanks to the filter and mask registers. So, to be able to receive the content of all the frames circulating in the physical bus, it is necessary to make an appropriate configuration on the masks and filters and to reset them to zero.
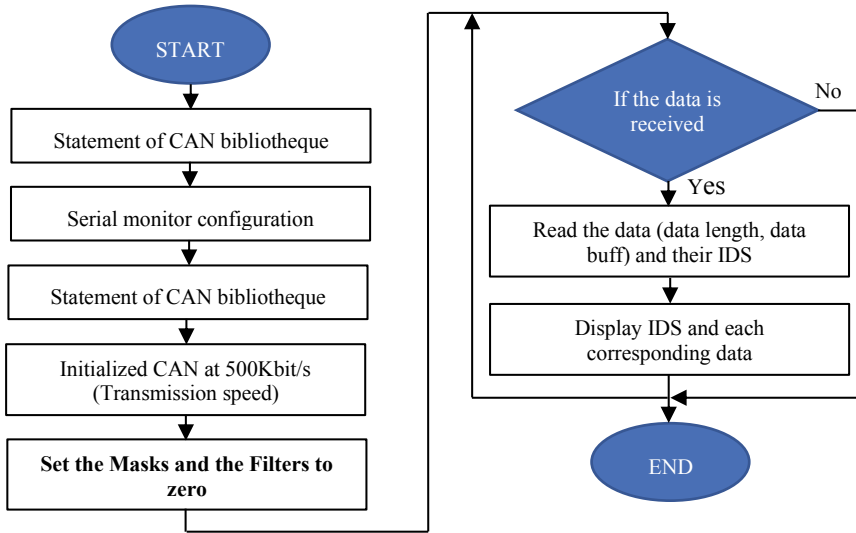
Fig. 2 Software structure for first program

At this stage, and after this recovery, we need a second task whose goal is to accept only a single frame of an identifier that we specified. This task was described in the second program which is identical to the first, except that the masks are at 1 and the filters take the value of the chosen identifier. This technique allows you to extract a single frame and hide the others. In each frame, it is necessary to repeat certain actions (press the accelerator pedal, for example) and observe the changes of bytes in the data corresponding to each identifier. Consequently, we can know the identifier and their responsible frame at each action in the vehicle. Figure 3 shows an example of changes in the content of the CAN message and the updated change at the byte level of an identifier (ID = 0×186) after certain actions. The strategy followed in the second program is illustrated in Fig. 4.

After identifying the frames and their identifiers responsible for the various actions, we need a third task, which is to intrude into the vehicle in an external way. Figure 5 shows the flowchart of the third program used to inject frames of an identifier specified in the CAN bus, with modification of the content of the data (byte change) to see the actions taken by the calculator (ECU) targeted in the attack.

3 Results

In this way, we were able to identify the frame containing the revolutions per minute (RPM) display data in the dashboard, as well as the frame which reacts with the dashboard lights (represents the signals from the car). After several attempts, it showed us that the first byte of these two frames which interacts with the variation of

COM6 — Before reaction

```
21    5C    30    13    6    0    21
------------------------------
Get data from ID: 0x186
20    7C    30    13    7    0    21
------------------------------
Get data from ID: 0x186
20    7C    30    43    8    0    21
------------------------------
Get data from ID: 0x186
20    7C    30    43    A    0    21
------------------------------
Get data from ID: 0x186
1F    AC    30    43    B    0    21
------------------------------
Get data from ID: 0x186
1F    AC    30    83    D    0    21
------------------------------
Get data from ID: 0x186
1F    AC    30    83    F    0    21
```

COM15 — After reaction

```
21 7B    84    35    53    54    21    A1
------------------------------
Get data from ID: 0x186
21 7C    2C    35    43    54    21    A1
------------------------------
Get data from ID: 0x186
21 7C    A0    35    43    54    21    A1
------------------------------
Get data from ID: 0x186
21 7C    DC    35    43    54    21    A1
------------------------------
Get data from ID: 0x186
21 7D    88    35    43    54    21    A1
------------------------------
Get data from ID: 0x186
21 7D    FC    35    43    53    21    A1
------------------------------
Get data from ID: 0x186
21 7E    34    35    33    53    21    A1
------------------------------
```

Fig. 3 Change of data of an identifier after an action

The Masks at 1 → Set the Filters to the chosen identifier value → If the content of the frame change — Yes → It's the responsible frame; No (loop back)
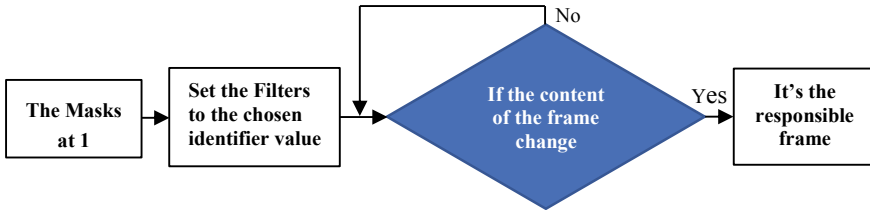
Fig. 4 Follow-up procedure in the second program

START → Statement of CAN bibliotheque → Serial monitor configuration → Statement of CAN bibliotheque → Initialized CAN at 500Kbit/s (Transmission speed) → **Define the chosen identifier** → i=0 → i < data length — No → END; Yes → Send the data vie the CAN Bus → END
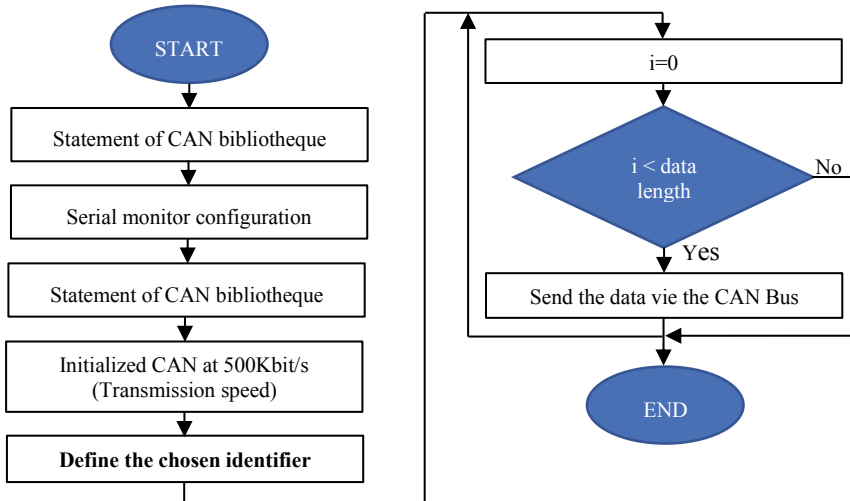
Fig. 5 Injection program of data into the CAN bus

the RPM display and the signals (left turn signal, right turn signal, distress, headlight, code, and pilot) in the dashboard. Figure 6 shows the experimental results. Table 1 shows some first-byte values for the two frames and their correspondence. The tests are carried out on a DACIA Dokker vehicle available at the Oujda Higher Technical School. Figure 7. defines the variance of the engine speed (RPM) and their values which correspond to the first byte.

Fig. 6 **a** Takes the control of the revolutions per minute. **b** Takes the control of the indicator lights

Table 1 First-byte values and its correspondence

Identifier value	Value of the first byte of the frame (Hex)	Correspondence
ID: 0 × 186 RPM tachometer	0 × 00	0000 Tr/min
	0 × 60	3000 Tr/min
	0 × C8	5500 Tr/min
	0 × FF	7000 Tr/min
	0 × 20	A left turn signal
ID: 0 × 5DE indicator lights	0 × 40	A right turn signal
	0 × 60	Distress
	0 × 01	Deadlights
	0 × 04	Sidelights
	0 × 06	Base light
	0 × 61	Deadlights + Distress
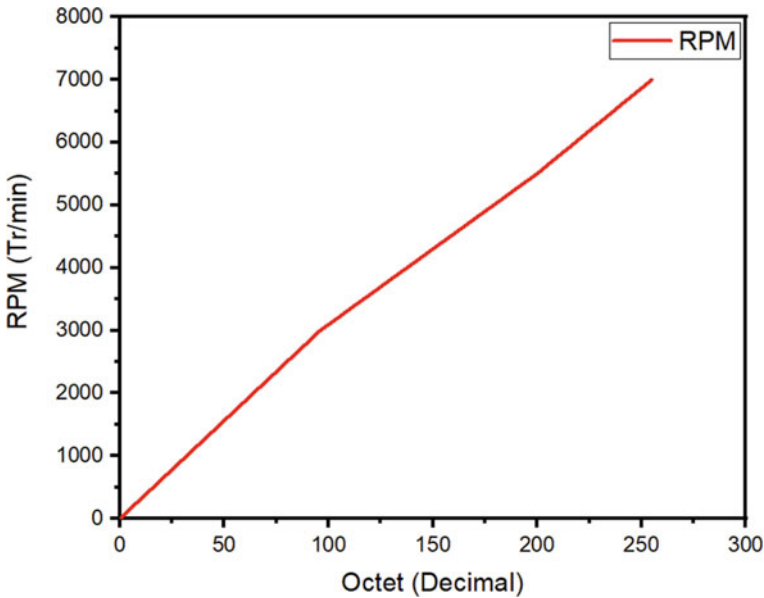
Fig. 7 Variation of the engine speed according to the values of the first targeted byte

4 Conclusion

Modern vehicle functionalities have become increasingly connected to the world around them to help the driver control his vehicle lightly and without fatigue. In this paper, we mentioned the vulnerabilities that we used to carry out an attack via the OBD2 port. The design code was divided into various sequences of processes, listening on the bus, identifying each ID corresponds to each action and in the end building the intrusion. Our study was carried out on a DACIA Dokker, we were able to modify externally the content of the dashboard such as the engine RPM display, light signals. We plan to expand the attacks to other computers (engine control unit, for example) and work to find solutions to fight against these attacks.

References

1. Pillai MA, Veerasingam S, Sai DY (2010) Implementation of sensor network for indoor air quality monitoring using CAN interface. In: IEEE international conference on advances in computer engineering. https://doi.org/10.1109/ACE.2010.85
2. Amarasinghe M et al (2015) Cloud-based driver monitoring and vehicle diagnostic with OBD2 telematics. In: 2015 Fifteenth international conference on advances in ICT for emerging regions (ICTer). IEEE. https://doi.org/10.1109/ICTER.2015.7377695
3. Karrouchi M, Nasri I, Snoussi H, Messaoudi A, Kassmi K (2021) Implementation of the vehicle speed and location monitoring system to minimize the risk of accident. In: International

conference on digital technologies and applications. Springer, Cham, pp 1539–1548. https://doi.org/10.1007/978-3-030-73882-2_140

4. Karrouchi M, Nasri I, Snoussi H, Messaoudi A, Kassmi K (2021) Black box system for car/driver monitoring to decrease the reasons of road crashes. In: The 4th international symposium on advanced electrical and communication technologies. https://doi.org/10.1109/ISAECT53699.2021.9668545

5. Abbott-McCune S, Shay LA (2016) Intrusion prevention system of automotive network CAN bus. In: 2016 IEEE international Carnahan conference on security technology (ICCST). IEEE, pp 1–8. https://doi.org/10.1109/CCST.2016.7815711

6. Ashraf J, Bakhshi AD, Moustafa N, Khurshid H, Javed A, Beheshti A (2020) Novel deep learning-enabled lstm autoencoder architecture for discovering anomalous events from intelligent transportation systems. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2020.3017882

7. Avatefipour O, Al-Sumaiti AS, El-Sherbeeny AM, Awwad EM, Elmeligy MA, Mohamed MA, Malik H (2019) An intelligent secured framework for cyberattack detection in electric vehicles' CAN bus using machine learning. IEEE Access 7:127580–127592. https://doi.org/10.1109/ACCESS.2019.2937576

8. Miller C (2019) Lessons learned from hacking a car. IEEE Des Test 36(6):7–9. https://doi.org/10.1109/MDAT.2018.2863106

9. Costantino G, Matteucci I (2019) CANDY CREAM hacking infotainment android systems to command instrument cluster via can data frame. In: 2019 IEEE international conference on computational science and engineering (CSE). https://doi.org/10.1109/CSE/EUC.2019.00094

10. Abbott-McCune S, Shay LA (2016) Techniques in hacking and simulating a modem automotive controller area network. In: 2016 IEEE international Carnahan conference on security technology (ICCST). IEEE. https://doi.org/10.1109/CCST.2016.7815712

11. Koscher K et al (2020) Experimental security analysis of a modern automobile. In: The ethics of information technologies. Routledge, pp 119–134. https://doi.org/10.1109/SP.2010.34

12. Zedan MJ (2020) Operating system realization for real-time visualization of CAN-Bus data streams using Xilinx ZyncSoC. J Electr Electron Inf 4(2). p-ISSN: 2549-8304 e-ISSN: 2622-0393

Real-Time Power Management Strategy of Battery/Supercapacitor Hybrid Energy Storage System for Electric Vehicle

Mohammed Amine Mossadak, Ahmed Chebak, Nada Ouahabi, and Abdelhafid AIT Elmahjoub

Abstract Due to the most important impact of the transportation sector on climate changes, the Li-ion batteries deployed in electric vehicles (EV) are considered as a suitable choice for reducing the toxic gases and save our climate. But, given the driving range limitations and the lifetime degradation of Li-ion batteries, the EVs based only on batteries still unable to be competitive against ICE vehicles. In this way, the integration of hybrid energy storage systems (HESSs) represents a trending research topic in EVs domain with the expectation to enhance the battery lifetime. However, the battery/supercapacitor topology requires a real-time energy management strategy that allows to manage the energy flux in the powertrain efficiently while optimizing the lifetime of the battery. This paper proposes a real-time power management control system based on two levels in which the high level is focused on power sharing between the HESS on-boarded in the EV using If-else rules and frequency-decoupling methods while taking into consideration the electrical limitations of each source. On the other hand, the low level is focused on ensuring the good regulation of the DC bus voltage and control of the battery/SC electric currents. The performance of the proposed strategy is validated over Extra Urban Driving Cycle (EUDC) using MATLAB/Simulink. The results obtained show an improvement in the lifetime of Li-ion batteries which validates the efficiency of the proposed framework.

Keywords Lithium-ion battery · Supercapacitor · Electric vehicle · Power management strategy · Frequency-decoupling

M. A. Mossadak (✉) · A. Chebak · N. Ouahabi · A. AIT Elmahjoub
Green Tech Institute, Mohammed VI Polytechnic University, Benguerir, Morocco
e-mail: mohammed.mossadak@um6p.ma

A. AIT Elmahjoub
ENSAM-Casa, Hassan II University, Casablanca, Morocco

1   Introduction

Nowadays, the negative and dangerous contribution of the transport sector on the environment is alarming and it is expressed by the rapid warming of our planet, the increase in the concentration of CO_2 and the depletion of the ozone layer, as well as by the increase in the demand for energy and the constant decrease of fossil fuels [1]. Therefore, finding a green solution has become an obligation. With the technological progress in electric motors, power electronics and especially batteries, the automotive market has started to see a growing penetration of electric vehicles based on lithium-ion batteries [2]. However, the battery electric vehicles (BEV) have many challenges to overcome, such as driving range, lifetime, and cost. To address these challenges, the integration of Hybrid Energy Storage Systems (HESSs) has attracted the attention of many researchers in recent years [3].

The HESS concept is based on combining two or three sources and deploying them in an electric vehicle to exploit the advantages of each source. Batteries, supercapacitors, and fuel cells are the main sources used in EVs. In the literature, there are several topologies and configurations for HESS, and they can be classified into four types: passive, semi-active, multi-mode, and fully active [4, 5]. In this study, we will focus on lithium-ion batteries and supercapacitor HESS while using fully active parallel topology. The choice of this topology is due to its qualities in terms of reliability, performance, and efficiency. Indeed, with this configuration we can reduce the size of the storage sources [6].

In fact, the battery and SC HESS require an energy management strategy to control and manage the power flow between the sources on-boarder not only that, but also in the entire powertrain system, in which the main objective is to satisfy the energy demanded by the load, improve the lifetime of the batteries and to enhance the driving range of the vehicle. In this way, the energy management strategies can be categorized into three types: rule-based strategy, optimization-based strategy, and data-driven strategy [7]. The flowchart below presents the types and sub-types of EMSs.

For many years, the If-else rules and frequency-decoupling methods have been handled separately. Therefore, in this work we tried to develop an EMS based on the advantages of If-else rules and frequency-decoupling techniques. So, our approach is based on two levels: power electronics control system and power sharing algorithm. The first level is dedicated to regulating the DC bus voltage and controlling the charging and discharging currents of each source while using a PI controller. The second level is responsible for sharing the electric power demand/regenerated between the on-boarded electrical sources. The algorithm of power sharing system merges two real-time strategies: rule-based strategy and frequency-decoupling method.

This paper is structured as follows: A general description and system modeling are provided in Sect. 2. Section 3 explains the proposed bi-level real-time power management control system. Then, the results of the simulation are presented and

discussed in Sect. 4. Finally, in Sect. 5 we conclude our study and discuss some perspectives.

2 System Description and Modeling

The objective of this section is to briefly describe the proposed HESS topology and to model the main elements of our study such as the Li-ion battery, the supercapacitor, the bidirectional DC/DC converters, and the electrical power demanded by the EV.

In this study, a fully active parallel topology has been used owing to its advantages in terms of efficiency, reliability, and performance. The on-boarded electric sources are linked to the DC bus via two bidirectional buck–boost DC/DC converters. Figure 1 shows the synoptic schematic of the system under study.

It should be noted that the traction system part (inverter + electric motor) is not included in our study, so they are modeled as a controllable electric current source.

2.1 Electric Vehicle Modeling

The electric vehicle model aims to convert a driving cycle speed profile to a power demand profile for HESS. The electric power required by a vehicle P_{ev} moving at speed V is calculated as the product between the total road forces and the vehicle speed, in which the total forces are the sum of the road slope force, inertial force, friction force, and aerodynamic drag force [7]. So, the P_{ev} is calculated as follows:

$$P_{ev} = \left[M_V.g.C_{rr} + \frac{1}{2}.\rho.C_d.A.V^2 + M_V.a_V \right].V \tag{1}$$
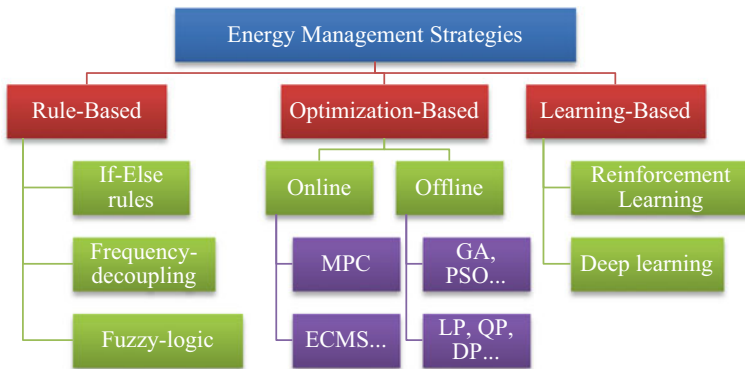
Fig. 1 Classification of EMSs techniques

where V is the vehicle speed; M_V is the mass of the vehicle; C_{rr} is the aerodynamic drag coefficient; g is the gravitational acceleration constant; and ρ_{air} is the air density.

2.2 Lithium-ion Battery Model

The equivalent circuit model of Li-ion battery in Fig. 2 is used due to its simplicity and its availability in MATLAB/Simulink library [8].

Where the charge model is as follows:

$$V_{bat} = E_0 - K \cdot \frac{Q_{max}}{0.1Q_{max} - q} \cdot i^* - K \cdot \frac{Q_{max}}{Q_{max} - q} \cdot q + A \cdot e^{-B \cdot q} \qquad (2)$$

And the discharge model is as below:

$$V_{bat} = E_0 - K \cdot \frac{Q_{max}}{Q_{max} - q} \cdot i^* - K \cdot \frac{Q_{max}}{Q_{max} - q} \cdot q + A \cdot e^{-B \cdot q} \qquad (3)$$

where E_0 is the constant voltage; K is the polarization constant in V/Ah; A is the exponential voltage; and B is the exponential capacity in Ah^{-1}.
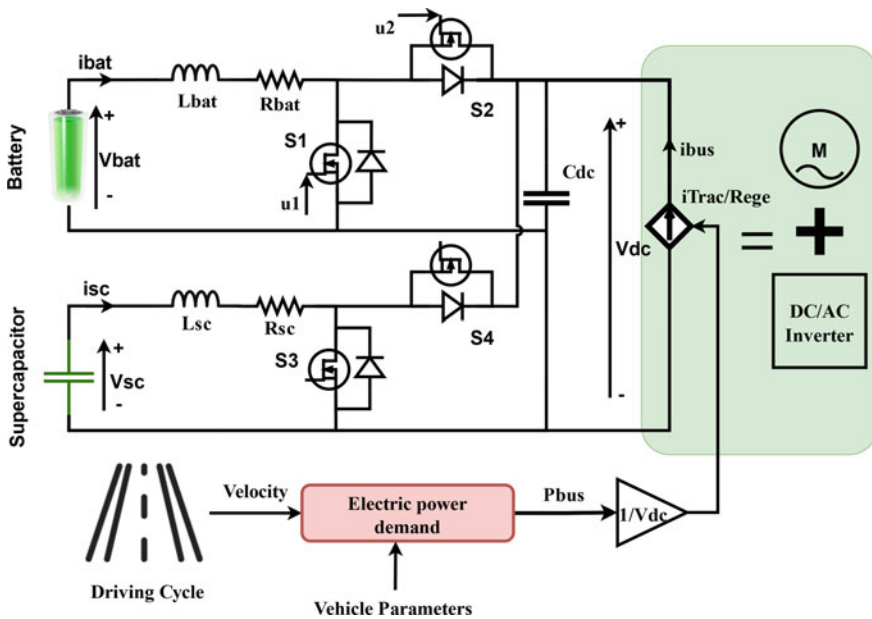
Fig. 2 Fully active parallel architecture of battery and supercapacitor HESS

2.3 Supercapacitor Model

The supercapacitor is modeled using a RC equivalent circuit composed of a capacitance (Csc), an equivalent parallel resistance R_p representing the self-discharging losses, and an equivalent series resistance R_s representing the charging and discharging resistance. This model is also available in MATLAB/Simulink libraries [9].

The supercapacitor output voltage U_t is given as follows:

$$U_t = R_s . i + \left[U_{C0} - \int_0^t \frac{i}{C} e^{\frac{t}{R_p.C}} dt \right] e^{-\left(\frac{t}{R_p.C}\right)} \tag{4}$$

where U_{C0} is the initial voltage of the ideal capacitor (C).

2.4 Bidirectional Buck–Boost DC/DC Converter Modeling

The bidirectional buck–boost DC/DC converter is composed of two MOSFET switches $S1$ and $S2$, an inductor L with an internal resistance R and a filter capacitor Cdc. The switches are controlled by applying a PWM signal to their gates which takes values between [0, 1] and determines the operation mode (boost or buck) of the converter [10]. Therefore, in discharging mode the converter works as a boost converter and in charging mode it operates as buck converter. A global model of the bidirectional buck–boost converter besides the battery and that combines the charging, and discharging modes are introduced as below:

$$\frac{di_{\text{bat}}}{dt} = \frac{V_{\text{bat}}}{L_{\text{bat}}} - \frac{R_{\text{bat}}}{L_{\text{bat}}} i_{\text{bat}} - (u_{12}) \frac{v_{\text{dc}}}{L_{\text{bat}}} \tag{5}$$

$$\frac{dv_{dc}}{dt} = (u_{12}) \frac{i_{\text{bat}}}{C_{dc}} - \frac{i_1}{C_{dc}} \tag{6}$$

where u_{12} is a control signal expressed as follows:

$$u_{12} = [K(1 - u_1) + (1 - K)u_2] \tag{7}$$

And K is a binary variable equal to 0 in charging mode and 1 in discharging mode. It should be noted that the two converters next to the HESS are, respectively, the same.

3 Power Management Control System

The philosophy behind the battery/SC hybridization is to operate the SCs to meet sudden and fluctuating load demands while the batteries meet the slow power demands in order to reduce the RMS of the battery current and therefore improve its lifetime [11]. In this section, we will highlight the proposed power management control strategy based on two levels: PI controller level and power sharing algorithm level.

3.1 PI Controller Design (Low Level)

Given the non-minimum phase of our bidirectional buck–boost converters, a cascade control architecture is adopted. In this way, the control of our converter will be conducted by two control loops. Furthermore, this type of control allows to improve the dynamic performance and the robustness of the controller [12]. The cascade control structure adopted is represented in Fig. 3.

The transfer functions $F1(s)$ and $F2(s)$ were obtained following a linearization of the nonlinear model in (6) and (7) of the converter around a point of operation in steady state [13]. The transfer functions are as follows:

$$F1(s) = \frac{I_{bat}}{\alpha_{bat}} = \frac{2 \cdot V_{bus}}{R \cdot (1 - \alpha_{bat})^2} \frac{1 + \frac{R \cdot C}{2}s}{1 + \frac{L_1}{R \cdot (1-\alpha_{bat})^2}s + \frac{L_1 \cdot C}{(1-\alpha_{bat})^2}s^2} \tag{8}$$
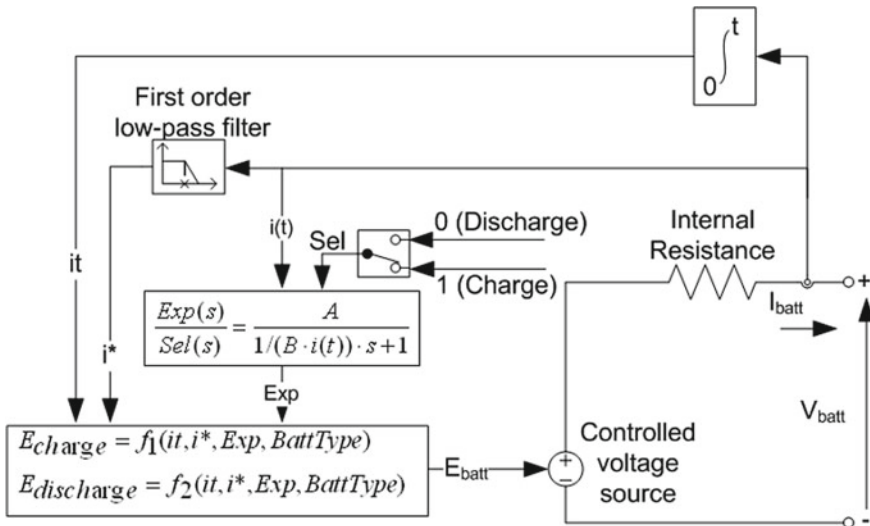
Fig. 3 Equivalent circuit model of Lithium-ion battery [10]
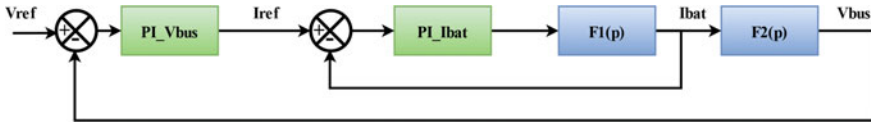
Fig. 4 Cascade control architecture of battery—DC/DC converter

$$F2(s) = \frac{V_{\text{bus}}}{I_{\text{bat}}} = \frac{R \cdot (1 - \alpha_{\text{bat}})}{2} \frac{\left(1 - \frac{L_1}{R \cdot (1 - \alpha_{\text{bat}})^2} s\right)}{1 + \frac{R \cdot C}{2} s} \tag{9}$$

To design our PI controller, we are based on the SISOTOOL graphical tool of MATLAB by introducing the transfer functions obtained and finding a compromise between stability, precision, and rapidity [13].

3.2 Power Sharing Algorithm (High Level)

The high level of our strategy proposed is dedicated to distributing the electric power demand/regenerated between the electric sources. The algorithm proposed is a combination of frequency-decoupling technic and rule-based methods. Indeed, the rule-based method allows to determine the operating modes of our sources according to their voltage levels and the power of the load. The flowchart in Fig. 4 describes the operation modes of our system.

After defining the operation mode using the rule-based strategy, a frequency-decoupling technic based on a high-pass filter (HPF) is used to extract the high-frequency electric powers for the supercapacitors and then the rest is for the batteries. The mathematical formula of a HPF is as follows:

$$\text{High Pass Filter}(s) = \frac{\tau s}{\tau s + 1} \tag{10}$$

We have chosen $\tau = 50$ s while based on the Ragone plot and the studies done in [11].

4 Simulation Results and Discussion

To evaluate the performance of the proposed real-time power management control strategy, the electric vehicle model based on Li-ion batteries and supercapacitor HESS is simulated in MATLAB/Simulink over EUDC driving cycle. Therefore, a comparative study between the proposed strategy and rule-based EMS has been

Table 1 Electrical characteristics of the HESS and DC bus voltage

Li-ion battery pack	(12.8 V; 30 Ah) 18 series and 4 parallel branches
DC bus voltage	400 V
Supercapacitor pack	(310 F; 2.7 V) 60 series cells 2 parallel branches

Table 2 Vehicle characteristics

Vehicle Mass Mv [Kg]	1500
Gravitational acceleration g [m/s^2]	9.81
crr [−]—road rolling resistance coefficient	0.011
ρ [kg/m^3]—air density at 20 °C	1.202
cd [−]—air drag coefficient	0.36
A [m^2]—vehicle frontal area	2.42

conducted to show the impact of our strategy on the battery life. The vehicle and the HESS characteristics are presented in Tables 1 and 2.

Using Eq. (1), the electric power demand is calculated through the EUDC illustrated in Fig. 5. We can observe in Fig. 6 that the electric power required by the vehicle is absolutely fulfilled by the battery and supercapacitor while using both methods. Not only that but also, we can confirm from Fig. 7 that the lifetime of the battery is improved using our strategy compared to rule-based method. Indeed, the RMS of the battery current managed by the proposed strategy has a smooth curve compared to the RMS of the battery managed just with (If-else)rules, which implies that in case of RB-EMS, the battery was exposed to more current fluctuations than the case of RBFD-EMS. We can also conclude that from Fig. 8 in which in case of RB-EMS the state of charge of the battery has been rapidly reduced compared to the battery managed by RBFD-EMS. For the DC bus voltage shown in Fig. 9, we observe that it is well-regulated at its reference (400 V) despite the small fluctuations due to load variations in Fig. 10.

5   Conclusion

In this paper, a bi-level real-time power management control strategy for electric vehicles powered by lithium-ion batteries and supercapacitors HESS has been proposed. A simulation under EUDC driving cycle while using MATLAB/Simulink environment has been established to evaluate the performance and effectiveness of the proposed strategy. A comparative study between a typical rule-based(If-else) energy management strategy and a combination of frequency-decoupling method and rule-based method. The simulation results obtained by implementing the proposed strategy illustrate the high efficiency of the power sharing between the on-boarded
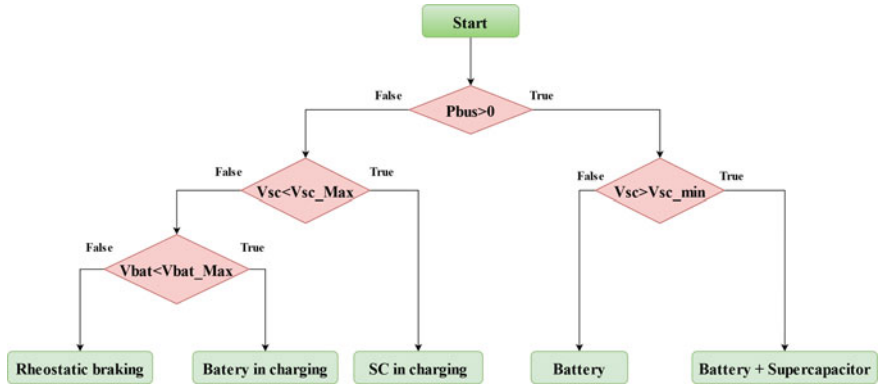
Fig. 5 Flowchart of the rule-based strategy
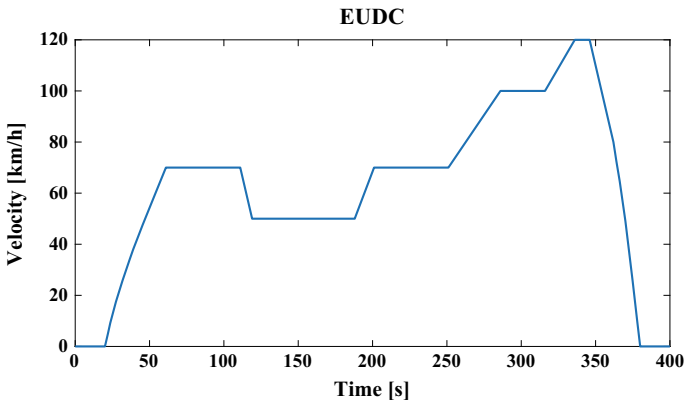
Fig. 6 Velocity of EUDC driving cycle
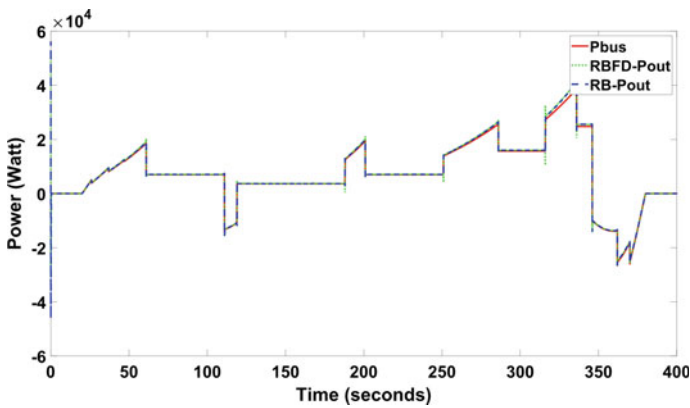
Fig. 7 Electric power demand by the vehicle and the power generated by the sources
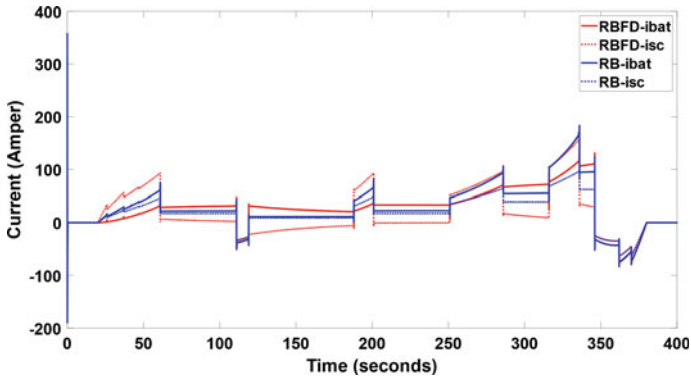
Fig. 8 Electric current of the on-boarded sources
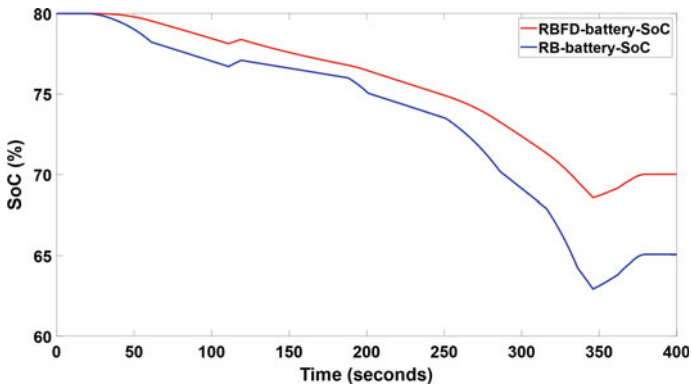
Fig. 9 State of charge of the on-boarded sources
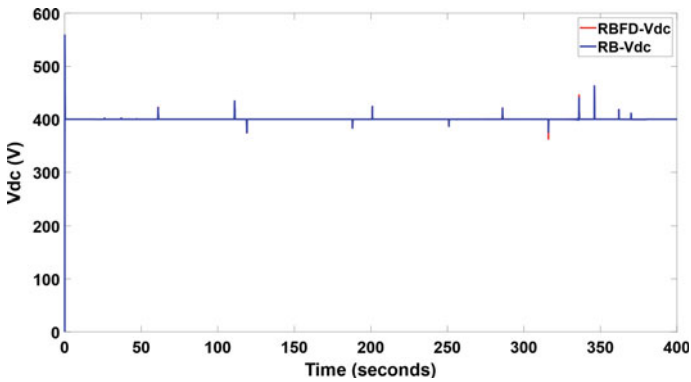
Fig. 10 DC bus voltage

sources while taking into consideration the electrical limitations of each source. This has allowed to reduce the stress on battery by minimizing the RMS of battery current, and in this way, it was possible to increase the lifetime of our Li-ion battery.

References

1. Xu B et al (2020) Ensemble reinforcement learning-based supervisory control of hybrid electric vehicle for fuel economy improvement. IEEE Trans Transp Electrif 6(2)
2. Balali Y, Stegen S (2021) Review of energy storage systems for vehicles based on technology, environmental impacts, and costs. Renew Sustain Energy Rev 135(July 2020):110185. https://doi.org/10.1016/j.rser.2020.110185
3. Sadeq T, Wai CK, Morris E, Tarboosh QA, Aydogdu O (2020) Optimal control strategy to maximize the performance of hybrid energy storage system for electric vehicle considering topography information. IEEE Access 8:216994–217007
4. Khaligh A, Li Z (2010) Battery, ultracapacitor, fuel cell, and hybrid energy storage systems for electric, hybrid electric, fuel cell, and plug-in hybrid electric vehicles: State of the art. IEEE Trans Veh Technol 59(6):2806–2814. https://doi.org/10.1109/TVT.2010.2047877
5. Cao J, Emadi A (2012) A new battery/ultracapacitor hybrid energy storage system for electric, hybrid, and plug-in hybrid electric vehicles. IEEE Trans Power Electron 27(1):122–132. https://doi.org/10.1109/TPEL.2011.2151206
6. Xiong R, Chen H, Wang C, Sun F (2018) Towards a smarter hybrid energy storage system based on battery and ultracapacitor—a critical review on topology and energy management. J Clean Prod 202:1228–1240. https://doi.org/10.1016/j.jclepro.2018.08.134
7. Zhang Q, Wang L, Li G, Liu Y (2020) A real-time energy management control strategy for battery and supercapacitor hybrid energy storage systems of pure electric vehicles. J Energy Storage 31:101721. https://doi.org/10.1016/j.est.2020.101721
8. Generic battery model—Simulink. https://www.mathworks.com/help/physmod/sps/powersys/ref/battery.html?s_tid=srchtitle
9. Implement generic supercapacitor model—Simulink. https://www.mathworks.com/help/physmod/sps/powersys/ref/supercapacitor.html
10. Herizi O, Barkat S (2016) Backstepping control and energy management of hybrid DC source based electric vehicle. In: 4th International symposium on environmental friendly energies applications (EFEA) 2016
11. Mossadak MA, Chebak A, Elmahjoub AA (2021) Multi-level architecture modeling of an intelligent energy management strategy for battery/supercapacitor electric vehicle. In: 2021 3rd International conference on transportation and smart technologies. TST 2021, pp 65–71, May 2021.https://doi.org/10.1109/TST52996.2021.00018
12. Xu Q, Vafamand N, Chen L, Dragicevic T, Xie L, Blaabjerg F (2021) Review on advanced control technologies for bidirectional DC/DC converters in DC microgrids. IEEE J Emerg Sel Top Power Electron 9(2):1205–1221
13. Abdelhedi R, Ammari AC, Sari A, Lahyani A, Venet P (2016) Optimal power sharing between batteries and supercapacitors in electric vehicles. In: 2016 7th International conference on sciences of electronics, technologies of information and telecommunications (SETIT), Dec 2016, pp 97–103. https://doi.org/10.1109/SETIT.2016.7939849

Controlling Powered Two-Wheeled Vehicles in Bends Using Machine Learning

Fakhreddine Jalti, Hajji Bekkay, and Abderrahim Mbarki

Abstract The artificial intelligence is experiencing tremendous development due to its multiple uses conquering wide range of fields; however, few predictive models are used to estimate the dynamics parameters of two-wheeled vehicles. This work focuses on the prediction of the parameters that contribute the most to vehicle dynamics in cornering using artificial intelligence techniques. Two artificial intelligence algorithms, namely the K-nearest neighbor (KNN) and the artificial neural network (ANN) were tested and evaluated for road adherence prediction (regression context) and predicting a scenario of losing control of the two-wheeled vehicle (classification context). The test results show that the KNN technique is more performant in the prediction of road grip, while the ANN technique is better in the prediction of a loss of control scenario. By adjusting the optimal conditions for our chosen ML model, we succeeded to reach an accuracy percentage that exceeds 98.77%.

Keywords Powered two wheeler (PTW) · Road grip · Critical driving events · Cornering speed · Machine learning

1 Introduction

Since the appearance of road traffic, drivers of powered two-wheeled vehicles (PTW) presented the most vulnerable category on the roads, indeed a fatal motorcycle accident is 22 times more likely to occur than a fatal accident by car. Several causes have been singled out for this alarming figure, including the lack of driver assistance systems. Moreover, the complexity of the dynamics of PTWs constitutes a brake on their development. As for AI in relation with this transport field, some works that

F. Jalti (✉) · H. Bekkay
Laboratory of Renewable Energy, Embedded System and Information Processing, National School of Applied Sciences, Mohammed First University, 60000 Oujda, Morocco
e-mail: jalti.fakhreddine@ump.ac.ma

A. Mbarki
National School of Applied Sciences, Mohammed First University, 60000 Oujda, Morocco

were conducted ([1–4]) still did not cover the full range of areas of assistance needed for PTW users.

The objective of this work is to present specialized driving aids for PTWs, which will take into account the environment especially the road grip in order to ensure good control of the vehicle in question. To do this work, several learning algorithms have been implemented in order to make a suitable choice with minimal error. In the context of the situations faced, it was necessary for us to work with classification and regression both by calling each time on two algorithms (K-nearest neighbor (KNN) and artificial neural network (ANN)).

After implementing these algorithms, we conducted a comparison that showed the efficiency of KNN regression algorithm in the prediction of road grip based on velocity. But in order to strengthen the method of maximum speed prediction, we have chosen the classification-based approach which is destinated to serve detect critical events in driving situations, where we concluded that ANN is more precise.

So, organization of the paper will be as follows: Sect. 2 is dedicated to the problem statement and resolution method, Sect. 3 describes the results that we get after implementing the learning models and their efficiency, and Sect. 4 the conclusion.

2 Problem Statement and Resolution Method Presentation

2.1 Problem Statement

Consisting of a triplet {Driver, Vehicle, Environment} our system is meant to deal with dangerous situations [5] which until now modeled by the dynamic parameters, this conventional method imposes an expensive equipment by the sensors however does not guarantee a prevention because, unlike the four wheeled, the PTW has a very weak capacity to correct their dynamic state if a critical situation is already presented.

So, in this article, we are focusing on risk prevention using AI to estimate adherence loss while taking a curve. We are using a dataset of PTW having undergone the experience of falling on bends [6]. We compared different technics of ML, then we studied the accuracy of our selected AI algorithm by simulating real-time data.

2.2 Experimental Protocol and AI Development

The detailed study [7] on detection of critical situations for PTW used instrumentalized motorcycle by a data collection unit in order to record the dynamic parameters.

Our developed model (Fig. 1) uses the results of the study on the driving sequences carried out: normal driving sequence and driving sequence with fall.
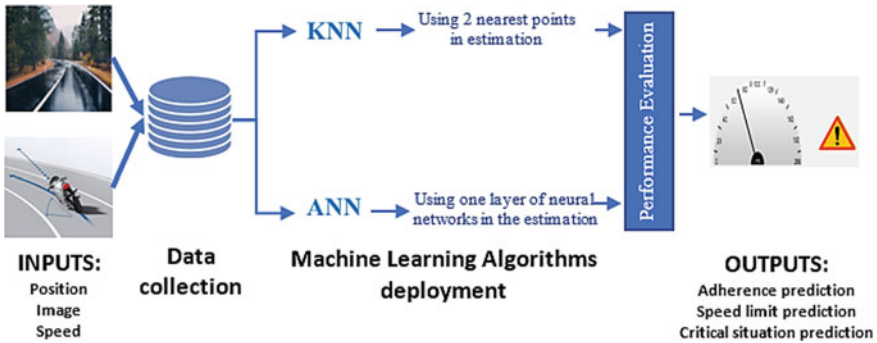
Fig. 1 AI learning model

The first part of the project is devoted to predict the value of road grip using the regression-based approach. And the second part is dedicated to the detection of PTW slippage based on the event classification approach.

Thus, our model consists of a machine learning module which uses two AI technics, regression and classification, in the prediction of critical scenarios.

2.3 Machine Learning Algorithms Implementation

In this section, we are giving an overview on different ML technics applied on our dataset; we can roughly consider such an algorithm a learning function f which maps the input variables X ($x_1, x_2, \ldots, x_n$) to output variables Y ($y_1, y_2, \ldots, y_n$).

$$Y = f(x) \tag{1}$$

In our computer development, we are using Python because of suitability to AI algorithms. Thus, to implement artificial neural network (ANN) and the K-nearest neighbor (KNN).

K-nearest neighbor uses "feature similarity" to predict the values of any new data point. This means that a value is assigned to the new point based on how it resembles the points in the training set. The output result is flowing by Eq. (2)

$$\tilde{y}(\tilde{x}) = \sum_{i=1}^{k} w_i \, y(x_i) \tag{2}$$

where w_i weighting factor for x_i, k number of neighbors, y response variable, and $\tilde{y}$ response variable prediction.

Fig. 2 Artificial neural
network methodology
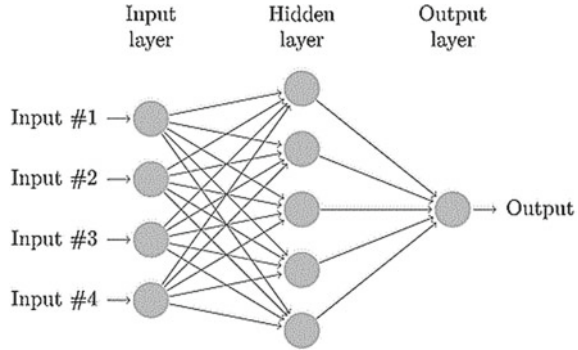
Artificial neural network. A multi-layered neural network or perceptron is a percep-
tron that works with additional perceptron, stacked in multiple layers, to solve
complex problems. Backpropagation is a method of updating the coefficients of
a neural network by taking into account the true and desired outputs. The deriva-
tive with respect to each coefficient w is calculated using the chain rule. Using this
method, each coefficient is updated by Eq. (3):

$$w = w - \alpha \frac{\partial L(\hat{y}, y)}{\partial w} \tag{3}$$

where $-\hat{y}, y, \alpha, L$ are, respectively, predicted value, real, learning rate, the cost
function to be minimized.

It is done in three stages: Calculation of the cost => calculation of the gradient
=> the gradient to update the coefficients w.

We therefore need to find the optimal structure of the model, that is to say: number
of simple layers (with single neurons), number of neurons in each layer, the activation
function to be used for neurons: the ReLU function (Fig. 2).

The flowchart (Fig. 3) presents the machine learning procedure used to link the
inputs with the outputs that are represented in Table 1. The initial parameters for the
KNN and ANN models are configured. Next, training and testing data are injected.
Finally, we get the expected output results.

With:

2.4 Performance Evaluation

Machine learning is the discipline of applying algorithms to a set of data in order
to extract models from it. Certainly, for a regression problem (our case), there are
different types of machine learning models. Therefore, we will study the different
ML algorithms in order to compare their performances.
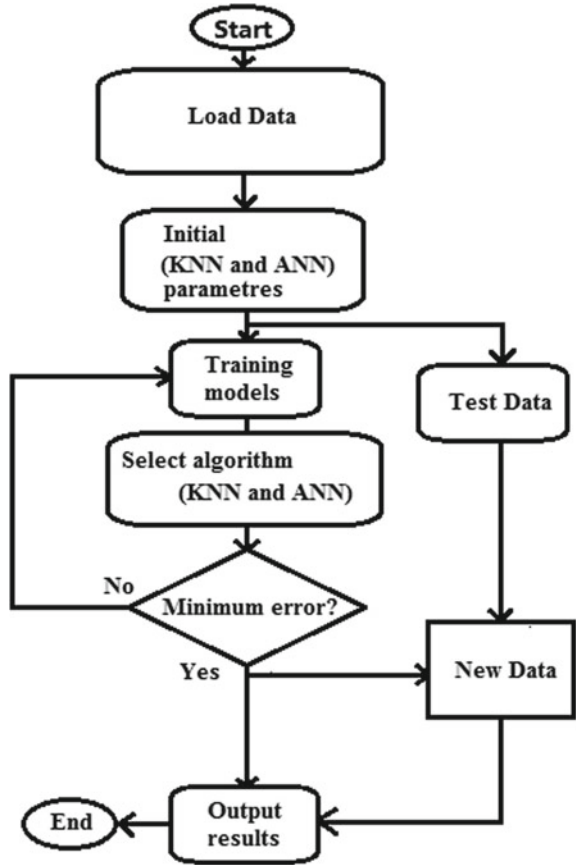
Table 1 Input and outputs
variables definition

	Inputs	Outputs
Regression	{Velocity, curvature}	{Road grip}
Classification	{Road state, Velocity, Curvature,	{Control loss}

As a conclusion, we found that we should improve the method of prediction by combining the regression algorithm with that of the classification of critical events; it should be based on a determining parameter: the recognition of the state of the road.

In this section, we use the following abbreviations for each algorithm: K-nearest neighbor (KNN), artificial neural network (ANN).

Quality of the prediction, we use Pearson's linear coefficient of determination, denoted R^2, which indicates how closely the predicted values correlate with the true values Eq. (4):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \tag{4}$$

The mean absolute error (MAE): Models are evaluated on the test data by the absolute error between the actual value and the predicted value Eq. (5):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

where n, y_i, $\hat{y}_i$, $\overline{y}_i$, are, respectively, the number of measurements, the value of measurement i, the corresponding predicted value, the mean of the measurements.

3　Results and Discussion

3.1　Road Grip Prediction Results Using Regression

In this section, considering the nature of output, we will be preferring regression ML technics by using the KNN and ANN algorithms that go with regression.

As an output (Table 1), the road grip is being evaluated using test data as a function of velocity (input). In Fig. 4, we can compare the prediction results given by the two technics (KNN ▽and ANN *) with experimental data (○).
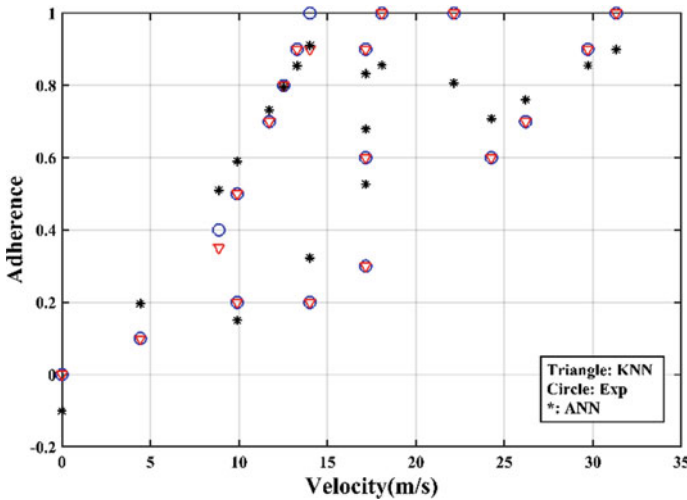
Fig. 4 Road grip as function of velocity for different ML regression algorithms
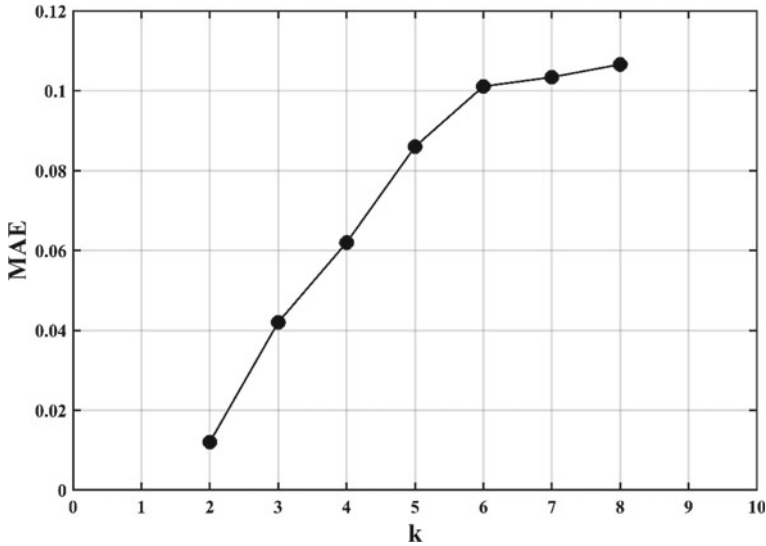
Fig. 5 K-nearest neighbor as a function of MAE

We can clearly observe that the predicted road grip given by the KNN model is the closer to the experimental values. More than that, ANN is instable given its probabilistic behavior. By calculating R^2 indicator, we found that the ANN algorithm gives different values that variate from 25 to 90%.

After realizing the efficiency of the KNN algorithm in front of the ANN, we are evaluating its precision by acting on its parameter K (Fig. 5).

We can clearly state on the negative correlation between the parameter K and the MAE indicator, that means the less we have close values (smallest k), the more the prediction of road grip is precise, thus our selected value of $k = 2$ which gives an accuracy of 98.77%

3.2 Improved Detection Method of Critical Events in Driving Mode

Prediction of grip on a curve using only velocity has its limits; the model needs to be reinforced with a determining module. We have chosen to use image recognition to determine road adherence, this technic is widely developed, and datasets are available.

Hence, we trained our classification algorithm on virtual dataset by using NHTSA formula in speed calculation. By doing so, we made it sure that our model is learning values close to theoretical maximum speed (Eq. 6).

Table 2 Variable's values definition

	Road state	Curvature	Speed	Critical situation
Description of variable	{0; 1; 2; 3} designating: {slippy, wet, normal, dried}	{0.01; 0.02; 0.03; 0.04; 0.05}	[0, 205 km/h]	{0; 1}

$$V_x = \sqrt{\frac{g}{\rho} \frac{\phi r + \mu_{\text{lat}}}{1 - \phi r \cdot \mu_{\text{lat}}}} \tag{6}$$

where μ_{lat}, ρ, v_x, ϕ_r are, respectively, coefficient of static friction, the curvature of the turn, the longitudinal speed, the tilt angle of the road.

As for ML algorithm, we roughly can express it as (Eq. 7):

$$Y = f(X) \tag{7}$$

With: $X = $ (Road state, curvature, Speed) and $Y = $ Critical situation

Where (Table 2):

In order to detect critical events (cases of fall) in the driving of a PTW, we used the Python software to implement K-nearest neighbor (KNN) and artificial neural network (ANN).

We split the dataset randomly into two subsets for which we variate the rate of training data to see its impact on precision R^2.

Figure 6 shows that the ANN model is more precise than the KNN. On the other hand, we can observe the positive correlation between the rate of training data and R^2 indicator of precision, it means that the more we increase the rate of training data, the more we get precision on the prediction of the critical situation state (Fig. 7).

The same statement can be made on the comparison between ANN and KNN using MAE indicator. On the other hand, the ANN could be configured by acting on the training data rate to get an accuracy of 88.13% (R^2 indicator).

3.3 Prevent Loss of Control When Cornering

Our developed model is able to predict the road grip using a regression algorithm, then distinguish whether the vehicle is losing control by the classification algorithm. In both cases, we can reach an excellent accuracy of prediction reaching 98.77%. Thus, the maximum velocity can be predicted to prevent a critical situation (Fig. 8).
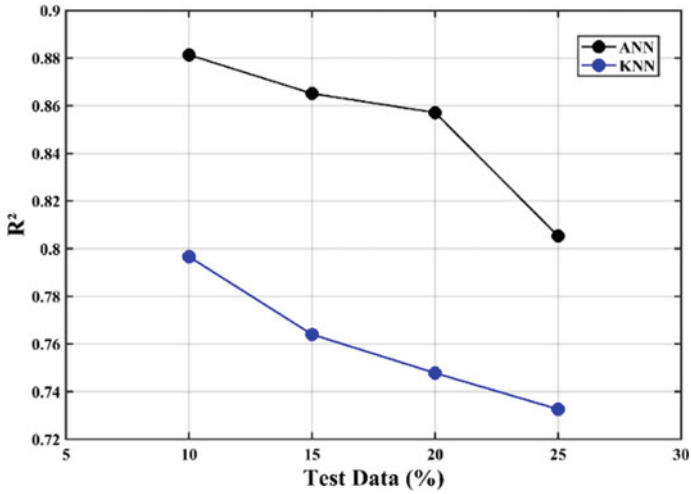
Fig. 6 Testing data as a function of R^2
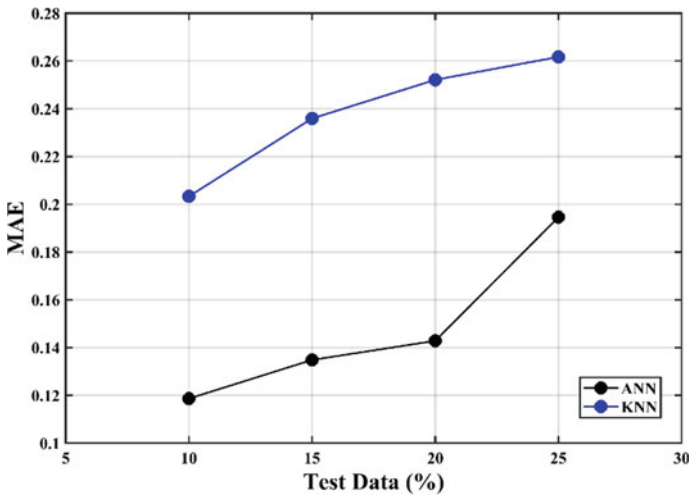
Fig. 7 Testing data as a function of MAE

4 Conclusion

The work carried out aimed to contribute to the reduction of the vulnerability of powered two-wheeled vehicles by predicting a critical situation.

We had the opportunity to conquer the field of machine learning and apply it to PTWs by analyzing and preparing data for machine learning as well as comparing the different types and algorithms of automatic/machine learning.
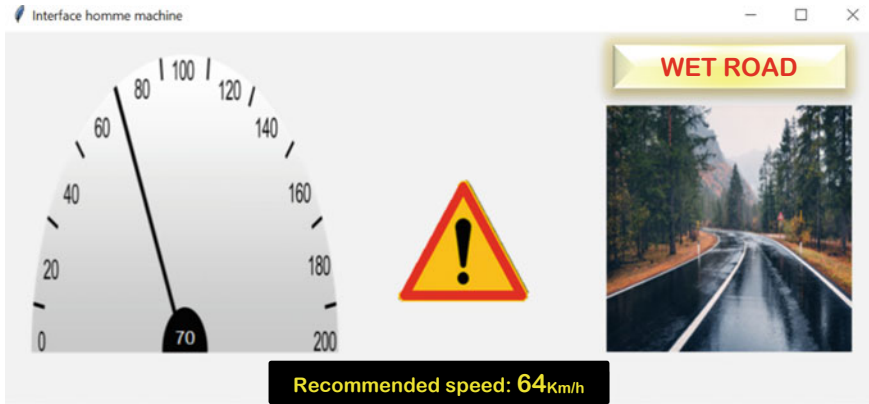
Fig. 8 HMI loss control prevention

Considering that the loss of control in bends and the unsuitable speed are one of the major causes of accidents for PTW; in the first, place we opted for the development of the learning model to predict road grip. Secondly, we presented the supervised classification approaches allowing to detect critical events in the driving of a PTW from data collected via simulated sensors and a camera.

The results obtained are satisfactory, with an accuracy reaching 98.77%; our developed model is able to predict whether the vehicle is about to face a critical situation, and with an HMI, the driver is capable to avoid the danger.

Our model is meant to be developed in further steps; we will detail more the image processing in order to have more precision on the prediction.

References

1. Suryana A, Familiana H (2019) Automatic braking system on motorbikes using the concept of kinematics non-uniform slowing down motion for safety of motorcycle riders on the highway. In: International conference on ICT for smart society (ICISS). IEEE, Bandung, Indonesia
2. Hecker S, Liniger A, Maurenbrecher H, Dai D, Van Gool L (2019) Learning a curve guardian for motorcycles. In: Intelligent transportation systems conference (ITSC). IEEE, Auckland, New Zeeland
3. Ferhat A (2015) Powered two-wheeler riding pattern recognition using a machine-learning framework. IEEE Trans Intell Transp Syst 16(1)
4. Ferhat A (2013) Riding patterns recognition for Powered two-wheelers users' behaviors analysis. In: 16th International annual conference on intelligent transportation systems, ITSC, The Hague, Netherlands. IEEE
5. Jalti F (2020) New approach for controlling PTW vehicle dynamics: characterization of critical scenarios. In: Bennani S, Lakhrissi Y, Khaissidi G, Mansouri A, Khamlichi Y (eds) WITS. Lecture notes in electrical engineering, vol 745. Springer, Singapore

6. Boubezoul A, Dufour F. Bouaziz S, Larnaudie B, Espié S (2019) Dataset on powered two wheelers fall and critical events detection. Data Brief 23:103828
7. Attal F (2015) Classification de situations de conduite et détection des événements critiques d'un deux roues motorisé. Mathématiques générales [math.GM]. Université Paris-Est. Français. NNT :2015PESC1003

Modeling and Analysis of a Fuel Cell-Battery Hybrid Electric Vehicle

M. Essoufi, Hajji Bekkay, and Abdelhamid Rabhi

Abstract In recent years, the lack of energy and the emission of greenhouse gases has led the research community to devote time and effort; to develop solutions for the new generation of vehicles. In this context, modeling and numerical simulation bring real added value because they reduce the cost and time of development. This article deals with the modeling and performance analysis of a fuel cell-battery hybrid electric vehicle (FCHEV). The power supply system consists of a fuel cell as a primary source and a battery as a secondary one. The various topologies are analyzed, and the commands used to control the current and voltage of the DC bus are presented. The performance of the proposed FCHEV model is verified by numerical simulations for the Urban Dynamometer Driving Schedule (UDDS), using MATLAB Simulink, considering a rule-based energy management strategy. Finally, the results show the performance and efficiency of the drive system under different operating modes of the FCHEV over a complete cycle: acceleration, deceleration, battery charge during acceleration, and regenerative braking.

Keywords Fuel cell · Hybrid vehicle · Battery · Energy management · Modeling · Rule-based strategy

1 Introduction

Greenhouse gas emissions are a direct cause of climate change. In its first assessment report, the Intergovernmental Panel on Climate Change (IPCC) predicted that carbon dioxide (CO_2) concentrations would double between 2025 and 2050 with-

M. Essoufi (✉) · H. Bekkay
Renewable Energy, Embedded System and Data Processing Laboratory,
National School of Applied Sciences, Mohamed First University, Oujda, Morocco
e-mail: m.essoufi@ump.ac.ma

A. Rabhi
Modelization, Information and Systems Laboratory, Picardie Jules Verne University, Amiens, France

out significant changes in human activity. In addition, the IPCC Fifth Assessment Report found that greenhouse gases and carbon dioxide emitted by fossil fuels have the greatest impact on global warming. In light of these findings, one of humanity's priority concerns for its survival and that of the planet is the reduction of greenhouse gas emissions, particularly CO_2 [6].

More specifically, among all the sectors responsible for pollution, we find the automobile sector. Indeed, the number of vehicles in the world is more than one billion. A number that is growing exponentially, particularly in recent years. Almost all vehicles on the road are powered by an internal combustion engine (ICE). However, even under optimal operating conditions, this type of engine can only achieve an efficiency of about 36% for a gasoline and 42% for a diesel. This means that a significant part of the energy supplied by the fuel is wasted. As a result, the internal combustion engine industry is responsible for about 32% of carbon dioxide (CO_2) emissions.

The need to minimize CO_2 emissions and greenhouse gases has led to the emergence of hybrid vehicles in recent years. One of them is the hybrid electric vehicle that uses the fuel cell as primary source. Zero emissions, high efficiency, low noise and high energy density are the main advantages of fuel cells. However, when used as the single source, fuel cells are not able to respond to fast fluctuations of power requirements due to their slow dynamics, and their non-reversibility does not allow for the storage of energy recovered by regenerative braking. Therefore, at least two energy sources must be combined. The fuel cell acts as the primary source, while the secondary source is an energy storage device such as batteries or supercapacitors, which supply the vehicle's power demand peaks, and store the recovered energy. However, the combination of these two sources requires an optimal topology, as well as adequate energy management strategy to ensure power distribution between these sources [7].

This paper is devoted to the modeling of the fuel cell-battery hybrid electric vehicle using MATLAB Simulink. To test the vehicle model, a deterministic rule-based energy management strategy is used. The objective of this modeling and simulation is to recognize the performance factors and behaviors of different vehicle parts to guide the design and optimization of such vehicles.

2   Vehicle Topologies

The main electrical power components of a fuel cell-battery hybrid electric vehicle are: the electrical machine (AC motor), the fuel cell system and secondary energy source (Battery). In order to connect these elements on one bus, it is necessary to define an electrical architecture.

In all topologies, the AC motor is systematically equipped with a DC–AC converter to control its torque. The converter is reversible in power to supply the necessary energy during the traction phases and to restore the energy produced by the electric machine during the braking phases of the vehicle [10].
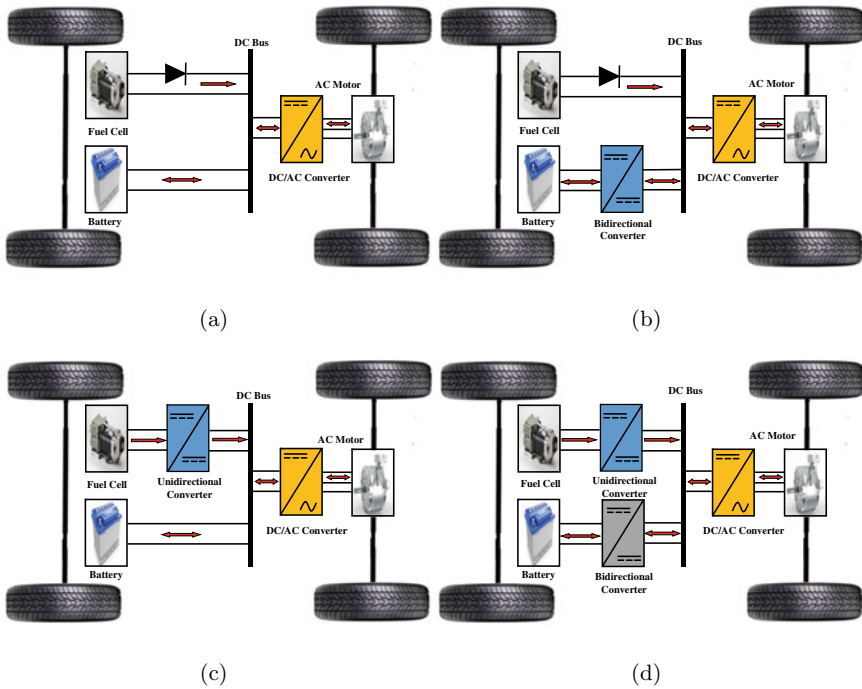
Fig. 1 Topologies of fuel cell-battery hybrid electric vehicle

Figure 1 shows the different FCHEV topologies reported in the literature:

- In the first topology (Fig. 1a), the fuel cell and the battery are connected directly to the DC bus. With this topology, its difficult to control the DC bus voltage, but the benefit is its simplicity [10].
- The second topology (Fig. 1b) shows that the fuel cell is connected directly to the DC bus and imposes its voltage, while the battery is interfaced via a bidirectional DC–DC converter. This topology is rarely used due to the higher losses of the bidirectional converter [13].
- In the next topology (Fig. 1c), the fuel cell is connected to the DC bus via an unidirectional DC–DC converter, while the battery is connected directly [13]. In this topology, the battery imposes its voltage on the DC bus. This configuration is widely used because of its simplicity, the reduced number of converters and the ease of power split control.
- The last one (Fig. 1d) shows that all sources are interfaced by DC–DC converters. This configuration is the most preferred because it offers two degrees of control, allows to freely control the DC bus voltage, to protect the sources, and to better control the power flow.
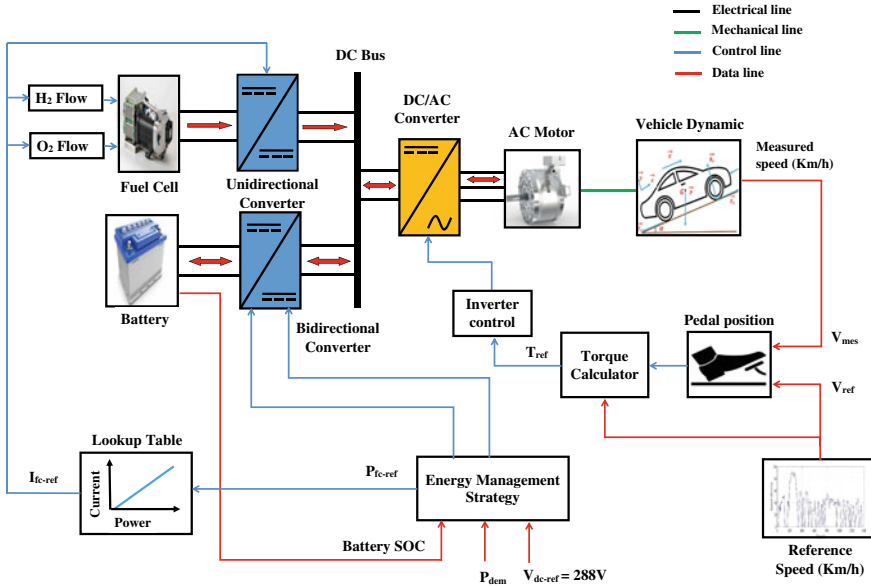
Fig. 2 General block diagram of the vehicle modeling

3 Hybrid Electric Vehicle Modeling

In this study, the architecture of Fig. 1d is adopted. The general block diagram of the implemented vehicle modeling system is presented in Fig. 2.

3.1 Vehicle Model

All applied forces to the vehicle in motion are illustrated in Fig. 3 [3].

The vehicle dynamics take into account the specific coefficients of the studied vehicle and the imposed dynamic performances. Thus, according to Newton's second law, and based on Fig. 3, the vehicle dynamics can be expressed by the Eq. (1).

$$M_v \frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \vec{R}_r + \vec{F}_{air} + \vec{F}_r + \vec{F}_t + \vec{P} \tag{1}$$

The principal forces acting on the vehicle are:

$$\vec{F}_{air} = -\frac{1}{2}\rho_{air}V^2 A C_x \vec{x} \tag{2}$$

Fig. 3 Forces applied to the vehicle
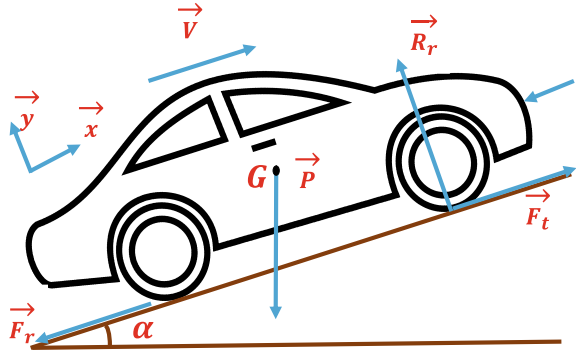
$$\overrightarrow{F_{\mathrm{r}}} = -PC_{\mathrm{r}}\cos\alpha\,\overrightarrow{x} \tag{3}$$

$$\overrightarrow{P} = -M_{\mathrm{v}}g\sin\alpha\,\overrightarrow{x} \tag{4}$$

where $\overrightarrow{F_{\mathrm{air}}}$ represents the aerodynamic force, which is the resistance force of the air in the direction of the vehicle motion. $\overrightarrow{F_{\mathrm{r}}}$ is the resistance of the wheels on the floor, and $\overrightarrow{P}$ the vehicle's gravity force [3].

From the Fig. 3 and the Eqs. (1–4), the expression of the mechanical tensile force F_{t}, which is the traction force of the vehicle is defined by the Eq. (5).

$$F_{\mathrm{t}} = M_{\mathrm{v}}\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{1}{2}\rho_{\mathrm{air}}V^2AC_{\mathrm{x}} + M_{\mathrm{v}}g\sin\alpha + M_{\mathrm{v}}gC_{\mathrm{r}}\cos\alpha \tag{5}$$

The mechanical power P_{m} is expressed as a function of the total traction torque and the rotation speed of the wheels as follows:

$$P_{\mathrm{m}} = C_{\mathrm{t}} \cdot \Omega_{\mathrm{wheel}} \tag{6}$$

where $C_{\mathrm{t}} = F_{\mathrm{t}} \cdot r$ and $\Omega_{\mathrm{wheel}} = \frac{v}{r}$.

So the mechanical power needed to move the vehicle is given by the Eq. (8).

$$P_{\mathrm{m}} = v \cdot F_{\mathrm{t}} \tag{7}$$

$$P_{\mathrm{m}} = v\left(M_{\mathrm{v}}\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{1}{2}\rho_{\mathrm{air}}V^2AC_{\mathrm{x}} + M_{\mathrm{v}}g\sin\alpha + M_{\mathrm{v}}gC_{\mathrm{r}}\cos\alpha\right) \tag{8}$$

Fig. 4 Model of the fuel cell [9]
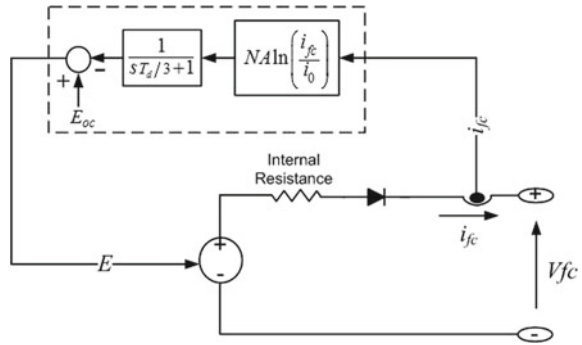
3.2 Fuel Cell Model

The fuel cell converts chemical energy (hydrogen and oxygen) into electrical energy. This electrochemical reaction is a reversed electrolysis of water. The electrical energy is produced by an redox reaction between oxygen (oxidizer) and hydrogen (reducer). The oxidation takes place at the anode and the reduction at the cathode. These two reactions are isolated by a membrane that plays the role of electrolyte. The cathode is supplied with oxygen gas (generally air), and the anode is supplied with hydrogen gas. The chemical equations of oxidation–reduction are the following [1]:

- Anode oxidation
  $H_2 \longrightarrow 2H^+ + 2e^-$
- Cathode reduction
  $O_2 + 2H^+ + 2e^- \longrightarrow H_2O$
- Global reaction
  $H_2 + 1/2O_2 \longrightarrow H_2O + heat$

Figure 4 depicts the dynamic model of the fuel cell available in MATLAB/ Simulink [9]. This model is represented by a controlled voltage source in series, with an internal resistor and a diode to protect the fuel cell against reverse currents. The voltage at the fuel cell output is given by the Eq. (9) [9].

$$V_{fc} = E - R_{ohm} \cdot i_{fc} \tag{9}$$

where the controlled voltage source (E) is defined by the Eq. (10).

$$E = E_{oc} - NA \ln\left(\frac{i_{fc}}{i_0}\right) \cdot \frac{1}{sT_d/3 + 1} \tag{10}$$

The oxygen and hydrogen flow rates for a given reference current are calculated by Eqs. (11) and (12) [9].
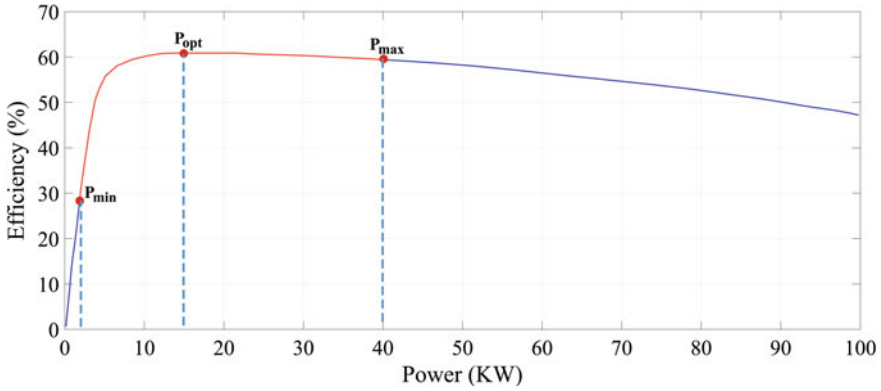
Fig. 5 Fuel cell system efficiency curve [2]

$$U_{O_2} = \frac{60,000 R T N I_{fc}}{z F P_{air} V_{lpm\,(air)} O_2 \%} \tag{11}$$

$$U_{H_2} = \frac{60,000 R T N I_{fc}}{z F P_{fuel} V_{lpm\,(fuel)} H \%} \tag{12}$$

Figure 5 shows the efficiency curve of the fuel cell system. From this figure, we can notice that there is a zone where the efficiency is maximum (red area), so its interesting to operate the fuel cell in this zone to reduce hydrogen consumption [2].

3.3 Battery Model

A battery is an electrochemical cell that transforms chemical energy into electrical energy. Its use in electric vehicles is justified by its high energy density compared to fuel cells. In this model, the lithium-ion battery is used because of its better response compared to other types of batteries and its wide use in the transportation field. For the simulations, the dynamic battery model available in MATLAB/Simulink presented in Fig. 6 is used [11].

Depending on the operating mode, the value (V_{batt}) is calculated by two different equations [11]:

- Discharge mode:

$$V_{discharge} = E_0 - R \cdot i - K \frac{Q}{Q - it} \cdot (it + i^*) + A \exp(-B \cdot it) \tag{13}$$
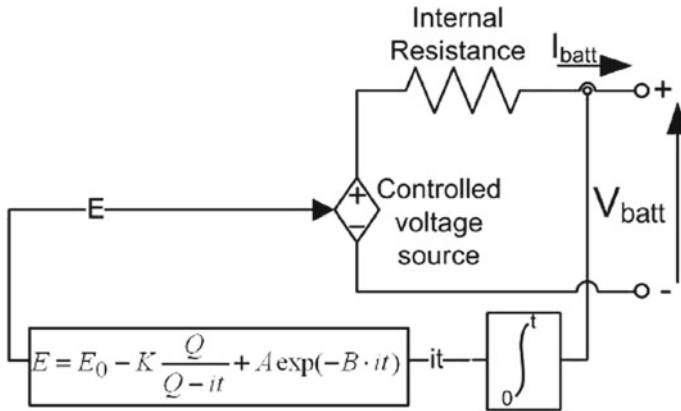
Fig. 6 Nonlinear battery model [11]

- Charge mode:

$$V_{\text{charge}} = E_0 - R \cdot i - K \frac{Q}{it - 0.1Q} \cdot i^* - K \frac{Q}{Q - it} \cdot it + A \exp(-B \cdot it)$$

(14)

The estimation of the battery state of charge is calculated by the Eq. (15).

$$\text{SOC}_{\text{batt}} = 100 \left(1 - \int \frac{i_{(t)} dt}{Q} \right)$$

(15)

4 Energy Management Strategy

The problem of energy management is to find the optimal power split of the vehicle's energy sources to minimize hydrogen consumption and increase its lifetime. In the literature, the different approaches used in this regard can be classified into three categories.

4.1 Rule-Based Strategy

Rule-based energy management strategies are based on the definition of a set of rules, which are designed according to the experience of the designer. These rules can be defined by deterministic rules or by fuzzy rules. This technique is easy to implement in real time but it is difficult to achieve optimal results [12].

4.2 Optimization-Based Strategy

Optimization-based strategies are identified based on the mathematical modeling approach chosen to formulate the energy management problem for a hybrid vehicle. The objectives of these strategies can be quantified by a cost function that can represent fuel economy, energy source sustainability and other objectives. Equality and inequality constraints are defined in the optimization function [4].

4.3 Learning-Based Strategy

Due to the development of artificial intelligence, learning-based energy management strategies have recently received more and more attention and can overcome the challenge and achieve optimal results. These strategies can be divided into several categories based on the feedback available to the algorithm to learn over time. In the literature, different learning-based strategies, such as clustering learning, neural network learning and reinforcement learning, are used to supervise the energy flow in a fuel cell hybrid electric vehicle [8].

In order to test the performance of our vehicle model, the deterministic rule-based energy management strategy is implemented, as shown in the flowchart in Fig. 7. The reference power of the fuel cell is determined based on the power demanded by the vehicle and the variation of the battery SOC. The management strategy must therefore supervise the battery's state of charge, in real time, in order to respond to the vehicle's power demands while avoiding overcharging and deep discharges.

5 Simulation Results and Discussion

To evaluate the performance of the proposed vehicle model, a real-time rule-based energy management strategy is implemented. The fuel cell hybrid electric vehicle model is simulated using the MATLAB/Simulink environment for the Urban Dynamometer Driving Schedule (UDDS). The results of the simulation are presented and show the efficiency of the model to simulate the different operating modes as well as the proposed energy management is shown the correct distribution of power between the sources, and the battery state of charge is maintained within an acceptable range.

Figure 8 shows the reference speed and the measured speed of the vehicle with the UDDS driving cycle. This driving cycle lasts 1369 s and the distance covered is 11.26 km with an average speed of 30.57 km/h, and a maximum speed of 91.24 km/h [5]. The measured speed follows the reference speed with a minimum error, which confirms the proper functioning of the accelerator pedal system to generate the reference torque to achieve the desired speed.
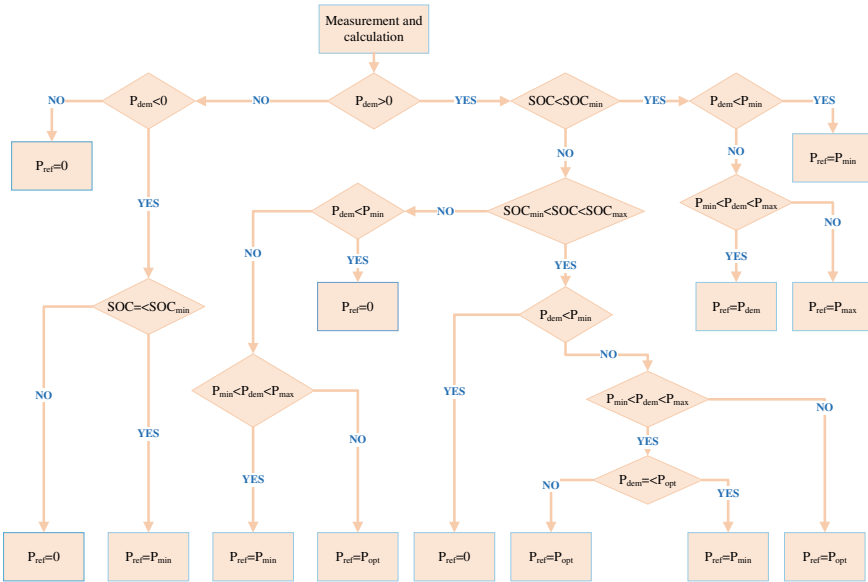
Fig. 7 Rules-based energy management algorithm
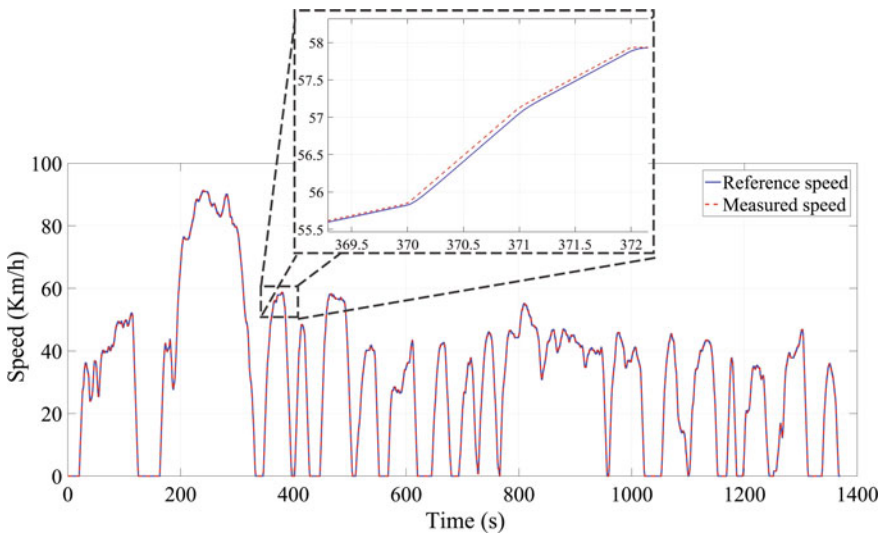
Fig. 8 Speed vehicle for UDDS drive cycle (Reference, measured)
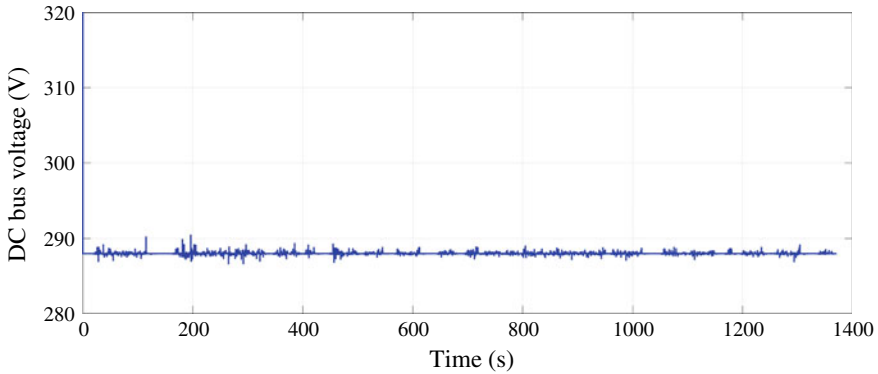
Fig. 9 DC bus voltage
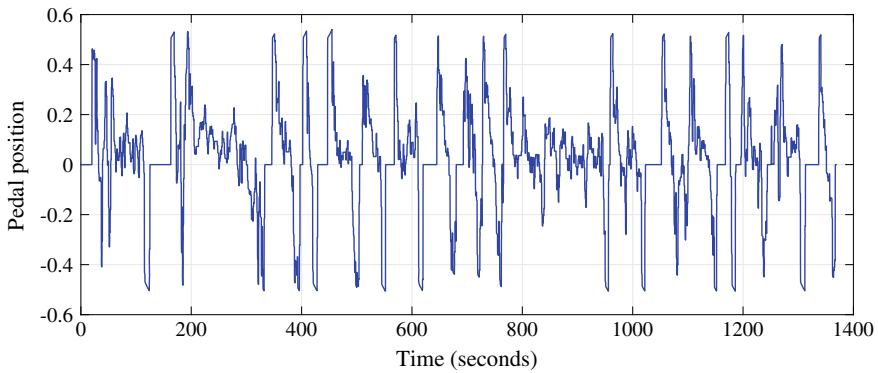
Fig. 10 Accelerator pedal position

Figure 9 shows the evolution of the DC bus voltage, which is kept equal to the reference value of 288 V by the control loop of the bidirectional DC–DC converter. The DC bus has a smooth form, and the perturbations caused by load variation do not influence the constant form of this voltage.

Figure 10 depicts the pedal position, which is determined by the proportional-integral control unit. The basic principle is to compare the desired speed (the drive cycle used) and the actual speed of the vehicle, and from the difference, the traction or braking command (between $+1$ and -1) is inferred.

The traction or braking command, which represents the pedal position, is multiplied by the value of the maximum motor torque to calculate the required torque. On the other hand, the torque available to the motor is determined using a lookup table that has the vehicle speed as input (Fig. 11). Between these two torques, the one with the lower value is chosen as the reference torque for the motor (Fig. 12).
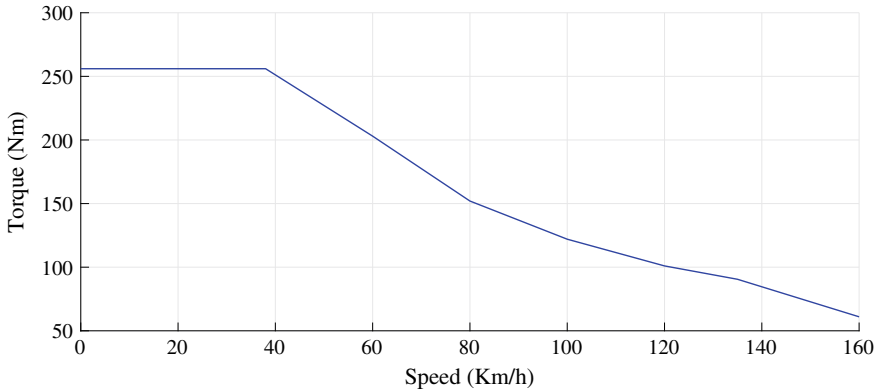
Fig. 11 Motor torque speed characteristic
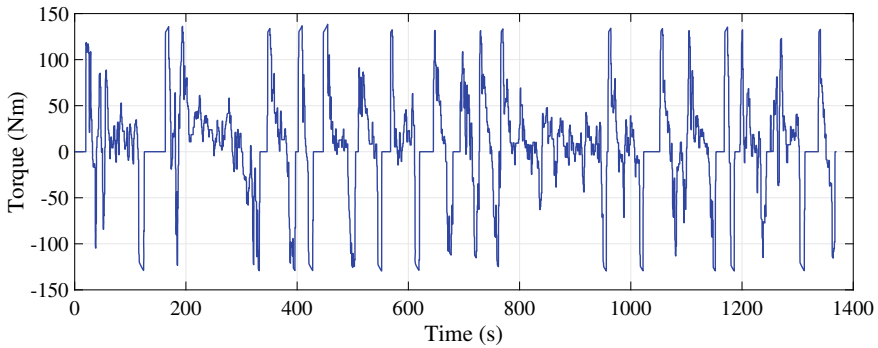
Fig. 12 Reference torque

Figure 13 shows the current demand of the electric motor measured in the DC bus, which corresponds to the reference torque. The three operating modes of the vehicle are shown in the figure :

- Stop mode: zero current.
- Braking mode: negative current.
- Traction mode : positive current.

Figure 14 illustrates the evolution of the power supplied by the two sources as well as the power of the electric motor. The power required by the electric motor is always the sum of the two powers (Fuel cell and battery). The battery is charged by recovering energy from braking, and also assists the fuel cell during the acceleration phases of the vehicle.

The battery state of charge is shown in Fig. 15. The SOC value is bounded by $SOC_{min} < SOC < SOC_{max}$, which varies around its initial value during the driving cycle and the difference between the initial and final SOC is $\Delta SOC = 2.5\%$. According to the Fig. 15, the battery is not solicited enough during the driving cycle,
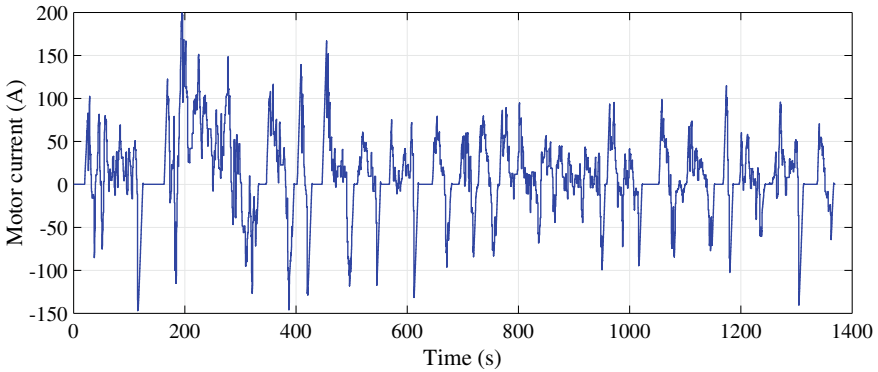
Fig. 13 Motor current
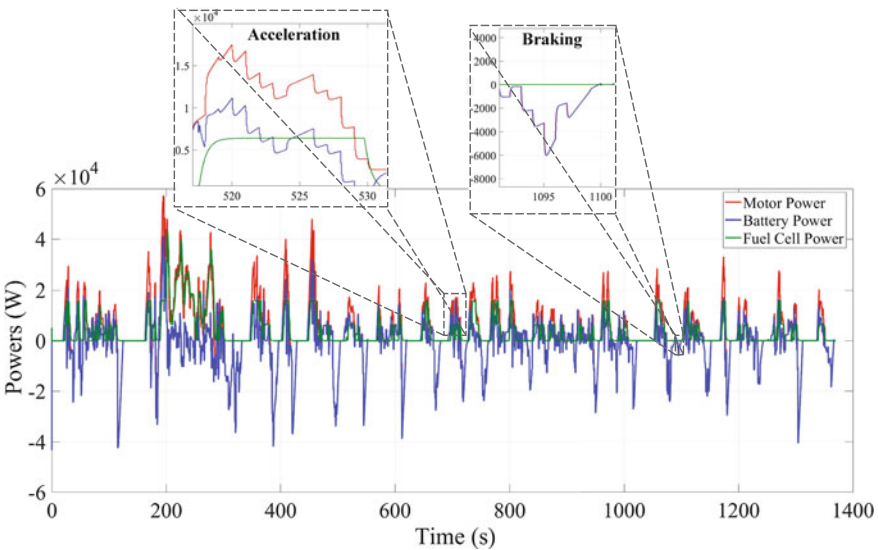
Fig. 14 Powers of battery, motor and fuel cell

except in the case where it recovers the power from braking and supplies the power demand peaks. Therefore, to improve battery utilization, the algorithm of the energy management strategy needs to be developed.

The effectiveness of the proposed vehicle model is validated. Indeed, the different electrical and mechanical elements modeled simulate the behavior of the vehicle. The rule-based energy management strategy perfectly splits the power between the sources. The resulting battery state of charge is maintained within the admissible ranges.
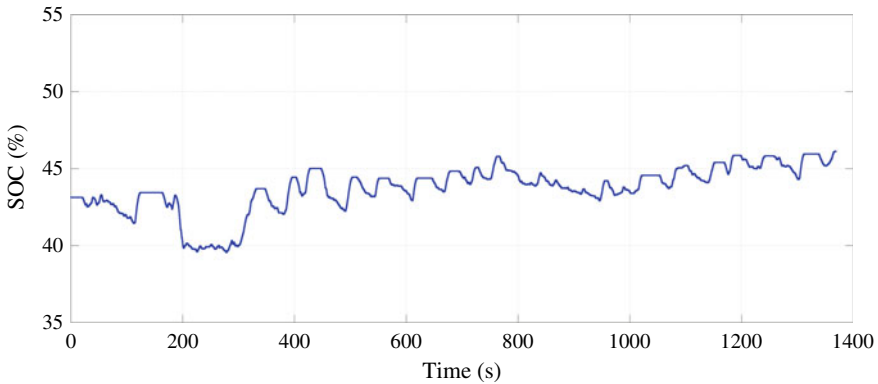
Fig. 15 Battery state of charge

6 Conclusion

In this paper, a fuel cell-battery hybrid electric vehicle model has been proposed and simulated using MATLAB/Simulink. The drive cycle Urban Dynamometer Driving Schedule (UDDS) is selected to evaluate the performance and efficiency of the proposed model.

To test the model, an energy management strategy based on deterministic rules is proposed, and these rules are deduced by expertise. This strategy is based on the comparison of the power demand of the motor and the power of the fuel cell in order to operate it in a high efficiency zone and to maintain the battery state of charge in admissible ranges.

The simulation results give insight into the behavior of the different vehicle components and the prediction of the power demand of the electric motor. In order to improve energy management strategy, to reduce consumption, and increase the lifetime of the different sources, it is necessary to base on this modeling study.

References

1. Abdin Z, Webb CJ, Gray EM (2016) PEM fuel cell model and simulation in Matlab-Simulink based on physical parameters. Energy 116:1131–1144
2. Edwards R, Mahieu V, Griesemann JC, Larivé JF, Rickeard DJ (2004) Well-to-wheels analysis of future automotive fuels and powertrains in the European context. 1072–1084
3. Essoufi M, Hajji B, Rabhi A (2020) Fuzzy logic based energy management strategy for fuel cell hybrid electric vehicle. In: 2020 International conference on electrical and information technologies (ICEIT). IEEE, pp 1–7
4. Ettihir K, Boulon L, Agbossou K (2016) Optimization-based energy management strategy for a fuel cell/battery hybrid power system. Appl Energy 163:142–153
5. Ganji B, Kouzani AZ, Trinh HM (2010) Drive cycle analysis of the performance of hybrid electric vehicles. In: Li K, Fei M, Jia L, Irwi GW (eds) Life system modeling and intelligent

computing. Lecture notes in computer science, vol 6328. Springer, Berlin, Heidelberg, pp 434–444

6. Ko J, Ahuja LR, Saseendran SA, Green TR, Ma L, Nielsen DC, Walthall CL (2012) Climate change impacts on dryland cropping systems in the Central Great Plains, USA. Climatic Change 111(2):445–472

7. Li H, Ravey A, N'Diaye A, Djerdir A (2017) A review of energy management strategy for fuel cell hybrid electric vehicle. In: 2017 IEEE vehicle power and propulsion conference (VPPC). IEEE, pp 1–6

8. Liu Y, Liu J, Qin D, Li G, Chen Z, Zhang Y (2020) Online energy management strategy of fuel cell hybrid electric vehicles based on rule learning. J Cleaner Prod 260:121017

9. Njoya S, Tremblay O, Dessaint LA (2009) A generic fuel cell model for the simulation of fuel cell vehicles. In: 2009 IEEE vehicle power and propulsion conference. IEEE, pp 1722–1729

10. Raga C, Barrado A, Lazaro A, Quesada I, Miniguano H, Zumel P, Sanz M (2014) Optimal sizing of propulsion systems applied to fuel cell based vehicles. In: 2014 IEEE energy conversion congress and exposition (ECCE). IEEE, pp 4797–4803

11. Tremblay O, Dessaint LA, Dekkiche AI (2007) A generic battery model for the dynamic simulation of hybrid electric vehicles. In: 2007 IEEE vehicle power and propulsion conference. IEEE, pp 284–289

12. Wang Y, Sun Z, Chen Z (2019) Development of energy management system based on a rule-based power distribution strategy for hybrid power sources. Energy 175:1055–1066

13. Wu J, Zhang N, Tan D, Chang J, Shi W (2020) A robust online energy management strategy for fuel cell/battery hybrid electric vehicles. Int J Hydrogen Energy 45(27):14093–14107

Interleaved Boost Converter Control Technique Improvements for Fuel Cell Electric Vehicles

Soufyane Benzaouia, Nacer M'Sirdi, Abdelhamid Rabhi, Brahim Khalil Oubbati, and Smail Zouggar

Abstract This study presents some control technique improvements for an interleaved boost converter for fuel cell electric vehicles. *Self-tuning PI* controller and an adaptive *PI* controller using *fuzzy logic* technique have been designed and developed to improve the voltage tracking accuracy. Several scenarios and tests including load disturbances were carried out to evaluate and to compare the performance of the improved controllers with the conventional one. The obtained results show a superior and high performance of the improved techniques compared to the fixed gains *PI* controller without adding any extra sensors.

Keywords Interleaved boost converter (*IBC*) · Proportional–integral (*PI*) controller · Self-tuning *PI* controller · *Fuzzy-logic* concept

1 Introduction

Recently, electric vehicles using fuel cell start gaining more and more attention and importance and representing a promising alternative to the traditional vehicles [1] (Fig. 1). One among the major drawbacks of fuel cells is their low output voltage,

S. Benzaouia (✉) · N. M'Sirdi
LIS UMR CNRS 7020 of Aix Marseille University, Marseille, France
e-mail: soufyane.benzaouia@gmail.com

N. M'Sirdi
e-mail: Kouider-Nacer.Msirdi@lis-lab.fr

A. Rabhi
Laboratory of Modelisation, Information and Systems—MIS, UPJV, Amiens, France
e-mail: abdelhamid.rabhi@u-picardie.fr

B. K. Oubbati
LTSS Laboratory, Laghouat University, Laghouat, Algeria

S. Zouggar
Laboratory of Electrical Engineering and Maintenance—LEEM, University Mohammed 1st, High School of Technology Oujda, Oujda, Morocco
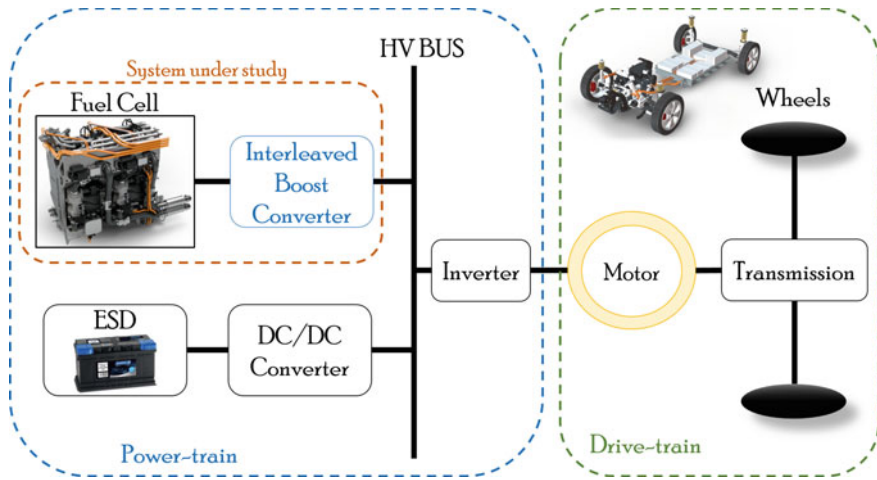
Fig. 1 System configuration of the fuel cell electric vehicle

and the *DC–DC* converter adaptation stage is required so that the output voltage can be step up to higher values. Several *DC–DC* converters topologies have been proposed in literature; one among this converters is the interleaved boost converter (*IBC*). This latter has the advantage of offering high efficiency, low current ripples and reliability [2].

Interleaved boost converter control is an important issue. In literature, many control techniques have been designed, developed and employed for controlling the output voltage. The regulators found in literature are the proportional–integral controller [3], the proportional–integral derivative controller [4] and linear quadratic controller [5]. The poor and the limited performance in over a wide load variation range has made the researchers thought of other more reliable and robust techniques. In [1], the authors propose a dual-loop control scheme based on *H-infinity* controller for three-phase interleaved boost converter. In [2], a robust controller based on extended state observer (*ESO*) is proposed, designed and validated experimentally for an interleaved boost converter for fuel cell application. A robust adaptive controller based on active disturbance rejection control algorithm is proposed in [6].

Proportional–integral controllers are simple and easy to implement, but are less efficient and cannot guarantee a high regulation performance in dynamic regime and face of external disturbance such as load variations, this is because the *PI* controllers do not adapt to changing conditions. *Self-tuning PI* control technique is one of the most interesting approaches to avoid the classical *PI* controller drawbacks. This approach allows an online adaptation without any extra sensors or development complexity. In this study, two *PI* controller improvements have been designed: the first one is called *self-tuning PI* controller and the second one is called *adaptive PI controller* using *fuzzy logic technique*. The two techniques allow avoiding the traditional manual gains tuning solution. A comparison result is provided to show

the efficiency of each technique in case of reference output voltage variation and load disturbance.

This article is organized as follows: Sect. 2 presents an overview, an analysis and the modeling of an interleaved boost converter, Sect. 3 focuses on the control strategy, Sect. 4 shows the obtained results with a discussion, and Sect. 5 concludes the paper.

2 Interleaved Boost Converter Overview, Analysis and Modeling

The N-phase interleaved boost converter (IBC) topology is shown in Fig. 2 [6], in each phase, we have one diode D_k, one switch S_k, one inductor L_k and a common capacitor C, $(k = 1, 2, \ldots, N)$. v_{in}, v_o, i_o and r_k represent, respectively, the input voltage, the output voltage, the output current and the circuit lumped parasitic resistance. i_{in} represents the input current, and it is expressed as follows:

$$i_{\text{in}} = \sum_{k=1}^{k=N} i_{Lk} \tag{1}$$

where i_{Lk} represents the current flowing through the inductor L_k. Considering a symmetrical circuit parameters, therefore,

$$L_k = L, \ r_k = r, \ d_k = d, \quad k = 1, 2, \ldots, N \tag{2}$$
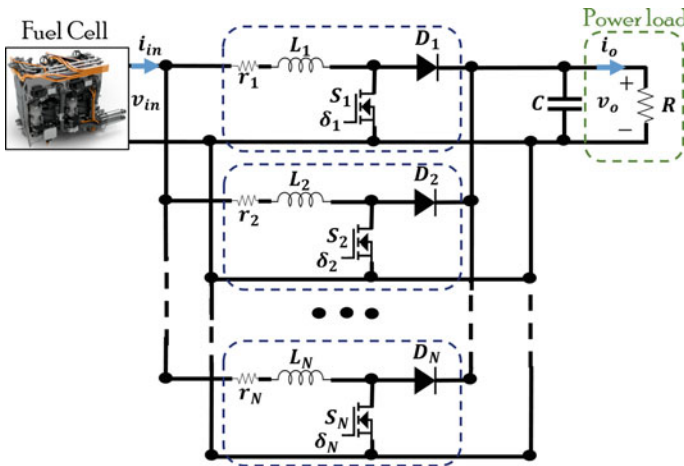
Fig. 2 N-phase IBC
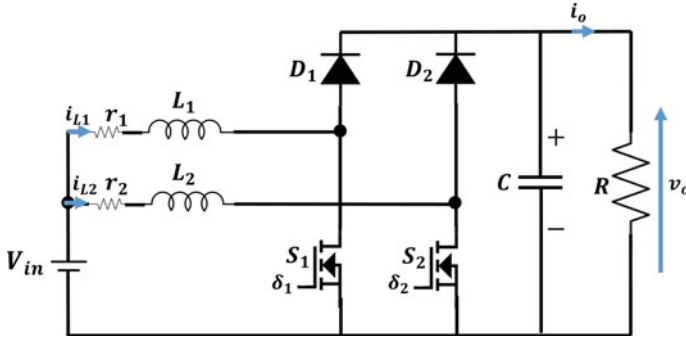
Fig. 3 Topology of a two-phase interleaved boost converter

d_k is the duty cycle/control signal for the switch S_k. δ_k is the switch state (ON–OFF).

In continuous conduction mode, the static voltage gain of an ideal IBC is given as follows:

$$M(d) = \frac{V_o}{V_{\text{in}}} = \frac{1}{1-d} \tag{3}$$

V_o and V_{in} represent, respectively, the dc values of v_o and v_{in}, and d represents the duty cycle.

In this paper, a DC–DC two-phase interleaved boost converter (Fig. 3) is considered ($N = 2$). The average mode of the two-phase IBC is given as follows [2]:

$$\begin{cases} \frac{d}{dt} i_{Lk} = \frac{1}{L_k}(v_{\text{in}} - r_k i_{Lk} - (1 - d_k)v_o), \quad k = 1, 2 \\ \frac{d}{dt} v_o = \frac{1}{C}\left(\sum_{k=1}^{k=2}(1 - d_k)i_{Lk} - i_o \right) \end{cases} \tag{4}$$

IBC operation: In case of an ON switch state ($\delta_k = 1$), the inductor starts charging via the input source, and capacitor discharges to supply the load. In case of an OFF switch state ($\delta_k = 0$), the input source and the inductor charge the capacitor and supply the load.

3   IBC Control Strategy

For fuel cell application, this converter kind suffers from two factors: the variation of the input voltage and the change of the load current [2]. In order to regulate the output voltage to a constant desired voltage, a cascade control configuration is employed. Figure 4 shows the cascade control block diagram of N-phase interleaved boost converter.
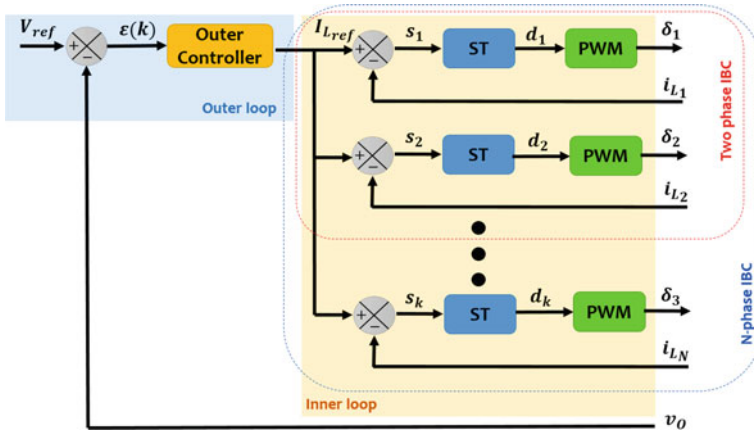
Fig. 4 Control scheme for interleaved boost converters

For a two-phase *IBC*, a dual-loop control strategy is used. It is composed of the following:

- An outer voltage control loop used to regulate the output voltage v_o to its reference V_{ref} and generating the reference current $I_{L\text{ref}}$ to the inner loops.
- Two current controllers for the inner current control loop, used to regulate the inductor currents i_{L1} and i_{L2} to their reference $I_{L\text{ref}}$ and generating the duty cycles d_1 and d_2. These latter are used to generate the control signals δ_1 and δ_2 for the *IBC* using a *PWM* block. For *N*-phase *IBC*, the *ON–OFF* signals are phase shifted by 360°/N, which means by π in two-phase *IBC* case [6].

In the conventional cascade control strategy, the *IBC* is controlled using an entirely *proportional–integral* controllers approach [7]. In the next subsections, improved techniques for the outer voltage control loop will be discussed.

3.1 Outer Voltage Control Loop

- *Fixed gains (standard) PI controller*

  The outer control loop is used for regulating the output voltage v_o to its given reference V_{ref}. The designed control law using constant gains *proportional–integral* (*PI*) controller is given as follows [2]:

$$
\begin{cases}
u = k_p \varepsilon_o(t) + k_i \int \varepsilon_o(t)\mathrm{d}t \\
\varepsilon_o(t) = V_{\text{ref}}(t) - v_o(t)
\end{cases}
\tag{5}
$$

  where k_p and k_i are the *PI* control parameters.

Fig. 5 *Self-tuning PI controller scheme*
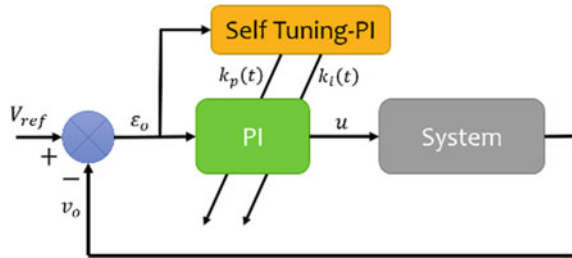
- **Self-tuning PI controller**

  Self-tuning PI controller (Fig. 5) is a technique used to vary the *PI* controller gains over a predetermined range for different operating conditions [8]. In a *PI* controller, the proportional gain k_p impacts the overshoots and the rise time response, and the integral gain k_i impacts the steady-state error. The *self-tuning PI* controller form is given as follows:

$$\begin{cases} u = k_p(t)\varepsilon_o(t) + k_i(t) \int \varepsilon_o(t)\mathrm{d}t \\ \varepsilon_o(t) = V_{\mathrm{ref}}(t) - v_o(t) \\ k_p(t) = k_{p(\max)} - \left(k_{p(\max)} - k_{p(\min)}\right)\mathrm{e}^{-[ke(t)]} \\ k_i(t) = k_{i(\max)}\mathrm{e}^{-[ke(t)]} \end{cases} \quad (6)$$

  The parameter k indicates the rate of $k_p(t)$ variation between the maximum and the minimum values.

- **Adaptive PI controller using Fuzzy Logic Technique**

  Fuzzy logic technique can be also employed for generating the *PI* controller parameters (Fig. 8). k_i is the parameter influencing the controller performance [2]. In case of step load disturbance, increasing the k_i value would reduce the voltage drop and accelerate the voltage recovery; however, in case of step voltage reference, the increase of k_i could introduce an overshoot in step response. In the applied adaptive *PI* controller, the *fuzzy logic* controller receives as input the error between the reference and the measured output voltage ($\varepsilon_o(t)$) and provides as output the change in the integral gain ($\Delta k_i(t)$). The adaptive *PI* controller expression is given as follows:

$$\begin{cases} u = k_p\varepsilon_o(t) + k_i(t) \int \varepsilon_o(t)\mathrm{d}t \\ \varepsilon_o(t) = V_{\mathrm{ref}}(t) - v_o(t) \\ k_i(t) = k_i^0 + \rho_i(t) \\ \rho_i(t) = \rho_i(t-1) + \Delta k_i(t) \end{cases} \quad (7)$$

  The parameter k_i^0 is a time-invariant constant while the *PI* controller is working, and $\rho_i(t)$ is time varying and updating parameter in real time, in other words $\rho_i(t)$
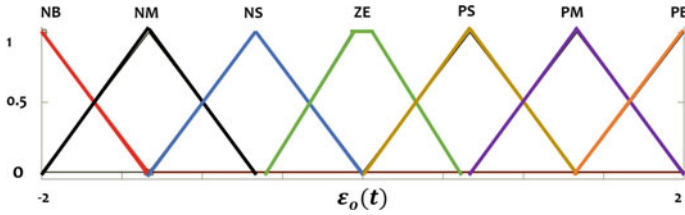
Fig. 6 Membership function for $\varepsilon_o(t)$

represents the adaptation for of $k_i(t)$. The fuzzy membership functions for input and output variables are shown in Figs. 6 and 7. The fuzzy rules are given in Table 1.
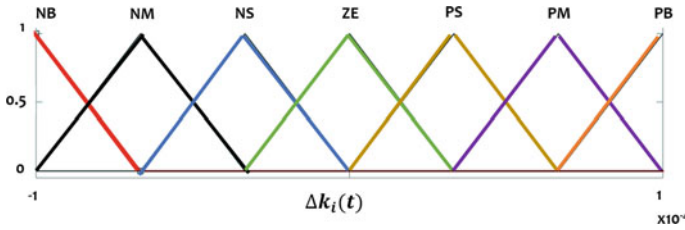
Fig. 7 Membership function for $\Delta k_i(t)$

Fig. 8 Adaptive *PI* controller scheme using *fuzzy logic* technique
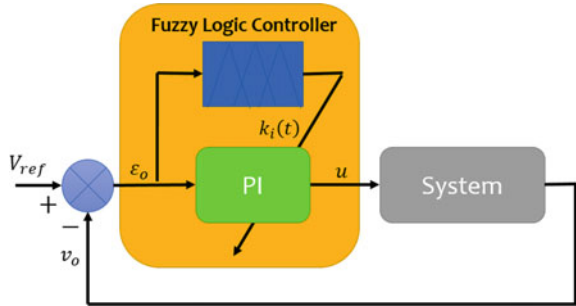
Table 1 *Fuzzy logic* control rules

$\varepsilon_o(t)$	NB	NM	NS	ZE	PS	PM	PB
$\Delta k_i(t)$	NB	NM	NS	ZE	PS	PM	PB

3.2   Inner Current Control Loop

The current control is designed using *super-twisting sliding mode* control concept [6, 9], and the inner-loop controllers form are given as follows:

$$
\begin{cases}
\qquad s_1 = I_{Lref} - i_{L1} \\
\qquad s_2 = I_{Lref} - i_{L2} \\
d_1 = \lambda |s_1|^{\frac{1}{2}} \text{sign}(s_1) + \alpha \int \text{sign}(s_1) \mathrm{d}t \\
d_2 = \lambda |s_2|^{\frac{1}{2}} \text{sign}(s_2) + \alpha \int \text{sign}(s_2) \mathrm{d}t
\end{cases}
\tag{8}
$$

The controllers' expression (8) is considered as a nonlinear *PI* controller form. s_1 and s_2 are the sliding surfaces for the inner-loop controllers, and λ and α are the super-twisting controller parameters.

In order to reduce the chattering effect, the sign function has been replaced by the hyperbolic tangent function tanh. The inner-loop control low becomes as follows:

$$
\begin{cases}
\qquad s_k = I_{Lref} - i_{Lk}, \quad k = 1, 2 \\
d_k = \lambda |s_k|^{\frac{1}{2}} \tanh(s_k) + \alpha \int \tanh(s_k) \mathrm{d}t \\
\qquad \tanh(s_k) = \frac{\sinh(s_k)}{\cosh(s_k)} = \frac{e^{s_k} - e^{-s_k}}{e^{s_k} + e^{-s_k}}
\end{cases}
\tag{9}
$$

The stability analysis as well as the asymptotical convergence of this controller can be found in [6] and [9].

4   Results and Discussion

In this part, we provide the comparison results between the fixed gains *PI* controller, the *self-tuning PI* controller and the *adaptive PI controller* using *fuzzy logic technique*. The *IBC* and the control strategy have been simulated in *MATLAB/Simulink* using *SimPowerSystem* library. The circuit interleaved boost converter parameters are $L = 400\ \mu H$, $r_L = 0.43\ \Omega$, $C = 1000\ \mu F$, $r_c = 0.04\ \Omega$, and the switching frequency is equal to 25 kHz [2]. The reference output voltage is chosen to vary from 48 V between $t = 0$ and $t = 0.5$ s to 56 V between $t = 0$ and $t = 0.5$ s and finally to 66 V between $t = 1$ and $t = 1.5$ s.

Figures 9, 10, 11, 12 and 13 show the variation of the desired output voltage and the output voltage response using the fixegains and the improved *PI* controllers. The obtained results at t $= 0$, 0.5 and 1 s correspond to the output voltage response under step variation of the desired output voltage, and at $t = 1.25$ s, the obtained results correspond to the output voltage response under step variation of the load disturbance.

During all simulation time, the input voltage was fixed at 18 V. Between $t = 0$ and 1.25 s the load resistance value is equal to 100 Ω. From the obtained results, it
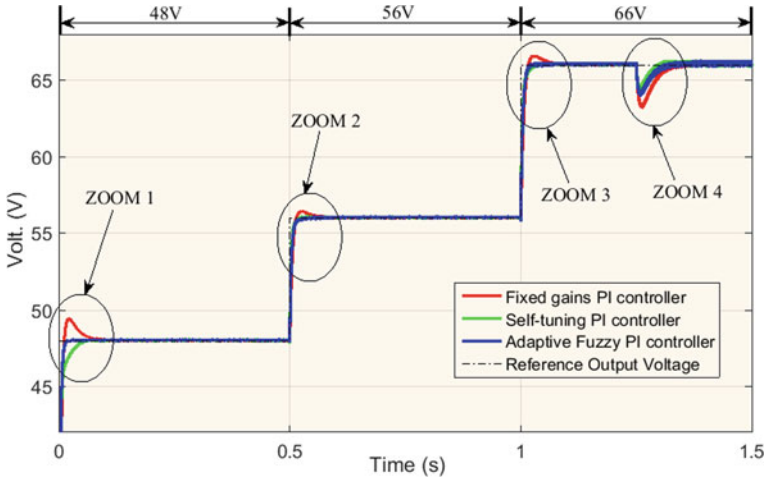
Fig. 9 Results for the output voltage response

Fig. 10 Zoom 1
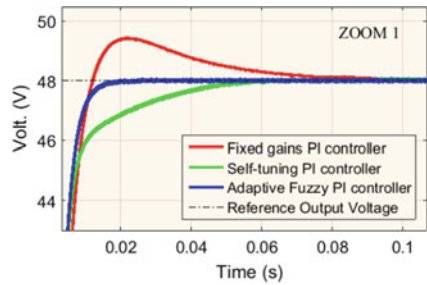
Fig. 11 Zoom 2
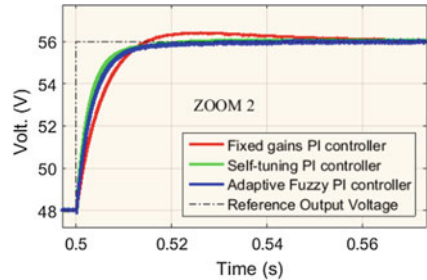
can be seen that the improved *PI* techniques are able to reduce the voltage overshoot observed when using fixed gains *PI* controller and also ensuring fast convergence response compared to the conventional technique. At $t = 1.25$ s, a step resistance load disturbance is applied, and the resistance value has been changed from 100 to 50 Ω. The obtained results show superior performance of the improved *PI* controller

Fig. 12 Zoom 3
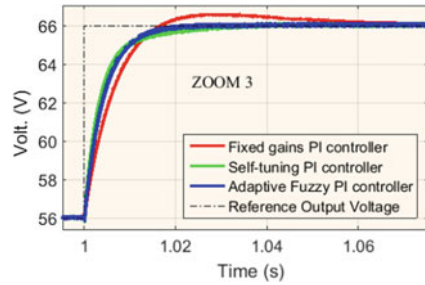
Fig. 13 Zoom 4
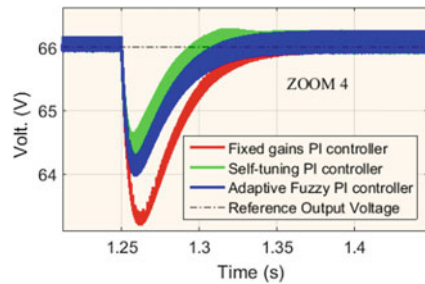
techniques with a good acceleration voltage recovery compared to the fixed gains *PI* controller.

5 Conclusion

This paper presents an analysis of *PI* controller improvements applied on an interleaved boost converter. The obtained results show that the *self-tuning PI* controller and the *adaptive PI controller* using *fuzzy logic technique* provide superior performance compared to fixed gains *PI* controllers in case of reference voltage variation and load disturbance. The studied improved techniques have also the advantage of not requiring any extra sensors and avoiding the unnecessary complexity.

References

1. Hammoudi MY et al. (2020) Practical implementation of H-infinity control for fuel cell-interleaved boost converter. Int J Model Simul 40(1):44–61
2. Zhuo S et al. (2020) Extended state observer-based control of DC–DC converters for fuel cell application. IEEE Trans Power Electron 35(9):9923–9932
3. Kabalo M et al. (2011) Modeling and control of 4-phase floating interleaving boost converter. In: IECON 2011–37th Annual conference of the IEEE industrial electronics society, IEEE

4. Guo L, John YH, Nelms RM (2009) Evaluation of DSP-based PID and fuzzy controllers for DC–DC converters. IEEE Trans Indus Electron 56(6):2237–2248
5. Arango E et al. (2007) Fuel cell power output using a LQR controlled AIDB converter. In: 2007 International conference on clean electrical power, IEEE
6. Zhuo S et al. (2021) Robust adaptive control of interleaved boost converter for fuel cell application. IEEE Trans Indus Appl 57(6):6603–6610
7. Somkun S, Sirisamphanwong C, Sukchai S (2015) A DSP-based interleaved boost DC–DC converter for fuel cell applications. Int J Hydrogen Energy 40(19):6391–6404
8. Zaky MS (2015) A self-tuning PI controller for the speed control of electrical motor drives. Electr Power Syst Res 119:293–303
9. Huangfu Y et al. (2019) An advanced robust noise suppression control of bidirectional DC–DC converter for fuel cell electric vehicle. IEEE Trans Transp Electrification 5(4):1268–1278

IoT and Machine Learning Methods

Synergy of Sentinel-1 and Sentinel-2 Satellites for Surface Soil Moisture Retrieval Over Wheat Crops in Semi-arid Areas

Nadia Ouaadi, Jamal Ezzahar, Lionel Jarlan, Saïd Khabba, and Pierre Luis Frison

Abstract Surface soil moisture is a key parameter of crop monitoring, water stress detection and irrigation management, particularly in the mediterranean region where the water resources are very limited. The aim of the study is the synergy of radar Sentinel-1 and optical Sentinel-2 data for surface soil moisture (SSM) retrieval over wheat crops. The backscattering coefficient derived from Sentinel-1, is modeled using the Water Cloud Model (WCM) combined with Oh model. The normalized difference vegetation index (NDVI) computed from Sentinel-2 is used as descriptor of vegetation in the WCM. The combined model is calibrated and validated using Sentiel-1/2 data and in situ measurement collected from two irrigated wheat fields located in the Haouz plain in the center of Morocco. The calibration is done at VV and VH polarizations and at 35.2° and 45.6° of incidence angles. Hereafter, an inversion approach is developed basing on the combined model for surface soil moisture retrieval. Results showed that SSM is retrieved with significant statistical metrics at VV polarization with $R = 0.65$, RMSE = 0.08 m^3/m^3, and bias = − 0.01 m^3/m^3 at 35.2° of incidence angle and $R = 0.57$, RMSE = 0.09 m^3/m^3 and bias = 0.01 m^3/m^3 at 45.6°.

N. Ouaadi · S. Khabba
Faculty of Science Semlalia, LMFE, Cadi Ayyad University, Marrakech, Morocco
e-mail: khabba@uca.ac.ma

N. Ouaadi · L. Jarlan
CESBIO, University of Toulouse, Toulouse, France
e-mail: lionel.jarlan@ird.fr

J. Ezzahar (✉)
MISCOM, ENSA, Cadi Ayyad University, Safi, Morocco
e-mail: j.ezzahar@uca.ma

J. Ezzahar · S. Khabba
CRSA, Mohammed VI Polytechnic University UM6P, Benguerir, Morocco

P. L. Frison
LaSTIG//MATWAS, University of Paris-Est, Champs Sur Marne, Paris, France
e-mail: pierre-louis.frison@u-pem.fr

1   Introduction

The south-mediterranean regions are threatened by increasing water scarcity. In fact,
the countries of these regions are already classified in shortage or total water shortage.
Climate change and drought intensify the problem. On the other hand, the demand
for water is increasing for an ever-growing population and therefore the quantities
of agricultural irrigation demanded are increasing. Agriculture deserves exceptional
attention in this context as it is the main consumer of water in semi-arid regions (up
to 90% in Morocco for instance) and optimization of water use therefore becomes a
crucial aspect of research. Special attention has been given to wheat in this study as
it represents the most cultivate cereal in the world, and particularly in Morocco when
it presents 75% from 59% of the useful agricultural area reserved for cereals. One
of the key parameters in the water-optimization process is the surface soil moisture
(SSM). Active microwave remote sensing is widely used for the estimation of this
parameter. The sensitivity of the microwave backscattering coefficient to SSM is
heavily documented in the literature for bare or covered soils [1]. Over bare soils,
radar signal is mainly affected by SSM and surface roughness. Numerous electromag-
netic backscatter models are used for the analysis and understanding of radar signal
sensitivity to soil parameters [2, 3] and thereafter for SSM retrieval using different
approaches [4]. Over a vegetation canopy, the estimation of SSM is more compli-
cated because of the interaction of the microwaves with the canopy components,
which makes the backscatter signal more sophisticated and ambiguous to understand.
However, several approaches and models are developed based on the resolution of
the radiative transfer equation [5]. The most common is the semi-empirical Water
Cloud Model (WCM) [6]. The vegetation is described in the WCM by one or two
vegetation parameters. Descriptors such as vegetation water content [7], biomass
[8], Leaf Area Index (LAI) [9], and normalized difference vegetation index NDVI
[3], are used. These parameters are directly computed from in situ measurements
or derived from optical *satellite imagery*. Coupling to a bare soil model, WCM is
widely used in SSM and biophysical parameters retrieval especially for crops such
as wheat [3, 8]. With the launch of the new *C*-band radar satellite Sentinel-1 and
afterwards, the optical Sentinel-2, open access products are provided for the globe
with high spatial and temporal resolutions (5–6 days at 10 m spatial resolution).
This availability allows the coupling of synthetic aperture radar (SAR) and optical
data for soil moisture mapping at the field scale every 6 days. Within this context,
the objective of this work is to develop an approach for SSM retrieval based on the
synergy of Sentinel-1 and Sentinel-2 data. The WCM coupled with Oh model [10]
is used and the approach is tested over a study area located in the center of Morocco.
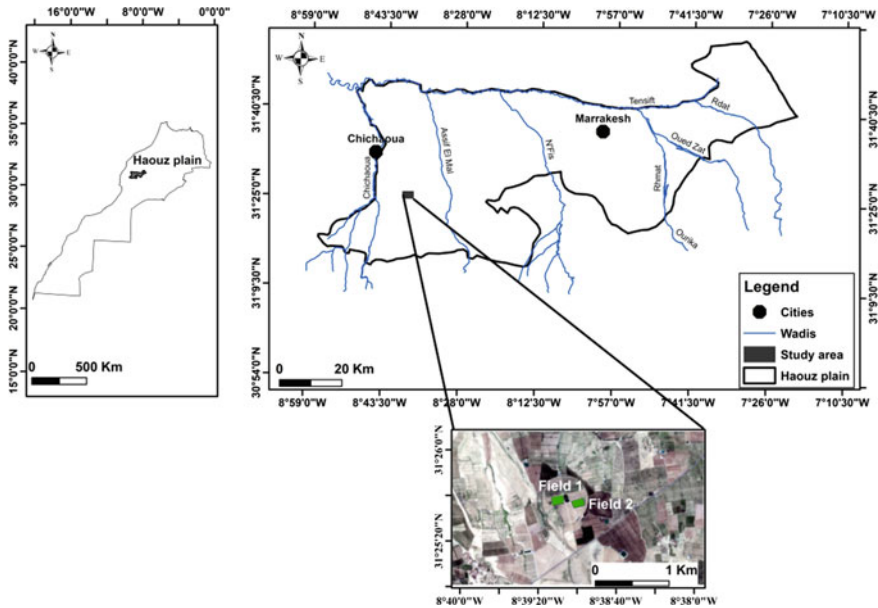
Fig. 1 Location of the study sites in the Haouz plain in the center of Morocco

2 Study Area

The study site is located in the province of Chichaoua at 65 km west of Marrakech city, in the Haouz plain, center of Morocco (Fig. 1). The region is characterized by a semi-arid Mediterranean climate with a dry summer season. Two fields of about 1.5 ha each (Field 1 and Field 2 in Fig. 1), are selected within a private farm and monitored during two crop seasons 2016–2017 and 2017–2018. The fields are sown with winter wheat and irrigated using the drip technique. The sowing is generally at the end of November (25–27) and the harvest occurs from mid-May to mid-June.

3 Methodology

Field measurements

In situ measurements of SSM and surface roughness are collected over the two irrigated wheat plots. SSM is automatically measured every 30 min using TDR sensors buried at 5 cm of depth. The soil roughness is measured once a week during the early stages of wheat development (from the sowing until the wheat cover the soil) using a pin profiler of 1 m in length. The root mean square surface height (Hrms) is computed by averaging six teen measurements carried out every week. The SSM

values are ranging between 0.04 and 0.35 m^3/m^3 while the Hrms values are between 0.7 and 1.5 cm.

Satellite measurements

Sentinel-1 is a *C*-band synthetic aperture radar mapping the world in 175 orbits. The main operational mode (IW) provides Ground Range Detected (GRDH) products with a revisit time of six days. In this mode, the sensors provide acquisitions in VV and VH polarizations. The backscattering coefficient is extracted by processing GRDH products using the Orfeo Tool Box command (OTB). Over the study area, two incidence angles are available (35.2° and 45.6°) thanks to the pass of two orbits.

Sentinel-2 is an optical satellite provides images every 5 days at a resolution ranging between 10 and 60 m according to the spectral band. Products corrected from atmospheric effects are downloaded from the Theia site web (https://theia. cnes.frhttps://theia.cnes.fr/). The Normalized Difference Vegetation Index (NDVI) is computed from individual bands 4 and 8 and averaged over each field.

Radar Signal Modeling

The backscattering coefficient is modeled using WCM combined with Oh model. NDVI is used as descriptor of wheat in the coupled model. The total backscattering coefficient $\left(\sigma^0_{pq,\,canopy}\right)$ is given as the incoherent sum of two contributions:

$$\sigma^0_{pq,\,canopy} = \sigma^0_{pq,\,vegetation} + L^2_{pq}\sigma^0_{pq,\,soil} \tag{1}$$

where pq is the polarization mode (*V* or *H*). $\sigma^0_{pq,\,vegetation}$ denotes the contribution from the vegetation computed using the WCM:

$$\sigma^0_{pq,\,vegetation} = A_{pq}NDVI\cos\theta\left(1 - L^2\right) \tag{2}$$

$$L^2_{pq} = e^{-2B_{pq}NDVI\sec\theta} \tag{3}$$

L^2 is two-way transmissivity factor of the canopy. θ is the incidence angle and A_{pq} and B_{pq} are the model's coefficients that depend on the canopy type, sensor's frequency and incidence angle.

$\sigma^0_{pq,\,soil}$ is underlying bare soil backscatter calculated using Oh model as following:

$$\sigma^0_{VV,\,soil} = g * (\cos\theta)^3 * (\Gamma_V + \Gamma_H)\big/\sqrt{p} \tag{4}$$

$$\sigma^0_{VH,\,soil} = q * g * (\cos\theta)^3 * (\Gamma_V + \Gamma_H)\big/\sqrt{p} \tag{5}$$

$$\text{Where}: q = 0.23 * \left(1 - e^{(-k*Hrms)}\right) * \sqrt{\Gamma_0} \tag{6}$$

$$\sqrt{p} = 1 - e^{(-k*\text{Hrms})} * \left(\frac{2\theta}{\pi}\right)^{\frac{1}{3}*\Gamma_0} \tag{7}$$

$$g = 0.7 * \left(1 - e^{\left(-0.65*(k*\text{Hrms})^{1.8}\right)}\right) \tag{8}$$

Γ_V, Γ_H, and Γ_0 denote the Fresnel coefficients given by the following expressions:

$$\Gamma_V = \frac{\varepsilon_r \cos\theta - \sqrt{\varepsilon_r - \sin^2\theta}}{\left(\varepsilon_r \cos\theta + \sqrt{\varepsilon_r - \sin^2\theta}\right)^2}, \ \Gamma_H = \frac{\cos\theta - \sqrt{\varepsilon_r - \sin^2\theta}}{\left(\cos\theta + \sqrt{\varepsilon_r - \sin^2\theta}\right)^2} \text{ and}$$

$$\Gamma_0 = \left|\frac{1 - \sqrt{\varepsilon_r}}{1 + \sqrt{\varepsilon_r}}\right|^2 \tag{9}$$

where k is the number of wave and ε_r is the relative dielectric constant of the soil surface computed following Hallikainen et al. [11].

The coupled WCM and Oh model is calibrated and validated using the data of both fields during 2016–2017 and 2017–2018 season, respectively. The calibrated model is then implemented in an inversion procedure for SSM estimation and the inverted SSM is compared to the field measurements.

4 Results and Discussion

4.1 Backscatter Time Series

Figure 2 illustrates the temporal evolution of the simulated backscattering coefficient using the coupled WCM and Oh model over Field 1 at 45.6° of incidence angle and for both agricultural seasons (2016–2017 and 2017–2018). The contributions from the vegetation and from the attenuated soil are also displayed. Sentinel-1 data are superimposed on the simulations for both VV and VH polarizations.

Results of the simulation show that the modeled signal is in a good agreement with the measured radar backscatter (from Sentinel-1) until the end of March. After this date, the simulated signal is observed to decrease while the observed backscatter is increasing again. This mechanism is obviously clearer on VH than on VV because of the high sensitivity of VH to vegetation canopies. The end of March is corresponding to the heading phase. During this period, the canopy structure changes due to the thick and wet head layer formed on the top of the wheat field. This behavior has been marked by some studies before. For instance, [12] reported that minimum backscatter is reached at the heading stage and the signal increased again as the crops continued to mature. The contribution of the attenuated soil dominates the signal at the beginning of the season when the soil is not completely covered. With wheat development, the
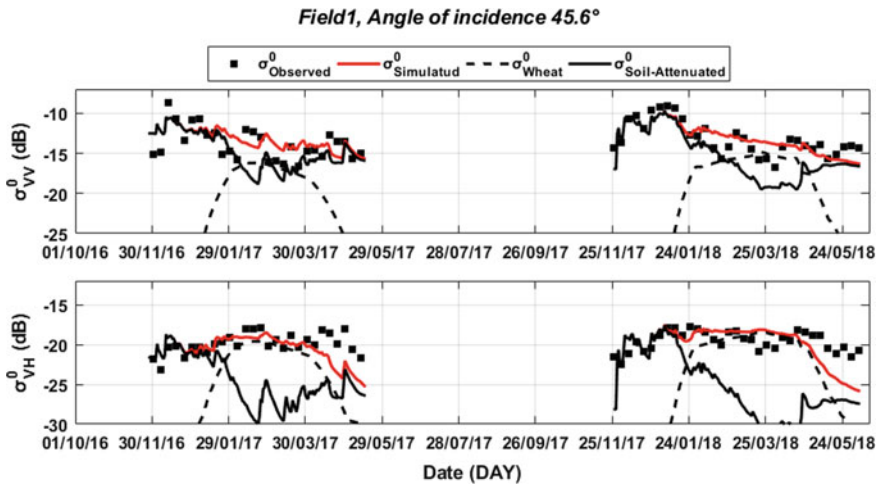
Fig. 2 Time series of the different contributions simulated using the WCM fitting the Sentinel-1 observations at VV and VH polarizations over Field 1 for both seasons (2016–2017 and 2017–2018) at 54.6° of incidence angle

contribution of vegetation volume increases to dominate the backscatter mechanism from February 1 until the end of the season when vegetation begins to dry out with the onset of senescence. Attenuated soil is important at VV for the whole season whereas at VH it becomes low when the wheat is well developed. This is an expected behavior because of the vertical structure of the stems. Similar results are obtained at 35.2° of incidence angle (not shown) with higher values at 35.2° than at 45.6° for both polarizations.

4.2 Surface Soil Moisture Retrieval

The results of implementing the already calibrated model in an inversion procedure for SSM retrieval are shown in Fig. 3. The figure illustrates the estimated SSM compared with in situ measurement taken over Field 1 at 35.2° and for both polarizations. The best results are evidently obtained at 35.2° of incidence angle at VV polarization. At this angle of incidence, SSM is retrieved with acceptable accuracy at VV ($R = 0.65$, RMSE = 0.08 m^3/m^3 and bias $= -0.01$ m^3/m^3). The results at VH are poorer than at VV because of the higher contribution of vegetation in this configuration ($R = 0.53$, RMSE = 0.09 m^3/m^3 and bias $= -0.04$ m^3/m^3). For 45.6° of incidence angle, the SSM is poorly estimated (especially at VH) compared to 35.2° due to the effect of vegetation on the C-band signal at higher incidence angles. At VV, the obtained statistical metrics are $R = 0.65$, RMSE = 0.08 m^3/m^3, and bias $= 0.01$ m^3/m^3. Overall, the obtained results are in good accordance with previous studies used the same approach over the wheat [9, 13].
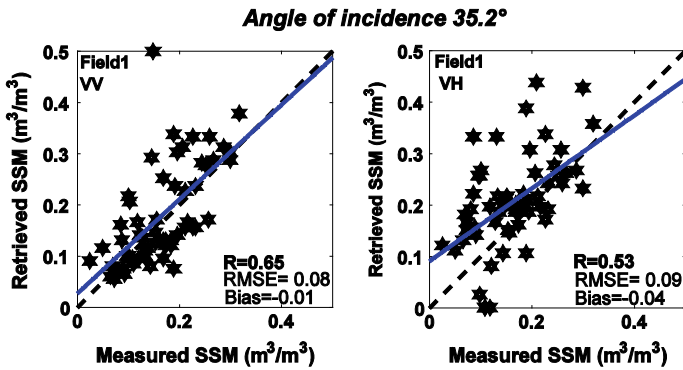
Fig. 3 Retrieved surface soil moisture (SSM) versus in situ measurements at VV and VH polarizations at 35.2° of incidence angle for the study site: Field 1 using the NDVI as descriptor of vegetation in the WCM

5 Conclusion

The backscattering coefficient is modeled using the Water Cloud Model and Oh model. The combined model is calibrated and validated over irrigated wheat crops based on Sentinel-1 backscattering coefficient. The NDVI derived from Sentinel-2 is used as descriptor of the wheat in the model. The time series of the modeled total backscatter, vegetation and attenuated soil contributions are analyzed together with Sentinel-1 data. The results of this analysis show that the model can acceptably simulate the *C*-band radar backscatter until the heading stage when a noticeable disagreement is observed between the simulations and the observations: the cycle related to the heads layer. Additionally, an inversion approach of the model is implemented for SSM retrieval using data from Sentinel-1 (radar) and Sentinel-2 (optics). The proposed approach presents the advantage of SSM estimation over wheat field using spatial data only with high temporal and spatial resolution thanks to ESA's satellites Sentinel-1/2. Although the obtained results are sound, they can be improved if the model can reproduce the cycle after the heading. For this reason, further works on wheat backscatter modeling are needed, in particular, the head's layer needs to be added to the various contributions in the modeling of the backscatter from wheat crops.

References

1. Ezzahar J et al (2020) Evaluation of backscattering models and support vector machine for the retrieval of bare soil moisture from Sentinel-1 data. Remote Sens 12(1):72
2. Fung AK, Zongqian L, Chen KS (1992) Backscattering from a randomly rough dielectric surface. IEEE Trans Geosci Remote Sens 30(2):356–369
3. Oh Y, Sarabandi K, Ulaby FT (1992) An empirical model and an inversion technique for radar scattering from bare soil surfaces. IEEE Trans Geosci Remote Sens 30(2):370–381
4. Amazirh A et al (2018) surface soil moisture at high spatio-temporal resolution from a synergy between Sentinel-1 radar and landsat thermal data: a study case over bare soil. Remote Sens Environ 211:321–337
5. Karam MA, Fung AK, Lang RH, Chauhan NS (1992) Microwave scattering model for layered vegetation. IEEE Trans Geosci Remote Sens 30(4):767–784
6. Attema EPW, Ulaby FT (1978) Vegetation modeled as a water cloud. Radio Sci 13(2):357–364
7. Hosseini M, McNairn H (2017) Using multi-polarization C- and L-band synthetic aperture radar to estimate biomass and soil moisture of wheat fields. Int J Appl Earth Obs Geoinf 58:50–64
8. Bai X et al (2017) First assessment of Sentinel-1A data for surface soil moisture estimations using a coupled water cloud model and advanced integral equation model over the Tibetan Plateau. Remote Sens 9(7):1–20
9. El Hajj M et al (2016) Soil moisture retrieval over irrigated grassland using X-band SAR data. Remote Sens Environ 176:202–218
10. Oh Y, Sarabandi K, Ulaby FT An empirical model and an inversion technique for radar scattering from bare soil surfaces. IEEE Trans Geosci Remote Sens 30:370–381. https://doi.org/10.1109/36.134086
11. Hallikainen MT et al. (1985) Microwave dielectric behavior of wet soil-part I: empirical models and experimental observations. IEEE Trans Geosci Remote Sens GE-23(1):25–34
12. Picard G, Le Toan T, Mattia F (2003) Understanding *C*-band radar backscatter from wheat canopy using a multiple-scattering coherent model. IEEE Trans Geosci Remote Sens 41(7):1583–1591
13. Ouaadi N et al. (2020) Monitoring of wheat crops using the backscattering coefficient and the interferometric coherence derived from sentinel-1 in semi-arid areas. Remote Sens Environ 251:1–20

Smart Greenhouse with Plant Diseases Classification Using Transfer Learning and Deep CNNs

Adel Mellit and Hajji Bekkay

Abstract In this paper, we apply transfer learning (TL) method with three deep convolutional neural networks (DCNNs) for plant diseases classification. First, a smart greenhouse designed at the RELab, University of Jijel (Algeria), is described. Subsequently, a dataset (images of tomato leaf diseases) has been used to train and test the three classifiers (ResNet-50, VGG-16 and AlexNet). Finally, the three classifiers have been evaluated, based on the confusion matrix, and the best accuracy was obtained by the VGG-16 classifier with a classification rate of 95.45%.

Keywords Smart greenhouse · Plant diseases · Classification · Transfer learning · Deep learning · IoT

1 Introduction

It is well known that plant diseases can seriously affect the crop as well as the quality of agriculture. Thus, the process of producing high-quality crops is very significant to meet the growing demand of food products around the world. Designing an automatic and effective method to detect and classify plant diseases is a crucial task [1, 2].

Recently, many researchers are attracted to the application of artificial intelligence (AI) techniques, including deep learning-based algorithms in this field. For instance, in [3], the author designed a deep convolutional neural networks DCNN-based model for tomatoes disease classification. The classification rate was 98.49%. Various deep CNN configurations were applied to classify disease of tomato leaf and rust corn [4]. The authors used the Internet of things (IoT) technique to collect and notifying users. In [5], the authors used the EfficientNet model for plant disease classification and the

A. Mellit (✉)
Faculty of Science and Technology, Department of Electronics, University of Jijel, BP.98, 18000 Jijel, Algeria
e-mail: adelmellit2013@gmail.com

H. Bekkay
Energy, Embedded Systems and Information Processing Laboratory, National School of Applied Sciences of Oujda, Mohammed First University, BP665, 60000 Oujda, Morocco

accuracy was 97%. A new approach was developed for automatic classification and detection of plant diseases from leaf images; the precision ranges between 91 and 98% plant diseases from leaf images [6]. A good extended review on the application of deep CNNs is given in [7], in which various structures were evaluated.

In our previous study [2], we have designed a smart greenhouse prototype using the Internet of things (IoT) technique, and only one type of deep CNN is considered (the accuracy was 88%); thus, the main objective of this work is to make a comprehensive comparative study between three DCNN configurations using transfer learning (TF) method. This will help users to select the appropriate method in terms of accuracy and simplicity.

This paper is organized as follows: Sect. 2 provides materials and methods; it includes the greenhouse presentation, the dataset collected and used to develop the classifier, the used language, and the evaluated metrics. A brief introduction to deep CNN configurations is given in the same section. Results and discussion are provided in Sect. 3, in this section, we investigate and evaluate the three classifiers.

2 Materials and Methods

2.1 Greenhouse Description and Database

Figure 1 shows the designed smart greenhouse, it consists mainly of an electronic board and other sensors and actuators; more details can be found in [2]. A low-cost Wi-Fi camera was used to capture the tomato plant images and sending them via firebase to the cloud. Then, stored images will be used for plant diseases classification.

Figure 2 depicts an example of the dataset [8] used to train and evaluate the three deep CNNs.

Image data augmentation technique is employed to expand the training dataset, in order to improve the performance and ability of the DCNN-based algorithms to

Fig. 1 Prototype of the designed smart greenhouse [2]

Fig. 2 Example of collected images (dataset), numbers 0, 1, 2, 3, 4, 5, and 6, mean the classes: Bacterial spot (D1), black leaf mold (D2), gray leaf spot (D3), healthy (H), late blight (D4), and powdery mildew (D5), respectively

generalize. Figure 3 shows an example of the transformed images based on random horizontal flip technique [9].

2.2 Deep CNN Configurations

This subsection describes briefly the deep CNN configurations. A basic block diagram of the classifier based on a deep CNN is shown in Fig. 4. It consists basically of various layers: convolution layers, pooling layer, dropout layer, flatten layer, fully connected, and softmax layer. The compared DCNN architectures are given below:

- *AlexNet*: It consists of five convolutional layers and three fully connected layers [10].
- *VGG*: There are various type of VGGNet (VGG-16, VGG19, etc.) which differ only in the total number of layers in the network [11].
- *ResNet*: There are multiple versions of ResNet architectures. It consists of various layers, and the most commonly used are ResNet-50 and ResNet101 [12].
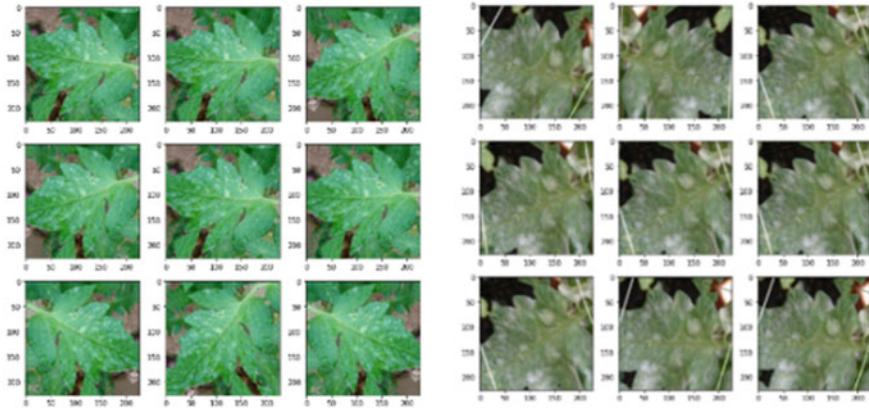
Fig. 3 Augmented images
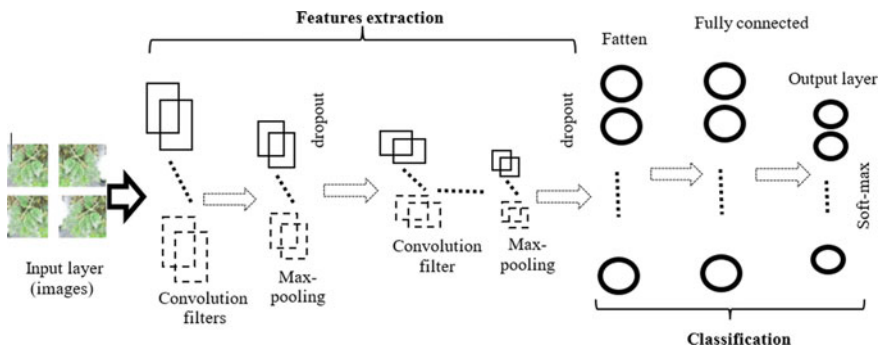
Fig. 4 Block diagram of the basic deep CNN classifier

To develop the three classifiers, transfer learning method is used [13].

2.3 Programming Language and Tools

DCNN classifier-based models have been implemented online using Google Colab, a cloud platform that provides Jupyter netbook services. Codes are written in Python language; in addition, libraries such as TensorFlow and Keras have been employed. Figure 5 shows Google Colab interface used for running the investigated classifiers. The Google Colab can be easily utilized and provides a virtual machine with GPUs (up to 16 GB of RAM).
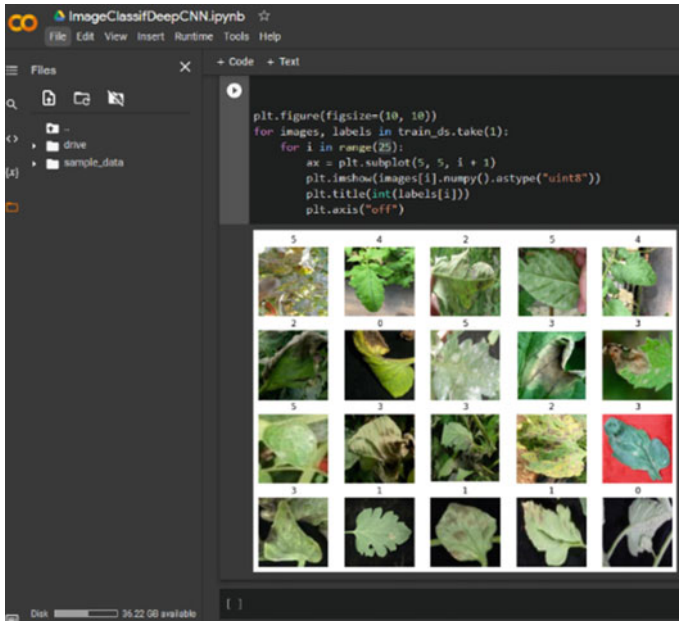
Fig. 5 Google Colab interface used for running the examined DCNN-based classifiers

2.4 Performance Evaluation Metrics

To evaluate the performance of the three investigated DCNN classifiers, the confusion matrix technique, denoted by CM, has been used [14]. It permits to calculate the precision, the accuracy, the sensitivity, and the F-score. The corresponding equations are given below:

$$\text{Accuracy} = \frac{\sum_i \text{CM}(i, i)}{\sum_i \sum_j \text{CM}(i, j)} \tag{1}$$

$$\text{Precision}_i = \frac{\text{CM}(i, i)}{\sum_j \text{CM}(j, i)} \tag{2}$$

$$\text{Recall}_i = \frac{\text{CM}(i, i)}{\sum_j \text{CM}(i, j)} \tag{3}$$

$$F1 - \text{score} = 2\frac{(\text{sensitivity} \times \text{Recall})}{(\text{sensitivity} + \text{Precision})} \tag{4}$$

3   Results and Discussion

The dataset was divided into two subsets, 80% (3120 images) for training and 20% (780 images) for testing of the three DCNNs. For example, Fig. 6 represents the loss and the accuracy plots during the training and the validation (case of the DCNN model #2). As can be seen from Fig. 6a after 75 epochs, the loss is very close to 0.001 and the accuracy is also closer to 1. From Fig. 4b, we can clearly observe that the accuracy reached 0.99 after 35 epochs, while the loss reached 0.25 after 175 epochs. Globally, both training and validation results are very acceptable. To assess the performance of the explored classifiers, we have calculated the confusion matrix for each classifier, and the results are reported in Table 1.

From Table 1, it can be seen that the accuracy ranges between 90.25 and 95.45%, the best accuracy is obtained by the VGG-16 model. It should be noted that data 308 images have been used to evaluate the classifiers accuracy.

Figure 7 displays the confusion matrix for the three DCNN classifiers. Taking the case of the CM (b) (model #2: VGG-16), the classification accuracy is 294/308 (95.45%). From the first line of the CM, the recall of D1(disease class 1) is 52/52 (1.00); from the first column, the precision is 52/55 (0.945), and then the $F1$-score can be calculated based on the value of the recall and the precision as follows $2 \times (0.954 \times 1.00)/(1.00 + 0.945) = 0.971$, and so on for other lines and columns (see Eqs. 1–4). The cases of diseases D4 and D5 the same results are obtained by the three compared classifiers.
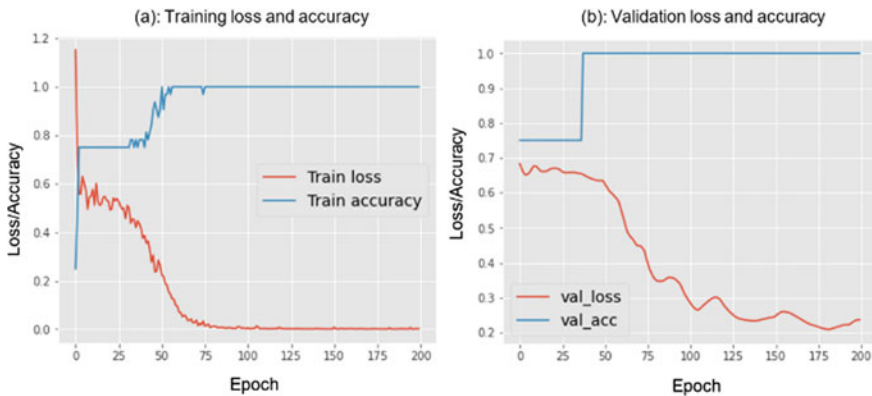
Fig. 6 Loss and accuracy evolution of the second DCNN (model #2), for disease classification of tomatoes. **a** Training and **b** validation

Table 1 Error metrics: precision, recall, *F*1-score, and accuracy for tomato diseases classification using the three DCNNs (model #1 ResNet-50, model #2 VGG-16, and model #3 AlexNet)

Classes	Precision	Recall	*F*1-score	Accuracy (%)
Model# 1 (ResNet-50)				
D1	0.91	0.92	0.91	
D2	0.93	0.88	0.90	
D3	0.89	0.98	0.93	92.53
H	0.93	1.00	0.96	
D4	1.00	0.96	0.98	
D5	1.00	0.78	0.88	
Model #2 (VGG-16)				
D1	0.95	1.00	0.97	
D2	1.00	0.93	0.96	95.45
D3	0.90	1.00	0.95	
H	0.97	1.00	0.99	
D4	1.00	0.96	0.98	
D5	1.00	0.78	0.88	
Model #3 (AlexNet)				
D1	0.88	0.87	0.87	
D2	0.89	0.83	0.86	90.25
D3	0.84	0.98	0.90	
H	0.97	0.97	0.97	
D4	1.00	0.96	0.98	
D5	1.00	0.78	0.88	
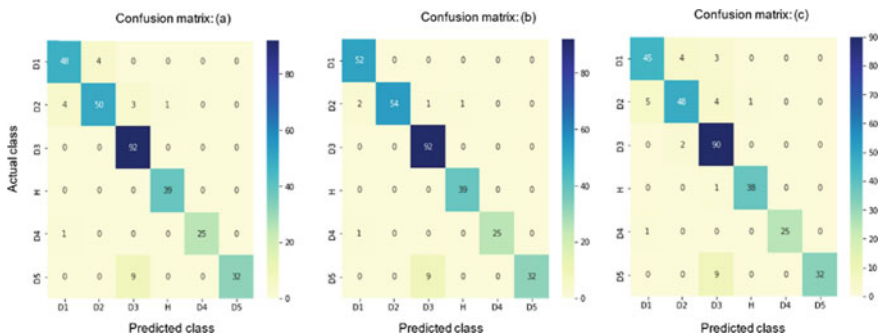
Fig. 7 Confusion matrices of the three investigated DCNN classifiers based on transfer learning approach: **a** AlexNet, **b** VGG-16, and **c** ResNet-50

4   Conclusion

In this paper, a smart greenhouse prototype is presented, while the main contribution was the examination of the well-known deep CNN classifiers based on transfer learning (AlexNet, VGG-16 and ResNet-50) for plant diseases classification. The case of the tomato plant diseases was considered in this study.

Results showed that the VGG-16 outperforms the other investigated classifiers (AlexNet and ResNet-50). Using this configuration VGG-16, the accuracy is improved by 7.45% compared with the one of 88% obtained previously in [2].

As further action, more efforts should be put on the database preparation, dataset size, images quality, and designing more advanced algorithms to improve the accuracy.

References

1. Zhang P, Yang L, Li D (2020) EfficientNet-B4-ranger: a novel method for greenhouse cucumber disease recognition under natural complex environment. Comput Electron Agric 176:105652
2. Mellit A, Benghanem M, Herrak O, Messalaoui A (2021) Design of a novel remote monitoring system for smart greenhouses using the internet of things and deep convolutional neural networks. Energies 14(16):5045
3. Trivedi NK, Gautam V, Annd A, Aljahdali HM, Villar SG, Anand D, Goyal N, Kadry S (2021) Early detection and classification of tomato leaf disease using high-performance deep neural network. Sensors 21(23):7987
4. Fatima N, Siddiqui SA, Ahmad A (2021) IoT-based smart greenhouse with disease prediction using deep learning. Int J Adv Comput Sci Appl 12:7
5. Akshai KP, Anitha J (2021) Plant disease classification using deep learning. In: 3rd International conference on signal processing and communication, pp 407–411
6. Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D (2016) Deep neural networks based recognition of plant diseases by leaf image classification. Comput Intell Neurosci 3289801
7. Li L, Zhang S, Wang B (2021) Plant disease detection and classification by deep learning—a review. IEEE Access 9:56683–56698
8. DatabaseTomate. https://data.mendeley.com/datasets/ngdgg79rzb/1. Last accessed 27 May 2021
9. Brownlee J (2019) How to configure image data augmentation in Keras. In: Deep learning for computer vision. https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks/. Last assessed 12 2019
10. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105
11. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: International conference on learning representations, San Diego, CA
12. Targ S, Almeida D, Lyman K (2016) Resnet in resnet: generalizing residual architectures. arXiv preprint arXiv:1603.08029

13. Marom ND, Rokach L, Shmilovici A (2016) November. Using the confusion matrix for improving ensemble classifiers. In: 26-th Convention of electrical and electronics engineers, Israel, pp 000555–000559
14. Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. J Big data 3(1):1–40

IoT-Based Data Acquisition and Remote Monitoring System for Large-Scale Photovoltaic Power Plants

Muhammet Şamil Kalay, Beyhan Kılıç, Adel Mellit, Bülent Oral, and Şafak Sağlam

Abstract The amount of solar capacity deployed worldwide has doubled in the past decades. The increasing use of solar energy makes photovoltaic (PV) power plants substantial. In PV power plants, reducing maintenance and operating costs positively affects efficiency. A failure in any module can reduce or interrupt the production of electrical energy, causing significant losses in both efficiency and revenue. Therefore, responding to a fault as quickly as possible in PV power plants is critical. The ability of the PV plant operator to react to potential faults is directly related to the rapid detection of faulty modules. In this paper, IoT-based data acquisition and monitoring system is designed to diagnose module failures and remotely monitor for PV power plant's performance. The current, voltage, module surface temperature, and solar radiation values are measured for each PV module. These data are transmitted wirelessly to long distances with LoRa modules. All data acquired in the central recording unit device are transferred to the Internet, enabling online access and also stored on the memory card.

Keywords Data acquisition · Monitoring · Photovoltaic systems · Fault detection and diagnosis · IoT · Data transmission

M. Ş. Kalay (✉) · B. Oral · Ş. Sağlam
Marmara University, Idealtepe Yolu. 15, 34854 Istanbul, Turkey
e-mail: mkalay@marmara.edu.tr

B. Oral
e-mail: boral@marmara.edu.tr

Ş. Sağlam
e-mail: ssaglam@marmara.edu.tr

B. Kılıç
Yıldız Technical University, Davutpasa, 34220 Istanbul, Turkey
e-mail: beykilic@yildiz.edu.tr

A. Mellit
Faculty of Sciences and Technology, University of Jijel, 18000 Jijel, Algeria
e-mail: adel_mellit@univ-jijel.dz

1 Introduction

Renewable energy systems are an alternative to fossil fuel-based energy production systems to meet the increasing energy demand and prevent environmental problems such as global warming and climate change. Global photovoltaic (PV) power capacity increased by more than 126–707 GW in 2020, accounting for 48% of renewable energy installations [1]. PV systems are installed at different scales and in different environments. PV systems are installed at different scales, from kW-sized rooftop systems to MW-sized power plants. PV plants have various malfunctions and maintenance requirements changes depending on the installation area's geographical and meteorological conditions. Manual methods are risky for the maintenance operator's health and time-consuming procedures. Wired communication in remote monitoring systems needs extra cabling and, therefore, additional cost. These reasons reveal the need for wireless remote data acquisition and monitoring systems in PV plants. Various monitoring PV systems based on Internet of Things (IoT) technique are presented in [2]. The main objective of this paper is to propose a wireless data acquisition and monitoring system to diagnose PV module failures and remotely monitor PV plant performance.

2 Method

The performance of PV system is affected by environmental variables such as solar radiation and module surface temperature. In addition, it is important to measure and monitor the current and voltage values produced by PV modules to understand the system's performance. Each PV module's surface temperature, current, voltage data, and solar radiation values are measured in this paper. Figure 1 shows a general schematic of the proposed data acquisition and monitoring system.
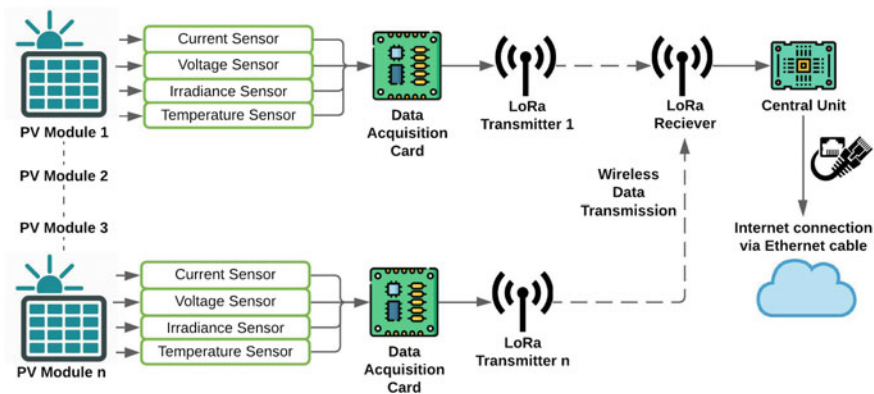
Fig. 1 General schematic of proposed data acquisition and monitoring system

Inverters are used to collect data in the majority of existing monitoring systems. This technique does not allow for understanding each module's performance data separately. The presented method provides data acquisition and monitoring at each PV for a low cost. This wireless communication technique can be used in large-scale PV systems at the module level.

2.1 Solar Radiation Measurement

Total solar radiation is defined as the total power of the sun striking a unit area (given in units of W/m^2) [3]. Pyranometer devices are generally used to measure total solar radiation. Due to the high cost of professional pyranometer devices, photodiode and reference cell methods are frequently used in low-cost systems. The presented system is intended to be low cost; therefore, solar radiation measurement is made with a BW34 model silicon photodiode instead of a pyranometer. The current produced by the photodiode is measured over the voltage drop across the resistor connected to the photodiode output. Due to the mV level change in the measured voltage value despite the large changes in the solar radiation value, the signal is amplified using an operational amplifier (LM358). Silicon photodiode is calibrated by comparing measurement results under the same conditions with a professional commercial pyranometer (KIMO CR-100).

2.2 Module Surface Temperature Measurement

The temperature change in the PV modules seriously affects the module efficiency and power output. PV modules work better at low temperatures. The expression of photocurrent (I_{ph}) as a function of radiation and temperature values is as follows:

$$I_{\mathrm{ph}} = \frac{E}{E_{\mathrm{ref}}}\big[I_{\mathrm{ph,ref}} + \mu_1(T - T_{\mathrm{ref}})\big] \tag{1}$$

I_{ref} is the current produced by a PV cell operating under standard test conditions (STC) reference conditions (provided by the manufacturer). E and T represent radiation and temperature. E_{ref} and T_{ref} represent standard irradiance values (1000 W/m^2) and standard temperature values (298.15 K), respectively. μ_1 is the temperature coefficient at the short-circuit current [4, 5].

In the developed system, the waterproof version of the DS18B20 model digital temperature sensor is used to measure the PV module temperature. The thermal conductive paste is applied to increase the measurement accuracy when mounting the sensor to the PV module. In addition, thermal insulators are used on surfaces that

do not contact the module, so the temperature sensor is less affected by the outdoor temperature during the measurement.

2.3 Voltage Measurement

Voltage divider circuits are the most commonly used method for measuring the voltage of PV modules. Low cost is one of the most important advantages of voltage divider circuits. The voltage value of the divider circuit is reduced to the determined level by using two series resistors. Resistor selection is made considering the open-circuit voltage (V_{oc}) specified in the catalog of the PV module. In this way, the module's output voltage is reduced to a certain level and measured over the resistor (4.7 and 47 kΩ used, assuming V_{oc} up to 50 V).

2.4 Current Measurement

Measurement of current is necessary for an effective PV monitoring system. The current measurement methods used may differ depending on the type of the application, the need for sensitivity, circuit size, and cost. Hall effect current sensor is chosen due to its advantages such as wide measuring range, high accuracy, and good linearity [6, 7] compared to other methods. A unidirectional, 185 mV/A sensitivity version of the ACS715 sensor model that can measure up to 20 A is selected [8].

2.5 Data Acquisition Card

Data acquisition (DAQ) card helps to convert analog signals into digital signals and interface with computers and similar environments. Arduino Nano is preferred in the presented system because of its cost-effectiveness, small size, ease of use, and accessibility. The microcontroller is chosen according to the number of analog inputs and resolution requirements. Three analog output sensors (voltage divider circuit, ACS715, radiation sensor) are connected to the analog pins of Arduino Nano, and the digital temperature sensor (DS18B20) is connected to the digital pins. The values obtained from all these sensors are interpreted in the Arduino software.

The developed data acquisition system is tested on a solar simulator. The solar panel output current is measured with an average deviation of $\pm$ 0.08 A relative to the reference ammeter. Figure 2a shows the comparative results for the current. The solar panel voltage measurement results are compared with the voltmeter, as shown in Fig. 2b. Temperature and solar radiation values are also measured and compared to those of reference devices, as shown in Fig. 2c and d. The average error in measuring
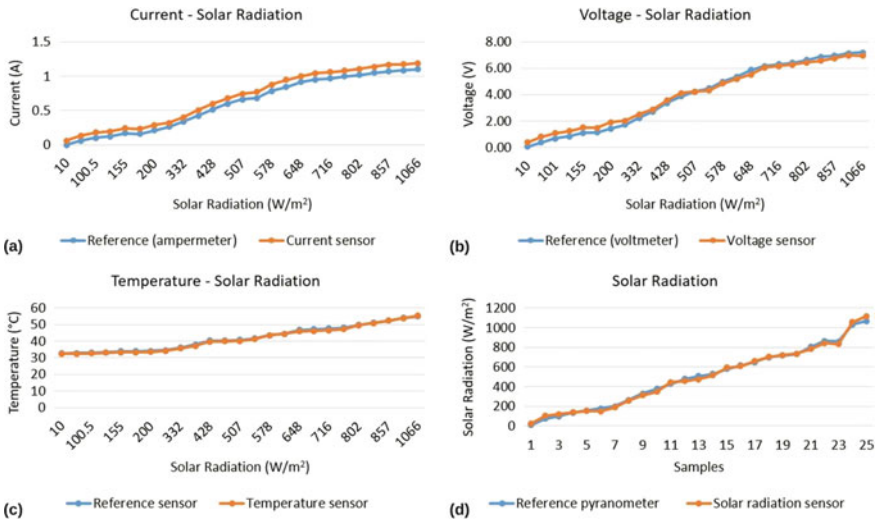
Fig. 2 Comparative measurement results; **a** current, **b** voltage, **c** temperature, and **d** solar radiation

solar radiation is 6%. Flowchart of proposed data acquisition and monitoring system is shown in Fig. 3.

2.6 Wireless Data Transmission

Wireless data transmission is one of the most important parts of this paper. DAQ cards are used for each PV module because the measurement is made at the module level. Since there are many PV modules in large-scale PV plants, wired data transmission from numerous DAQ cards to the central recording unit (CRU) causes high cost and cable complexity. Wireless data transfer aims to provide remote monitoring of PV module performances without cable clutter and additional costs. There might not be Internet infrastructure in large-scale PV plants, LoRa technology is chosen among the wireless data transmission technologies with low energy consumption, which communicates using radiofrequency. LoRa is a low-power wide-area network (LPWAN) technology that uses industrial, scientific, and medical (ISM) bands. LoRa technology has a low data rate [9]. Since the measurement data has several characters, the low data rate does not negatively affect the system. LoRa communication is based on the star topology. Star topology has less power consumption as it eliminates data transmission to intermediate devices. It is a more suitable topology for long-range communication than mesh networks [10]. The data of each PV module is stamped with the digital identity specific to the module before it is sent to the CRU (see Fig. 4). Thus, the data of many PV modules reaching the CRU does not interfere with each other. The tests have showed that the LoRa SX1278 module can transmit data up to

Fig. 3 Flowchart of proposed data acquisition and monitoring system
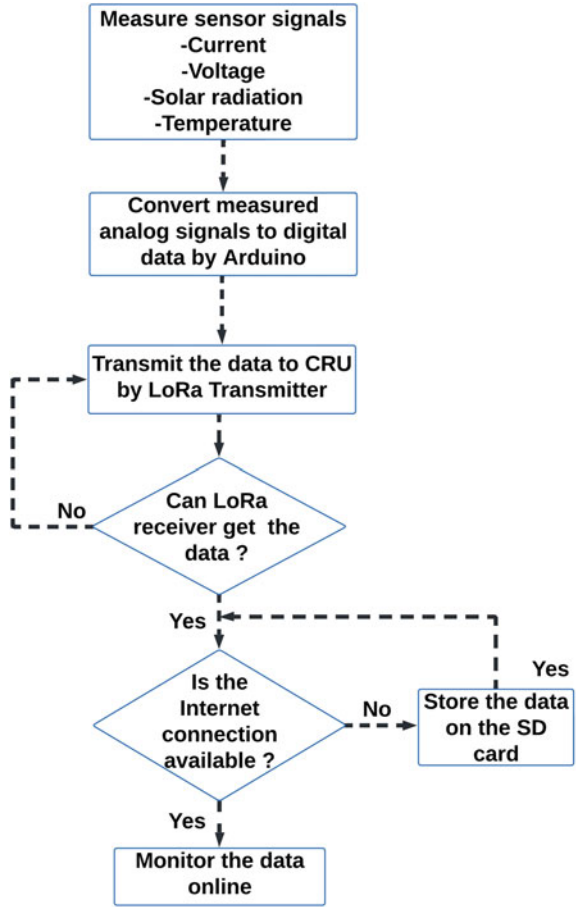
625 m without a line of sight. Assuming that the PV plant is square, and the CRU is located in the middle of this area, all measurement data in an area of 78 hec can be easily collected in a single device.

The spreading factor (SF) parameter of the LoRa module takes values between 6 and 12. A larger SF increases energy consumption, reduces the data rate, and improves the communication range. Communication is carried out at a distance of 625 m with a bandwidth of 20.8 kHz when the SF value is 11.

2.7 Data Storage and Analysis

The data of each module is transmitted to the CRU wirelessly, then to the Internet via an Ethernet connection. CRU saves the data on the SD card with date information, so

Fig. 4 Example of LoRa wireless communication architecture

Internet

Ethernet cable

625 m

883 m

625 m

75 hectares of total area

883 m

○ LoRa transmitter

● LoRa receiver (central unit)
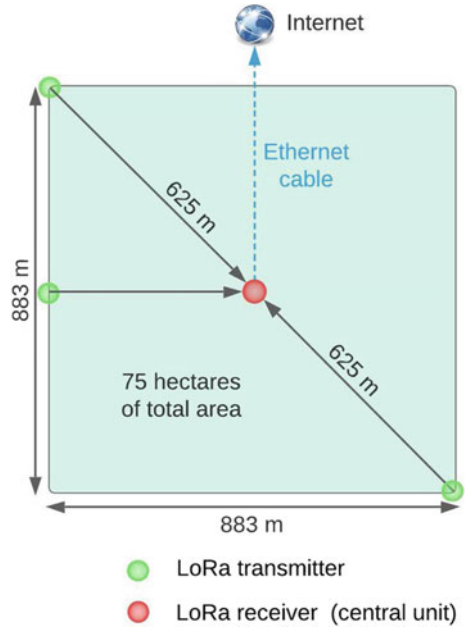
the data is not lost in case of any failure or interruption in the Internet infrastructure. Once the Internet connection is restored, this data is transferred to the Internet.

2.8 Energy Consumption and Cost of the System

Each DAQ card is powered by a rechargeable lithium-ion (Li-ion) battery. The rechargeable battery charges itself with a simple charging circuit from the PV module. Since the study aims to increase the efficiency of PV plants, components with low power consumption are selected. In Table 1, the currents drawn and the power consumed by the electronic components used for the DAQ card are given.

The LoRa module operates as a receiver; it typically draws current in the range of 10.8–12 mA while in transmitter mode 120 mA. LoRa modules on DAQ cards work in transmit mode.

The total cost of the data acquisition system per PV module is calculated as approximately $18.72. Assuming that $450\,W_p$ (watt-peak) PV modules are used, which costs about $200, the generation performance of the PV module can be monitored remotely with an additional cost of 9.36%.

Table 1 Power consumption of the components on the designed DAQ card

Component	Current drawn	Frequency of use	Energy consumed per hour	Energy consumed per day
ACS715	10 mA	Day time period (15 h)	0.05 Wh	0.75 Wh
LoRa SX1278	120 mA (transceiver mode)	10 s/min	0.10 Wh	1.50 Wh
Voltage divider	0.9 mA (in V_{oc} condition)	Continuous	0.45 Wh	0.18 Wh
Solar radiation sensor	–	–	–	–
Arduino nano	19 mA	Day time period (15 h)	0.13 Wh	2.56 Wh
Temperature sensor	1.5 mA	Day time period (15 h)	0.075 Wh	1.12 Wh
Total				6.11 Wh

3 Conclusion

This paper proposes a PV monitoring system based on IoT technique. Each PV module's current, solar radiation, and surface temperature values are measured and transferred to the CRU using LoRa technology with the digital identity information specific to each data acquisition card. Then, the data is transferred to the Internet via an Ethernet connection, and remote monitoring is provided. Assuming that the plant area is square, the developed system enables remote monitoring of each PV module's performance in an area of approximately 780,000 m^2. A negative situation that may occur in any module can be detected quickly. PV plant areas are varied and generally not in square shape. In this case, the communication architecture can be changed, and more central recording units should be used. The rechargeable battery that powers the DAQ card is charged with the electrical energy produced by the PV module on which it is located. In this way, the need for battery replacement and extra power supply is avoided. In the scenario where a 15-h sunshine duration is assumed, the designed system consumes approximately 6.11 Wh of energy. This energy consumption is commercially negligible compared to the gain from responding quickly to adverse situations.

References

1. International Renewable Energy Agency (2021). Renewable capacity statistics 2021, Abu Dhabi
2. Mellit A, Kalogirou S (2021) Artificial intelligence and internet of things to improve efficacy of diagnosis and remote sensing of solar photovoltaic systems: challenges, recommendations and future directions. Renew Sustain Energy Rev 143:110889
3. Solanki SK, Krivova NA, Haigh JD (2013) Solar irradiance variability and climate. Ann Rev Astron Astrophys 51(1):311–351
4. Xiao W, Dunford WG, Capel A (2004) A novel modeling method for photovoltaic cells. In: 35th Annual IEEE power electronics specialists conference, IEEE, Aachen, Germany
5. Tian H, Mancilla-David F, Ellis K, Muljadi E, Jenkins P (2012) A cell-to-module-to-array detailed model for photovoltaic panels. Sol Energy 86(9):2695–2706
6. Liu C, Liu JG (2014) Offset error reduction in Open Loop Hall Effect current sensors powered with single voltage source. In: International workshop on applied measurements for power systems proceedings, IEEE, Aachen, Germany
7. Xu C, Liu JG, Zhang Q, Yang Y (2015) International workshop on applied measurements for power systems, IEEE, Aachen, Germany, pp 19–24
8. Allegro MicroSystems. ACS715 Datasheet by Allegro MicroSystems, https://www.digikey.com/es/htmldatasheets/production/209355/0/0/1/acs715.html?site=US&lang=es&cur=USD. Last accessed 12 Jan 2022
9. Gupta T, Bhatia R (2020) Advanced communication technologies and intelligent systems communication technologies in smart grid at different network layers: an overview. In: International Conference on Intelligent Engineering and Management, IEEE, London, UK
10. Devalal S, Karthikeyan A (2018) LoRa technology—an overview. In: e 2nd International conference on electronics, communication and aerospace technology, IEEE, Coimbatore, India, pp 284–290

An Intrusion Detection System Using Machine Learning for Internet of Medical Things

Idriss Idrissi [ID]**, Mohammed Boukabous** [ID]**, Mounir Grari** [ID]**, Mostafa Azizi** [ID]**, and Omar Moussaoui** [ID]

Abstract The COVID-19 pandemic has impacted everyday life, the global economy, travel, and commerce. In many cases, the tight measures put in place to stop COVID-19 have caused depression and other diseases. As many medical systems over the world are unable to hospitalize all the patients, some of them may get home healthcare assistance, while the government and healthcare organizations have access to substantial sickness management data. It allows patients to routinely update their health status and have it sent to distant hospitals. In certain cases, the medical authorities may designate quarantine stations and provide supervision equipment and platforms (such as Internet of Medical Things (IoMT) devices) for performing an infection-free treatment, whereas IoMT devices often lack enough protection, making them vulnerable to many threats. In this paper, we present an intrusion detection system (IDS) for IoMTs based on the following gradient boosting machines approaches: XGBoost, LightGBM, and CatBoost. With more than 99% in many evaluation measures, these approaches had a high detection rate and could be an effective solution in preventing attacks on IoMT devices.

Keywords IoMT · IoT · IDS · HIDS · LightGBM · XGBoost · CatBoost

I. Idrissi (✉) · M. Boukabous · M. Grari · M. Azizi · O. Moussaoui
MATSI Research Lab, ESTO, Mohammed First University, Oujda, Morocco
e-mail: idrissi@ump.ac.ma

M. Boukabous
e-mail: m.boukabous@ump.ac.ma

M. Grari
e-mail: m.grari@ump.ac.ma

M. Azizi
e-mail: azizi.mos@ump.ac.ma

O. Moussaoui
e-mail: o.moussaoui@ump.ac.ma

1   Introduction

The worldwide coronavirus disease-2019 (COVID-19) pandemic is caused by the new SARS-CoV-2, also known as the coronavirus with its different variants. This disease is continually testing both the collective immunity and the global community's capacity to cope with this situation. SARS-CoV-2 has significantly hampered normal daily living, the global economy, international travel, and face-to-face activities. COVID-19 has been linked to depression in many individuals because of the severe quarantine put in place to halt its spread. Its social, political, and psychological effects were first assessed to learn more about this phenomenon. The pandemic's influence on ordinary life is portrayed, as are the role and ambitions of national governments in this unexpected situation [1]. According to WHO estimates, more than 5 million people have died because of the SARS-CoV-2. Every week, more than 50,000 people are lost. In the past year, 56 nations around the globe recorded an increase in fatalities of at least 10% [2].

On the other hand, this pandemic has demonstrated the weakness of current medical systems and urges the establishment of a medical platform that assists individuals with home health care and provides the government and healthcare institutions with extensive illness management tools. Patients may regularly consult and update their health status on the IoMT platforms and share it with hospitals. The government might supply equipment and designate quarantine stations (hotels or centralized quarantine facilities) if needed. The IoMT platform allows people to monitor their health using IoMT devices connected to the patient's body and get therapy without risking infection to others [3]. However, most of the available IoMT devices are not really designed with security handling, which makes them insecure and exposed to several threats. Unlike other businesses, healthcare systems and networks are more critical, and any security failure may actually result in loss of life.

Deep learning (a subset of machine learning [4]) has recently demonstrated successful results in a wide range of applications [5], including medical image processing [6], natural language processing [7], cybersecurity [8], and many other fields. Nevertheless, deep learning is not always the most efficient solution for tabular datasets [9], and machine learning may be better, such as gradient boosting machines (GBM) techniques like XGBoost, LightGBM, and CatBoost, which are some of the most well-known machine learning algorithms in use today [10]. Our IDS that we propose in this paper is targeting IoMT devices and is based on these GBM techniques.

The rest of the paper is organized as follows. In the second section, we present backgrounds and explore the related works. The third section exhibits the methodology used to build the models of our solution. Before concluding, the fourth section discusses the obtained results.

2 Backgrounds and Related Works

2.1 Internet of Medical Things (IoMT)

IoMT is a healthcare-specific extension of the Internet of Things (IoT). Wearable health monitoring equipment, wireless body area networks (WBAN), artificial Intelligence (AI), and cloud-based remote health analysis are all possible with IoMT as a health monitoring system. Using IoMT, functional components such as data collection, storage, transmission, and analytics may assist to limit the spread of infectious illnesses by providing an early warning system. Sensor data from end user devices, such as smartphones, tags, or healthcare devices, is sent to a cloud platform for analysis and decision-making [11].

2.2 Machine Learning (ML)

ML is a set of data analysis techniques that enable machines to learn from experience in the same way that humans and animals do [12]. Computational techniques are used in ML algorithms instead of a predefined equation as a model to allow the computer to "learn" information directly from data [13, 14]. The efficiency of the algorithms improves adaptively as the amount of data available for learning increases [15].

One of the most performed results techniques is "boosting," an ML technique used to convert weak learners into strong learners. A weak learner is an ML model that is only slightly better than random guessing, while a strong learner is significantly better than random guessing [16].

- Extreme gradient boosting (XGBoost) is an implementation of gradient boosting; it produces a prediction model in the form of an ensemble of weak prediction models. XGBoost is a three-way split based on decision trees. It uses a gradient boosting system to combine multiple models to produce a more accurate final model [17].
- Light gradient boosting machine (LightGBM) is another implementation of gradient boosting (developed by Microsoft); it is a two-way split based on decision trees. LightGBM can be faster than XGBoost and can handle large-scale data [18].
- CatBoost is an implementation of gradient boosting developed by Yandex, termed by the two terms "Category" and "Boosting." CatBoost is a three-way split based on decision trees. It uses a gradient boosting system to combine multiple models to produce a more accurate final model. CatBoost can handle categorical data and is designed to work with tabular data [19].

These three boosting algorithms are all very effective at producing strong predictive models. They have their own strengths and weaknesses, but the one that is best for a particular problem will depend on the specifics of the data and the problem.

2.3 Intrusion Detection System (IDS)

IDS systems monitor network traffic for suspicious behavior, recognize threats and issue alarms when such behavior is detected. They are a kind of a packet sniffer that looks for irregularities in network traffic or the host's files. The location of the IDS sensors determines whether an intrusion detection system is designated as a network "NIDS," a host "HIDS," or a hybrid that combines them [20]. Each of these intrusion detection systems has its own set of advantages and drawbacks. This implies that NIDS monitors real-time data and indicates problems as they occur [21]. It is also important to note that HIDS analyzes past data in order to capture sophisticated hackers who utilize unconventional tactics that may be difficult to detect in real time. Both HIDS and NIDS are great since they complement one another [22]. System security depends primarily on IDS that identify and react to attacks [23].

2.4 Dataset

A dataset is a digital tool that integrates several data. It can be presented in different formats (videos, images, texts, numbers, sounds, or even statistics). Their grouping forms a whole. In the field of machine learning, the dataset remains indispensable for the creation of ML models [24]. Temperature, SpO2, pulse rate, heart rate, IP addresses, and other characteristics will be assigned to the attributes, while each feature will be associated with a unique device, object, person, and such. In this work, we used a publicly available dataset for IoMT security named "WUSTL EHMS 2020 Dataset" [25], and it was generated using a real-time (EHMS) enhanced healthcare monitoring system testbed. Due to the lack of a dataset that includes both network traffic measurements and patient biometrics, this testbed gathers both. Man-in-the-middle attacks like spoofing and data injection are covered in this data collection.

2.5 Related Works

An approach based on the IoT in medical contexts was developed by [26] in order to create a global connection with the patient, sensors, and everything else. It is the primary objective of their globality feature to make the patient's life simpler and the healthcare procedure more successful. They devised an architecture that provides a wide range of options for communications, monitoring, and control. An anomaly-based IDS (AIDS) for IoMT networks was suggested by [27]. The objective of this new AIDS is to gather log files from the IoMT devices and the gateway, also as traffic from the edge IoMT network, using both host and network-based approaches. They used ML approaches to detect anomalies in the acquired data and thereby identify

malicious occurrences in the IoMT network, taking into consideration the computational cost. Zachos et al. [28] proposed a web-based security assessment framework (IoMT-SAF) to identify security enhancements in IoMT and analyze the protection and prevention provided by IoMT solutions. In [29], biometric and network metrics were combined to demonstrate that employing both kinds of characteristics outperforms using just one alone. Results reveal that performance has increased by 7–25% in certain circumstances, demonstrating the resilience of the proposed system in delivering appropriate intrusion detection.

3 Proposed Method

Patients in an IoMT network architecture are remotely monitored by doctors through dashboards accessed via cloud servers (Fig. 1). This enables doctors to continuously monitor patients' vital signs and other health data in real time, allowing them to make more informed decisions about their care. It also gives patients more control over their health and makes them feel more connected to their doctors. Our proposed IDS will be fully integrated into the IoMT device.

To build such IDS, we take advantage of the WUSTL EHMS dataset to construct ML models that can ultimately predict attacks on the IoMT devices. Figure 2 depicts the different building steps to create our proposed models.

To begin, we need to import the dataset using Python's Pandas library, and we read the CSV files as a dataframe. Then, the data preprocessing is done as a second step. In order to encode categorical data, such as "Dir," "Flgs," and "Sport," we must first pass through the processing stage. The mathematical equations model cannot use these values; thus, we have to define these variable types as categorical. In the third step, the dataset is divided into a training and a validation sub-dataset: one for training the model and one for testing its performance, known as "X" (80%) and "Y"
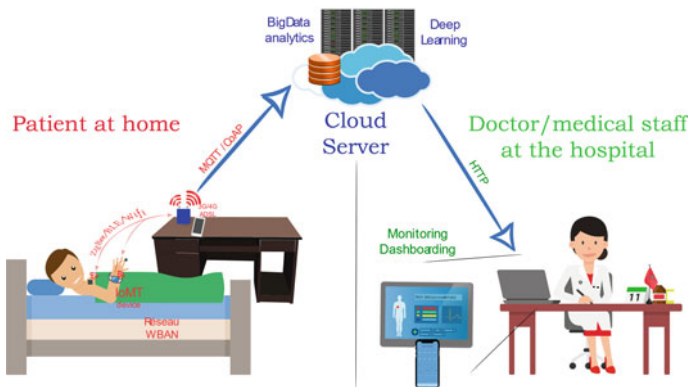
Fig. 1 Architecture of an IMoT network

Fig. 2 Building steps of our models
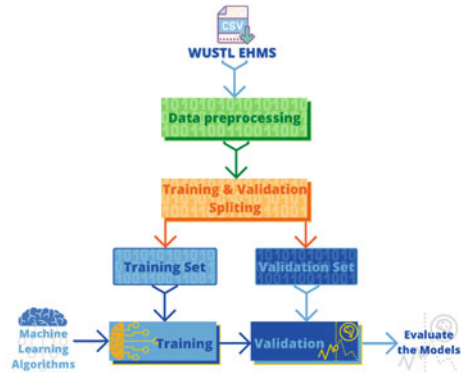
(20%). We used the *K*-folds cross-validator, which offers train and test indices to split the data into train/test datasets, and into *k* consecutive folds, to divide the data. Finally, as the last step, we tune and apply three different proposed ML algorithms to build the most accurate model using the training set and, eventually, test these models using different evaluating metrics on the testing sub-dataset.

4 Results and Discussion

4.1 Hardware and Software Characteristics

The following hardware specifications describe the laboratory server that we have used to perform our experimentation.

- CPU(s): Intel(R) Xeon(R) Silver 4210 CPU 2.20 GHz, (20 cores)
- RAM: 80 GB.

We used scikit-learn, a free and open-source ML framework in our experiments.

4.2 Evaluating the Results

Table 1 and Fig. 3 show the results of three different ML algorithms. The LightGBM algorithm outperforms both the XGBoost and CatBoost ones with an accuracy of 99.28%, a ROC_AUC of 97.98%, a recall of 94.79%, and a precision of 99.46%. Furthermore, the $F1$-score for the LightGBM algorithm is 97.07%, which is the highest of the three algorithms. This shows that the LightGBM algorithm is the best algorithm for this IoMT dataset. Indeed, LightGBM has the ability to learn the structures of the data (categorical features), while XGBoost and CatBoost only

Table 1 Implemented models results

Evaluating metrics	LightGBM	XGBoost	CatBoost
Accuracy (%)	99.28	98.57	98.74
ROC AUC (%)	97.98	94.76	95.91
Recall (%)	94.79	89.72	92.13
Precision (%)	99.46	98.50	97.71
F1-score (%)	97.07	93.91	94.91
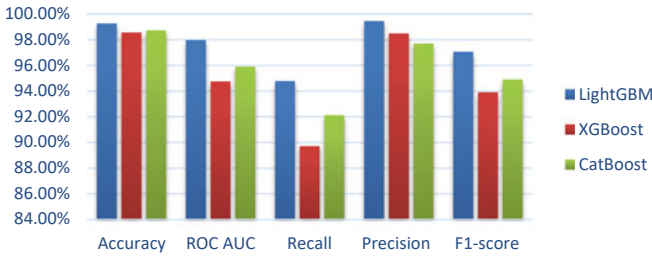
Fig. 3 Comparison of models results

focus on the prediction results. In addition, LightGBM can handle a large number of features, which is beneficial for data with high dimensionality.

Furthermore, when working with large datasets, LightGBM is significantly faster than XGBoost, which can be important when analyzing data. Finally, the results presented here suggest that LightGBM is potentially more accurate than XGBoost in our case study. As the accuracy of the IoMT IDS is critical in identifying potential threats, this could be of interest to the IoMT IDS.

5 Conclusion

An IoMT device is a key tool that is required to assist hospitals in managing this surge of patients. It may link medical staff with patients to assist them in the management of their treatments. Furthermore, IoMT devices may assist in monitoring patients' conditions and help them recover. As these devices grow more ubiquitously, so does the need for security. An IDS is strongly required to protect IoMT devices against attacks.

In this research, we proposed an IDS for IoMTs using gradient boosting machines (XGBoost, LightGBM, and CatBoost) approaches. Many evaluations revealed that these techniques had a high detection rate and could be an effective method of preventing IoMT devices attacks. In particular, the LightGBM algorithm outperformed other techniques in all studied metrics, demonstrating that LightGBM could

be an effective and rapid method of building a strong and real-time IDS for IoMT devices.

References

1. Zoumpourlis V, Goulielmaki M, Rizos E, Baliou S, Spandidos DA (2020) The COVID-19 pandemic as a scientific and social challenge in the 21st century. Mol Med Rep 22:3035. https://doi.org/10.3892/MMR.2020.11393
2. WHO Coronavirus (COVID-19) Dashboard|WHO Coronavirus (COVID-19) Dashboard with vaccination data. https://covid19.who.int/. Last accessed 05 Nov 2021
3. Yang T, Gentile M, Shen C-F, Cheng C-M (2020) Combining point-of-care diagnostics and internet of medical things (IoMT) to combat the COVID-19 pandemic. Diagnostics 2020 10:224. https://doi.org/10.3390/DIAGNOSTICS10040224
4. Boukabous M, Azizi M (2021) A comparative study of deep learning based language representation learning models. Indones J Electr Eng Comput Sci 22:1032–1040. https://doi.org/10.11591/ijeecs.v22.i2.pp1032-1040
5. Kherraki A, Ouazzani RE (2022) Deep convolutional neural networks architecture for an efficient emergency vehicle classification in real-time traffic monitoring. IAES Int J Artif Intell 11:110–120
6. Berrahal M, Azizi M (2020) Review of DL-based generation techniques of augmented images using portraits specification. In: 4th International conference on intelligent computing in data sciences, ICDS 2020, Institute of Electrical and Electronics Engineers (IEEE), pp 1–8. https://doi.org/10.1109/ICDS50568.2020.9268710
7. Boukabous M, Azizi M (2021) Review of learning-based techniques of sentiment analysis for security purposes. In: Innovations in smart cities applications, vol 4. Springer, Cham, pp 96–109. https://doi.org/10.1007/978-3-030-66840-2_8
8. Idrissi I, Boukabous M, Azizi M, Moussaoui O, Fadili H (2021) El: Toward a deep learning-based intrusion detection system for IoT against botnet attacks. IAES Int J Artif Intell 10:110–120. https://doi.org/10.11591/ijai.v10.i1.pp110-120
9. Shwartz-Ziv R, Armon A (2021) Tabular data: deep learning is not all you need
10. xgboost/demo at master dmlc/xgboost. https://github.com/dmlc/xgboost/tree/master/demo#machine-learning-challenge-winning-solutions. Last accessed 04 Nov 2021
11. Mohd Aman AH, Hassan WH, Sameen S, Attarbashi ZS, Alizadeh M, Latiff LA (2021) IoMT amid COVID-19 pandemic: application, architecture, technology, and security. J Netw Comput Appl 174:102886. https://doi.org/10.1016/J.JNCA.2020.102886
12. Boukabous M, Azizi M (2022) Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers. Indones J Electr Eng Comput Sci 25. https://doi.org/10.11591/IJEECS.V25.I2.PP
13. Hammoudi Y, Idrissi I, Boukabous M, Zerguit Y, Bouali H (2022) Review on maintenance of photovoltaic systems based on deep learning and internet of things. Indones J Electr Eng Comput Sci 26
14. What Is Machine Learning?|How It Works, Techniques and Applications—MATLAB and Simulink. https://www.mathworks.com/discovery/machine-learning.html. Last accessed 19 Oct 2021
15. Idrissi I, Azizi M, Moussaoui O (2020) IoT security with deep learning-based intrusion detection systems: a systematic literature review. In: 4th International conference on intelligent

computing in data sciences, ICDS 2020, Institute of Electrical and Electronics Engineers (IEEE), pp 1–10. https://doi.org/10.1109/ICDS50568.2020.9268713

16. XGBoost, Light GBM and CatBoost. A Comparison of Decision Tree…|by OCTAVE-John Keells Group|OCTAVE — John Keells Group|Sep, 2021|Medium, https://medium.com/octave-john-keells-group/xgboost-light-gbm-and-catboost-a-comparison-of-decision-tree-algorithms-and-applications-to-a-f1d2d376d89c. Last accessed 10 Nov 2021

17. Chen T, Guestrin C XGBoost: a scalable tree boosting system. https://doi.org/10.1145/2939672.2939785.

18. Al Daoud E (2019) Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. Int J Comput Inf Eng 13:6–10

19. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2017) CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst 2018-December 6638–6648

20. Idrissi I, Azizi M, Moussaoui O (2022) A stratified IoT deep learning based intrusion detection system. In: 2022 2nd International conference on innovative research in applied science, engineering and technology (IRASET), IEEE, pp 1–8. https://doi.org/10.1109/IRASET52964.2022.9738045

21. Idrissi I, Azizi M, Moussaoui O (2022) An unsupervised generative adversarial network based-host intrusion detection system for IoT devices. Indones J Electr Eng Comput Sci 25

22. Idrissi I, Mostafa Azizi M, Moussaoui O (2022) A lightweight optimized deep learning-based host-intrusion detection system deployed on the edge for IoT. Int J Comput Digit Syst 11:209–216. https://doi.org/10.12785/ijcds/110117

23. Idrissi I, Azizi M, Moussaoui O (2021) Accelerating the update of a DL-based IDS for IoT using deep transfer learning. Indones J Electr Eng Comput Sci 23:1059–1067

24. Berrahal M, Azizi M (2021) Augmented binary multi-labeled CNN for practical facial attribute classification. Indones J Electr Eng Comput Sci 23:973–979

25. WUSTL EHMS 2020 Dataset for internet of medical things (IoMT) Cybersecurity Research. https://www.cse.wustl.edu/~jain/ehms/index.html. Last accessed 05 Nov 2021

26. Jara AJ, Zamora MA, Skarmeta AFG (2010) An architecture based on internet of things to support mobility and security in medical environments. In: 2010 7th IEEE Consumer communications and networking conference CCNC 2010. https://doi.org/10.1109/CCNC.2010.5421661.

27. Zachos G, Essop I, Mantas G, Porfyrakis K, Ribeiro JC, Rodriguez J (2021) An anomaly-based intrusion detection system for internet of medical things networks. Electron 10:2562. https://doi.org/10.3390/ELECTRONICS10212562

28. Alsubaei F, Abuhussein A, Shandilya V, Shiva S (2019) IoMT-SAF: Internet of medical things security assessment framework. Internet Things 8:100123

29. Hady AA, Ghubaish A, Salman T, Unal D, Jain R (2020) Intrusion detection system for healthcare systems using medical and network data: a comparison study. IEEE Access. 8:106576–106584. https://doi.org/10.1109/ACCESS.2020.3000421

Predicting Intensive Care Unit Admission Using Machine and Deep Learning: COVID-19 Case Study

Mohammed Boukabous ⓘ**, Idriss Idrissi** ⓘ**, Mounir Grari** ⓘ**,**
Mostafa Azizi ⓘ**, and Omar Moussaoui** ⓘ

Abstract The COVID-19 pandemic has affected the entire world by causing widespread panic and disrupting normal life. Since the outbreak began in December 2019, the virus has killed thousands of people and infected millions more. Hospitals are struggling to keep up with large patient flows. In some situations, hospitals are lacking enough beds and ventilators to accommodate all of their patients or are running low on supplies such as masks and gloves. Predicting intensive care unit (ICU) admission of patients with COVID-19 could help clinicians better allocate scarce ICU resources. In this study, many machine and deep learning algorithms are tested over predicting ICU admission of patients with COVID-19. Most of the algorithms we studied are extremely accurate toward this goal. With the convolutional neural network (CNN), we reach the highest results on our metrics (90.09% accuracy and 93.08% ROC-AUC), which demonstrates the usability of these learning models to identify patients who are likely to require ICU admission and assist hospitals in optimizing their resource management and allocation during the COVID-19 pandemic or others.

Keywords COVID-19 · ICU · Machine learning (ML) · Deep learning (DL) · IoMT (Internet of medical things)

M. Boukabous (✉) · I. Idrissi · M. Grari · M. Azizi · O. Moussaoui
MATSI Research Lab, ESTO, Mohammed First University, Oujda, Morocco
e-mail: m.boukabous@ump.ac.ma

I. Idrissi
e-mail: idrissi@ump.ac.ma

M. Grari
e-mail: m.grari@ump.ac.ma

M. Azizi
e-mail: azizi.mos@ump.ac.ma

O. Moussaoui
e-mail: o.moussaoui@ump.ac.ma

1 Introduction

Coronavirus is capable of causing serious respiratory illness. As known, this virus is spread through contact with an infected person by his respiratory secretions, saliva, mucus, blood, or touching infected objects or surfaces [1–3]. COVID-19 has been declared a global pandemic by the World Health Organization (WHO), and health officials worldwide are scrambling to contain the virus and are concerned about the virus's potential to harm or kill patients in intensive care unit (ICU). In this context, Internet of medical things (IoMT) can be used in several ways to help manage patients. IoMT is the term given to the current trend of medical devices being connected to the Internet. It can be deployed to contain the spread of COVID-19 [4, 5], as well as helping to manage infected patients who are isolated, track the patients' symptoms, and provide information on how to treat those symptoms.

In some cases, patients with COVID-19 are admitted to ICU despite the absence of any serious symptom. In the absence of a test, health officials make decisions about ICU admission based on factors such as the patient's age and health history. Deciding whether or not to admit a patient with COVID-19 to the ICU is a difficult decision. The need for critical care beds in ICUs has exceeded capacity in many parts of the world, leading to the widespread use of makeshift or shared facilities, the use of general wards to treat ICU patients, and the allocation of ICU beds to patients based on their ability to pay rather than their medical needs, for instance, the price of treating a COVID-19 patient who requires ICU in Canada is estimated to be more than $50,000, compared to $8400 for a heart attack victim [6].

The objective of this study is to ensure the survival of COVID-19 patients at risk by predicting who will need ICU admission during their hospital stay using IoMT devices and machine learning or its subset, deep learning. Machine learning is a data analysis technique that enables computers to learn from previous experiences and improve their performance over time [7, 8]; deep learning on the other hand uses a neural network to learn how to perform a task [9–11]. In the medical field, machine and deep learning have been applied to a variety of tasks, including predicting patient outcomes, diagnosing diseases, and personalized treatment. The most used machine learning algorithms today are gradient boosting machines (GBM) techniques such as XGBoost [12], LightGBM [13], and CatBoost [14], and for deep learning are convolutional neural network (CNN) [15–17], recurrent neural network (RNN) [18, 19], long short-term memory (LSTM) [20], and gated recurrent unit (GRU) [21].

2 Related Work

In recent years, machine and deep learning techniques have been successfully applied to a wide range of biomedical tasks. Magunia et al. [22] developed a machine learning approach capable of accurately predicting the outcomes of COVID-19 patients admitted to the intensive care unit. The explainable boosting machine approach

was the most appropriate method as it allowed predictive modeling of COVID-19 ICU patient outcomes, circumventing the limitations associated with linear regression models. Subudhi et al. [23] compared the performance of 18 machine learning algorithms, where ensemble-based models outperformed other model types in terms of predicting five-day ICU admission and 28-day mortality from COVID-19. They found that C-reactive protein (CRP), lactate dehydrogenase (LDH), and O_2 saturation were important for ICU admission models whereas epidermal growth factor receptor (eGFR) < 60 ml/min/1.73 m^2, and neutrophil and lymphocyte proportions were the most significant predictors of death. For future infectious disease outbreaks like COVID-19, using such models might aid in clinical decision-making. Li et al. [24] developed a DL model and a risk scoring system to predict ICU admission and in-hospital death in COVID-19 patients. The study included 5.766 people under investigation for COVID-19 between February 7, 2020, and May 4, 2020. Oxygen saturation, lactate dehydrogenase, procalcitonin, ferritin, and C-reactive protein were the most important ICU predictors. The DL model predicted ICU admission with an AUC of 78% and mortality with an AUC of 84%. Cheng et al. [25] developed a machine learning-based risk prioritization tool for patients with COVID-19 who require ICU within the first 24 h. The study discovered that the tool had a sensitivity of 72.8%, a specificity of 76.3%, a precision of 76.2%, and an AUC-ROC of 79.9%. The tool could be used to optimize hospital resource management and patient flow planning.

3 Methodology Used

3.1 Dataset

The dataset used was gathered from Sírio-Libanês, São Paulo, and Brasilia hospitals and was annotated and anonymized following global best practices and guidelines [26]. This dataset includes comprehensive COVID-19 patient measurements (see Fig. 1). There are 54 features in the dataset divided into the following categories: demographic information is represented by three features; past diseases are represented by nine features; laboratory test results (such as blood, urine, nasal samples, etc.) by 36 features; and vital signs are represented by six features. These properties are enriched by other calculated variables in each time window, such as the mean and median of vital signs and laboratory tests, as well as the maximums and minimums (max–min) of these variables. For laboratory tests, there are 180 features, and for vital signs, 36 features.
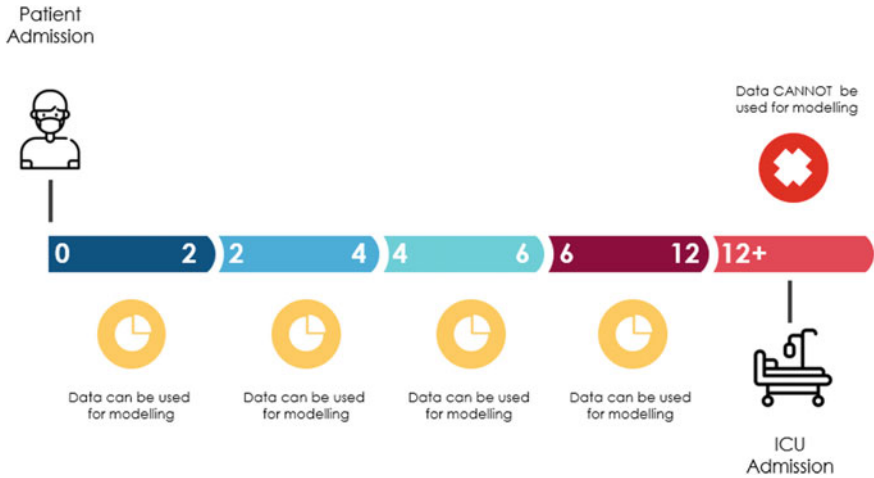
Fig. 1 Patients' data aggregated by time windows (hours) [26]

3.2 Proposed Methodology

The usage of machine and deep learning can help to predict the probability of ICU admission in COVID-19 patients. In this paper, we used multiple machine and deep learning algorithms to learn from the Hospital Srio-Libanês dataset. The advantage of these algorithms is the fact that they are capable to learn and generalize from datasets that are too large to be analyzed by a human.

The first step in our analysis is to import the dataset. We use python's Pandas library to read the CSV file as a data frame. This gives us a comprehensive dataset that we can use for our analysis. The next step is to do some data preprocessing which is an important step for many AI algorithms. The purpose of pre-processing is to clean and prepare the data so that the algorithm can learn from it as effectively as possible. This step can involve, for example, removing noise, transforming the data into a more appropriate format, and aggregating it [27]. Missing values in datasets are a common occurrence and often need to be addressed before any meaningful analysis can be conducted. One method for completing missing values is *k*-nearest neighbors imputation. This involves finding the *k*-nearest neighbors of each missing value and using the values of those neighbors to impute the missing value. This method is lightweight and easy to implement, and often produces good results.

In the third phase, the dataset is divided into two sub-datasets: one for training the model (80%) and one for testing its performance (20%). The data is then divided into *k*-consecutive folds in the case of machine learning algorithms, which allows for different methods to be tested on the training set. The most accurate model is then chosen and tested using different evaluating metrics on the testing set. We experimented with three machine learning and four deep learning algorithms to

determine which one fit our dataset the best. Figure 2 shows the various stages of our proposed model's development.

Table 1 displays the different epochs, batch sizes/units, and layers chosen in the different deep learning algorithms. It can be seen that the number of epochs, batch size, and layer count varies depending on the deep learning algorithm, with most algorithms using a batch size/units of 32.
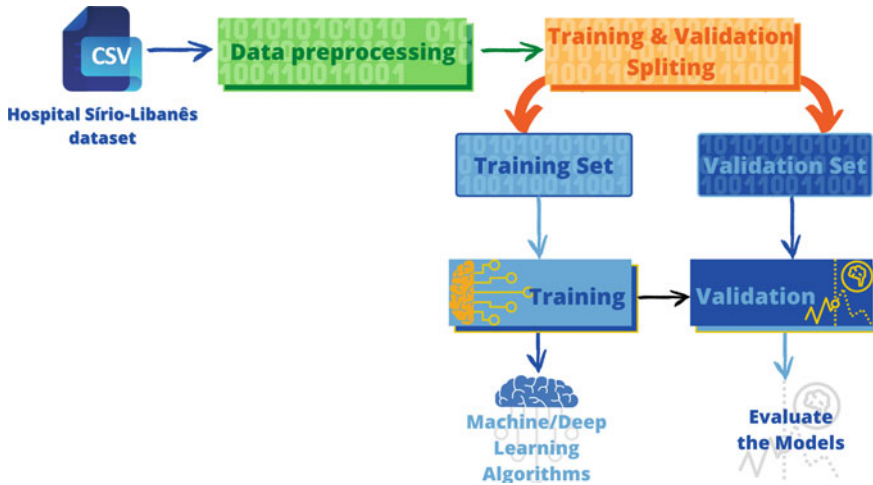
Fig. 2 Building steps of our models

Table 1 DL models layers and parameters

Model	Epochs	Batch size/units	Layers
CNN	100	16	Conv1D (kernel_size = 4) MaxPooling1D (pool_size = 2) Flatten 4 Dense (500, 250, 500, 250, 100, 10 units)
RNN	315	32	2 SimpleRNN (return_sequences = true, false) 2 dropout (0.1)
LSTM	155	32	4 LSTM (return_sequences = true) 4 dropout (0.2)
GRU	115	32	4 GRU (return_sequences = true) 4 dropout (0.2)

4 Results and Discussion

4.1 Hardware Characteristics

The following hardware specifications describe the laboratory server that was used to obtain our findings.

- CPU(s): Intel(R) Xeon(R) Silver 4210 CPU 2.20 GHz, (20 cores)
- RAM: 80 GB
- GPU: NVIDIA GeForce RTX 3090 (24 GB).

We used both Scikit-learn [28] and Keras [29] open-source Python libraries with the TensorFlow-GPU backend engine in our experiment.

4.2 Evaluating the Results

Table 2 shows the accuracy and receiver operating characteristic-area under the curve (ROC-AUC) results of both machine and deep learning models. The accuracy metric is a measure of how well a model correctly predicts the target class; it is calculated by dividing the number of correct predictions by the total number of predictions made [30]. ROC-AUC metric is a measure of how well a model distinguishes between two classes (binary classification). The curves (see Fig. 3) are created for the studied models by tracing the true positive percentage (sensitivity) against the false-positive percentage (specificity) at various thresholds [31].

As result (see Table 2), the CNN model has the highest accuracy (90.09%) and ROC-AUC value (93.08%). The RNN model has the second-highest accuracy (89.35%), followed by GRU (89.12%), LSTM (88.57%), XGBoost (88.40%), CatBoost (88.06%), and LightGBM (87.19%). ROC-AUC is second-highest for CatBoost (92.38%), followed by LSTM (92.22%), GRU (92.18%), RNN (92.09%), XGBoost (91.81%), and LightGBM (90.68%). This indicates that the deep learning models are performing better than the machine learning models for this dataset. The

Table 2 Implemented models results

	Model	Accuracy (%)	ROC-AUC (%)
Deep learning	CNN	90.09	93.08
	RNN	89.35	92.09
	LSTM	88.57	92.22
	GRU	89.12	92.18
Machine learning—ensemble	LightGBM	87.19	90.68
	XGBoost	88.40	91.81
	CatBoost	88.06	92.38

Fig. 3 ROC-AUC of our models
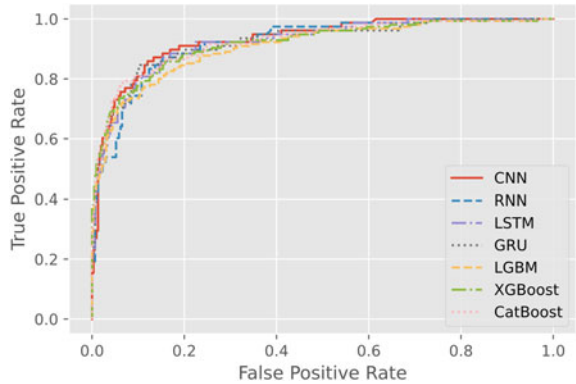
difference in accuracy ranges from 87 to 90%, and the ROC-AUC between 90 and 93% shows that deep learning models are more accurate for this type of data, but not by much.

As illustrated, the CNN model outperforms all other models in the ICU prediction. This is largely due to CNN's ability to recognize patterns over data for critical illness. The ICU prediction model developed using CNN has the potential to improve patient outcomes by providing clinicians with timely information about patients' conditions.

There is significant potential for the application of machine and deep learning algorithms in the prediction of ICU outcomes. These algorithms can predict patient outcomes with a high degree of accuracy and have the potential to improve patient care by providing individualized predictions for patients and helping to guide clinical decision-making.

5 Conclusion

ICU admission is associated with a high risk of death. Predicting ICU admission of infected patients with COVID-19 is important for selection, scheduling, and optimizing hospital resource allocation. In this paper, we investigated the performance of machine and deep learning models for predicting ICU admission of such patients. The obtained results showed that our established models achieve high levels of prediction performance for both accuracy and ROC-AUC (with an overall prediction accuracy over 87 and 90% for the ROC-AUC).

The ICU prediction model using the CNN achieves the highest result, with an accuracy of 90.09%, and a ROC-AUC of 93.08%, which makes the CNN a very effective candidate for predicting ICU admission of patients with COVID-19. Despite the potential for machine and deep learning to improve the quality of ICU prediction, there is still a lack of evidence demonstrating their generalized effectiveness. As future works, we will widen our study to explore more the ability of these algorithms

to improve ICU predictions and contribute to enhance patient admission management outcomes.

References

1. Marik PE, Iglesias J, Varon J, Kory P (2021) A scoping review of the pathophysiology of COVID-19. Int J Immunopathol Pharmacol 35:1–16
2. Goldman E (2020) Exaggerated risk of transmission of COVID-19 by fomites. Lancet Infect Dis 20:892. https://doi.org/10.1016/S1473-3099(20)30561-2
3. CDC says risk of COVID-19 transmission on surfaces 1 in 10,000|TheHill. https://thehill.com/policy/healthcare/546541-cdc-risk-of-covid-transmission-on-surfaces-is-low. Accessed 30 Dec 2021
4. Mohd Aman AH, Hassan WH, Sameen S et al (2021) IoMT amid COVID-19 pandemic: application, architecture, technology, and security. J Netw Comput Appl 174:102886
5. Idrissi I, Azizi M, Moussaoui O (2022) A stratified IoT deep learning based intrusion detection system. In: 2022 2nd International conference on innovative research in applied science, engineering and technology (IRASET), IEEE, pp 1–8
6. Average cost for COVID-19 ICU patients estimated at more than $50,000: report|CBC News. https://www.cbc.ca/news/health/cihi-covid19-canada-hospital-cost-1.6168531. Accessed 30 Dec 2021
7. Bishop CM (2006) Pattern Recognition and Machine Learning. Springer
8. Idrissi I, Azizi M, Moussaoui O (2022) A lightweight optimized deep learning-based host-intrusion detection system deployed on the edge for IoT. Int J Comput Digit Syst 11:209–216. https://doi.org/10.12785/IJCDS/110117
9. Idrissi I, Boukabous M, Azizi M et al (2021) Toward a deep learning-based intrusion detection system for iot against botnet attacks. IAES Int J Artif Intell 10:110–120
10. Kherraki A, Ouazzani R El (2022) Deep convolutional neural networks architecture for an efficient emergency vehicle classification in real-time traffic monitoring, IAES Int J Artif Intell 11:110–120. https://doi.org/10.11591/IJAI.V11.I1.PP110-120
11. Berrahal M, Azizi M (2022) Improvement of facial attributes' estimation using Transfer Learning. In: 2022 2nd International conference on innovative research in applied science, engineering and technology (IRASET), IEEE, pp 1–7
12. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings ACM SIGKDD international conference knowledge discovery data Mining 13–17-August-2016, pp 785–794
13. Ke G, Meng Q, Finley T et al. LightGBM: a highly efficient gradient boosting decision tree
14. Prokhorenkova L, Gusev G, Vorobev A et al. (2017) CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst 2018-December:6638–6648
15. LeCun Y, Bengio Y (1995) Convolutional networks for images, speech, and time series. Handb Brain Theory Neural Netw 3361
16. Berrahal M, Azizi M (2021) Augmented binary multi-labeled CNN for practical facial attribute classification. Indones J Electr Eng Comput Sci 23
17. Kherraki A, Maqbool M, El Ouazzani R (2021) Traffic scene semantic segmentation by using several deep convolutional neural networks. In: 2021 3rd IEEE Middle East and North Africa COMMunications conference (MENACOMM), IEEE, pp 1–6
18. Schuster M, Paliwal KK (1997) Bidirectional recurrent neural networks. IEEE Trans Signal Process 45. https://doi.org/10.1109/78.650093

19. Hammoudi Y, Idrissi I, Boukabous M et al. (2022) Review on maintenance of photovoltaic systems based on deep learning and internet of things. Indones J Electr Eng Comput Sci 26:1060–1072. https://doi.org/10.11591/ijeecs.v26.i2.pp1060-1072
20. Boukabous M, Azizi M (2021) A comparative study of deep learning based language representation learning models. Indones J Electr Eng Comput Sci 22:1032–1040
21. Cho K, Van Merriënboer B, Gulcehre C et al. (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. EMNLP 2014—2014 Conf Empir Methods Nat Lang Process Proc Conf 1724–1734
22. Magunia H, Lederer S, Verbuecheln R et al. (2021) Machine learning identifies ICU outcome predictors in a multicenter COVID-19 cohort. Crit Care 25
23. Subudhi S, Verma A, Patel AB et al. (2021) Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19. NPJ Digit Med 4
24. Li X, Ge P, Zhu J et al. (2020) Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables. PeerJ 8
25. Cheng FY, Joshi H, Tandon P et al (2020) Using machine learning to predict ICU transfer in hospitalized COVID-19 patients. J Clin Med 9. https://doi.org/10.3390/JCM9061668
26. COVID-19—Clinical data to assess diagnosis|Kaggle. https://www.kaggle.com/Sírio-Libanes/covid19. Accessed 17 Jan 2022
27. Boukabous M, Azizi M (2022) Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers. Indones J Electr Eng Comput Sci 25
28. Pedregosa F, Varoquaux G, Gramfort A et al. (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12
29. Keras: the Python deep learning API. https://keras.io/. Accessed 18 Aug 2020
30. Boukabous M, Azizi M (2022) Multimodal sentiment analysis using audio and text for crime detection. In: Institute of electrical and electronics engineers (IEEE), pp 1–5
31. Idrissi I, Azizi M, Moussaoui O (2022) An unsupervised generative adversarial network based-host intrusion detection system for IoT devices. Indones J Electr Eng Comput Sci 25

Neural Network-Based Precision Irrigation Scheduling and Crop Water Stress Index (CWSI) Assessment

Benzaouia Mohammed, Hajji Bekkay, Mokhtari Hassan, and Chaabane Khalid

Abstract The efficient use of water by irrigation systems is critical for sustainable agricultural development, food security, and overall economic growth. This is especially true in light of global population growth, climate change, and competitive water demands from other economic sectors. To meet this challenge, developing a precision irrigation (PI) strategy is targeted. The developed technique consists of applying the exact amount of water needed by the crop according to its phonological stage through assessing the crop water stress index (CWSI), climatic conditions, and soil water stock. The strategy is based on a neural network to improve the system's behavior and render it intelligent and automatic. The obtained results lead to increased efficiency, saving irrigation water, and consequently, increasing water productivity.

Keywords Precision irrigation · Irrigation scheduling · Crop water stress index (CWSI) · Soil moisture · Neural network controller

1 Introduction

Moroccan agriculture represents 7% of the national water and energy consumption, a number expected to increase in the coming years [1, 1]. Faced with the climatic and energy challenges raised with acuity by the use of conventional irrigation methods, the development and implementation of strategy-based intelligent precision irrigation systems became a necessity and a global vision [3, 3].

In general, irrigation schedulers are classified into three categories: the soil-based approach, the climate change-based approach, and the approach based on the evolution of parameters for a specific crop [3, 3]. The first approach (soil-based) consists in using the soil moisture measurement (SM), which represents the volume of water

stored by the soil in agricultural fields. In this case, the soil behaves as a reservoir that stores water. In [6], the authors propose an irrigation method that allows setting soil moisture percentage to a fixed value. Irrigation is scheduled, therefore, when this percentage is less than the pre-defined range. The second approach (based on climatic changes) is based on assessing parameters related to weather, such as ambient temperature, air humidity, solar irradiation, and precipitation. Therefore, the irrigation can be scheduled when these parameters reach minimum values so that the water is not evaporated rapidly. The last approach (plant-based) is founded on theories that quantify the water stress of plants through leaf water potential and surface temperature. A combination of several approaches is possible and allows better control and programming of the irrigation duration and time while exploiting artificial intelligence solutions (fuzzy logic, neural network, etc.), resulting in high-efficiency gains and reliability in reducing water and energy losses [7].

To schedule irrigation automatically and intelligently, in [8] a presentation of a real-time automated irrigation controller based on fuzzy inference implemented in LabVIEW, and the use of data recorded from the sensor network and communicated using the GSM/GPRS module. The proposed drip irrigation scheduler receives data from soil moisture, humidity, air temperature, and water level sensors to decide the required opening valve percentage. Other research focuses on developing an ON–OFF system and ANN controllers that are applied to the automated irrigation system. The control of the irrigation system used a Penman–Monteith evapotranspiration calculation and a soil moisture reference as comparative input [4].

In this paper, a combined plantsoil-based approach is implemented for the selected plant Lactuca sativa, the crop water stress index (CWSI) is used as an estimator to quantify the plant water status (crop water deficit) at any local point based on canopy temperature measurements by an infrared temperature sensor, as well as the soil water content measured by a capacitive sensor. These two variables represent the input of the neural network controller. The output is the valve opening rate; for a CWSI equal to 1, the opening rate is 100% since there is water stress, thus a requirement of irrigation; otherwise, the valve is closed. The main contributions of this article are:

- An intelligent irrigation scheduling system based on a neural network is designed.
- Crop water stress index (CWSI) and soil water content are simultaneously considered variables for irrigation scheduling strategy.

The remainder of this paper is organized into five main sections. In sect. 2, a description of the irrigation system is presented. The proposed irrigation scheduling strategy is discussed in sect. 3. Section 4 describes the development of the strategy based on a neural network. Section 5 describes the steps of operation of the irrigation system. The analysis of the results is further elaborated in sect. 6 and, finally, a conclusion.
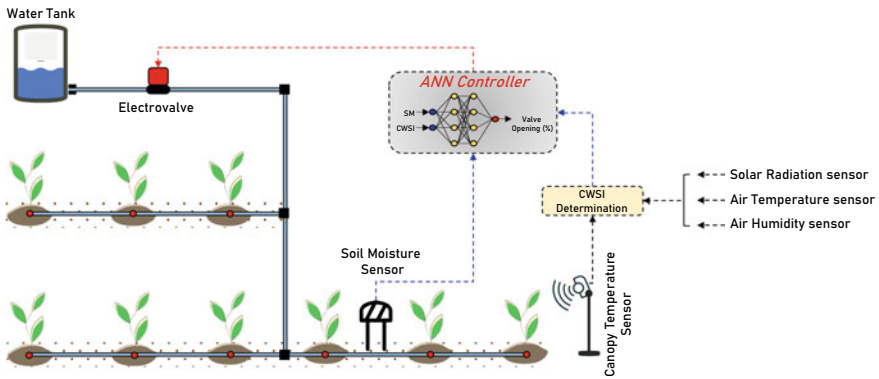
Fig. 1 Structure of the irrigation scheduling system

2 Irrigation System Description

The architecture of the irrigation system is illustrated in Fig. 1; it consists of two parts; the first part is the sensor network, including the temperature and air humidity sensor, DHT11 sensor can be used to sense these two variables. A pyranometer is used to measure the global solar irradiation reaching the ground, and the soil moisture sensor (MS10) is used to accurately measure the soil wetness rate (placed in the root zone of the plant). The second part is the control; a neural network-based controller is implemented to determine the valve opening rate, depending on the system's input variables, which are soil moisture and CWSI. After obtaining the controller output, the valve opening percentage is converted into a control signal.

3 Irrigation Strategy Development

The estimation and determination of the water requirements of a plant correspond to the evaluation of the crop water stress index (CWSI) through the measurement of the crop surface temperature. The relationship between surface temperature and water stress is based on the assumption that when a crop transpires, the evaporated water cools the leaves to a lower temperature than the air. When it is subjected to water stress, transpiration decreases and, consequently, leaf temperature increases. Two different techniques of calculation of CWSI can be distinguished, namely, the empirical and theoretical methods.

Empirical $CWSI_E$ is calculated through Eq. (1), which uses the ratio between the difference of the actual canopy temperature (T_c) and the non-water-stressed baseline (T_{ns}) obtained by identifying the canopy temperature of a well-watered crop transpiring at a maximum rate and the difference between the water-stressed baseline (T_d) obtained by the canopy temperature of a non-transpiring plant leaf

and the non-water-stressed baseline (T_{ns}). However, the determination of T_{ns} and T_d requires a precise measurement of artificial reference surfaces, which limits the use of $CWSI_E$ in real-time applications.

$$\text{CWSI}_E = \frac{T_c - T_{ns}}{T_d - T_{ns}} \tag{1}$$

In this paper, the theoretical $CWSI_T$ formulated by Eq. 2 is used:

$$\text{CWSI}_T = \frac{dT_e - dT_l}{dT_u - dT_l} \tag{2}$$

where dT_e is the difference between the actual canopy temperature and the air temperature ($T_c - T_{air}$), dT_l is the difference between the canopy temperature and the temperature of a well-watered plant, calculated by Eqs. (3)–(6). Assuming that the stomata are closed for a non-transpiring canopy, the difference between the canopy temperature and the non-transpiring plant canopy temperature dT_u can be determined by Eq. (7).

$$dT_l = \frac{R}{\beta + \Delta/P} - \frac{\text{VPD}}{P(\beta + \Delta/P)} \tag{3}$$

$$R = \frac{1}{4}(\alpha_s S_r + \tau \alpha_s S_r + 4L(\alpha_L - 1)) \tag{4}$$

$$\beta = \frac{2c_p g_h - \varepsilon \sigma T_{air}^3 (3\alpha_L - 4)}{\alpha g_v} \tag{5}$$

$$g_h = 0.189 \sqrt{\frac{u}{d}} \tag{6}$$

$$dT_u = \frac{R}{2c_p g_h - \varepsilon \sigma T_{air}^3 (3\alpha_L - 4)} \tag{7}$$

where R is the net radiation, β is the constant of psychometric, Δ is the relationship slop between saturation vapor pressure and air temperature, P is the atmospheric pressure, VPD is the vapor pressure deficit, α_s, α_L are, respectively, the absorptivity of thermal waveband, g_h is the boundary layer conductance to heat, c_p is the capacity of heat in the air, σ is the constant of Boltzmann, u is the wind speed, and d is the characteristic size of the sheet.

CWSI is rated between 0 and 1, with CWSI = 0 meaning there is no water stress (well-watered condition), and a value of 1 represents maximum water stress. Therefore, CWSI is used to quantify the instant water status of a crop and for scheduling irrigation times.

4 Neural Network Irrigation Scheduling Strategy

Inspired by the brain physiology of humans, artificial neural networks are software programs that can emulate neural networks. They use self-learning processes and can create relationships between the memory and the information itself. ANN-based control systems have been used in many fields such as electrical system control, aeronautics, predictive maintenance and control of robots, and various applications.

For precision irrigation application, the feed-forward back propagation neural network is used to estimate the output of our system. The architecture contains three types of layers. CWSI and soil moisture information are received on the input layer and an output layer having a single neuron and giving the result of the calculation and estimation of the percentage of solenoid valve opening. There is a specific number of hidden layers in the middle of the two preceding layers (the input and output layers). Choosing the optimal number of hidden layers is essential; an inappropriate number of neurons leads to poor information processing, resulting in weak performance and robustness. The ANN configuration of 2:5:1 is adopted, i.e., an input layer of two neurons, a hidden layer of five neurons, and an output layer of one neuron, as illustrated in Fig. 2.

The output neuron U_{out}, can be expressed by the following Equation:

$$U_{\text{out}} = g(U_{\text{net}}) \tag{8}$$

$$U_{\text{net}} = \sum_{j=1}^{M} W_{y_j} Y_{\text{out}_j} \tag{9}$$

$$Y_{\text{out}_j} = f\left(Y_{\text{net}_j}\right) \tag{10}$$

$$Y_{\text{net}_j} = \sum_{i=1}^{N} W_{x_{i,j}} x_i \tag{11}$$
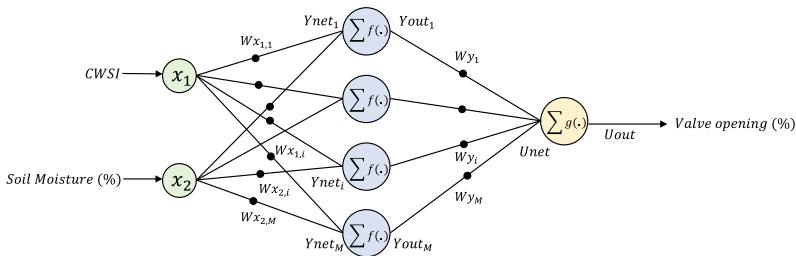
Fig. 2 Structure of the ANN controller

where $g(.)$ is the activation function of the output, U_{net} is the output layer, Y_{net} is the input of the hidden layer, Y_{out} is the output of the hidden layer, W_{y_j} is the weight between the hidden and the output layer, $W_{x_{i,j}}$ is the weight between the input and the hidden layer, N is the number of the input layer, and M is the number of hidden layer nodes.

The database contains 200 data values for each input (noting that the data are collated hourly) and it is divided as follows: 70% of the data is used for training, 15% for validation, and finally, 15% for testing. Figure 3 shows the training, validation, and test regression values for the dataset using the ANN models. It is found that the structure adopted is correct and can also be used to predict the output for other input data sets.
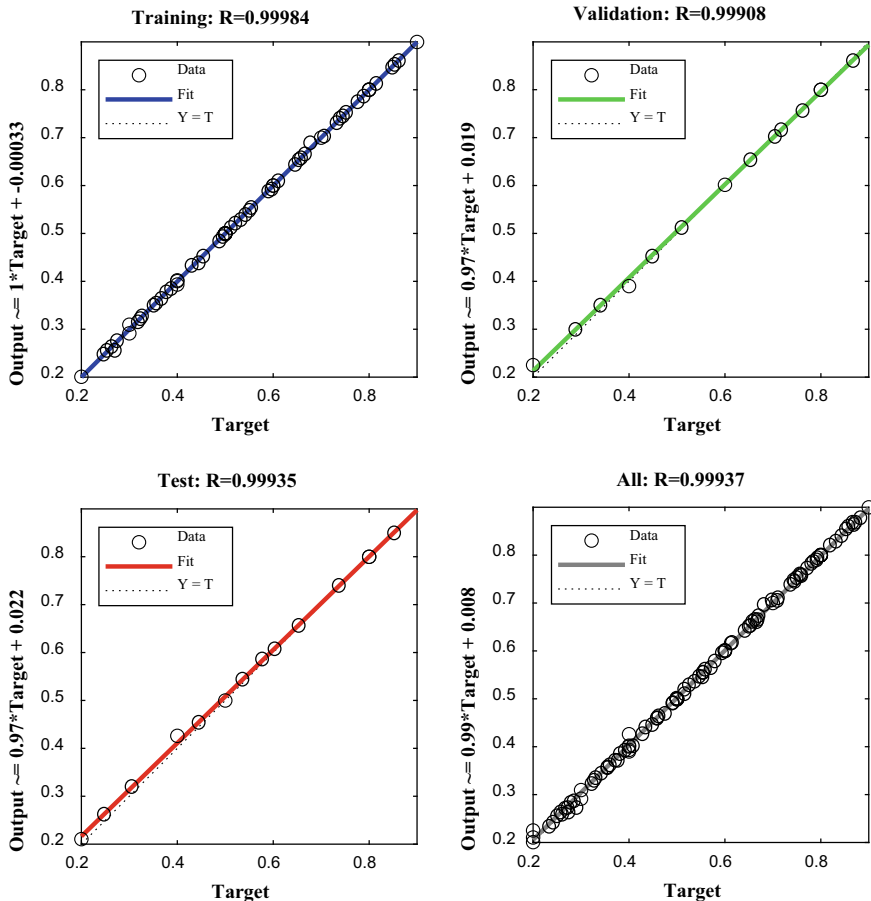
Fig. 3 Regression value for the ANN model's training, validation, and testing data
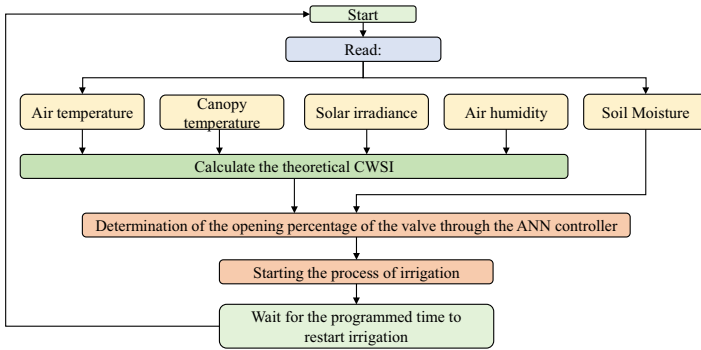
Fig. 4 Flowchart software of the proposed irrigation system

5 System Operation Steps

Figure 4 represents the different stages of operation of the irrigation system. In the first step, the parameters, solar irradiation, air temperature, and canopy and air humidity, are measured to calculate and determine the $CWSI_T$, another measurement of soil moisture is necessary. The second step is processing and analyzing information by the ANN controller to determine the valve opening percentage. Between each irrigation procedure, the system is programmed to wait 120 min after starting the last application of water to the soil since the rate of water aspiration is estimated to be around 120–135 min.

6 Results and Discussion

The behavior of the proposed irrigation system and strategy are examined and evaluated in Fig. 5. Typical seven-day data from the summer season (August 2021) have been used. The variations of meteorological conditions, including air humidity (H_a), air temperature (T_a), canopy temperature (T_c), and solar irradiance (S_r) are illustrated in Fig. 5a, b, c and d, respectively. It is noticed that solar irradiation increases results in air humidity diminution, while the temperature varies in the reverse way of the last variable. In addition, the temperature of the canopy follows the evolution of the air temperature. In this part, the $CWSI_T$ (Fig. 5e) is determined only for the whole day period, since in the night or dawn period the water does not evaporate, and the plant's needs are not exigent, it can be seen that this rate reaches a maximum during high weather conditions and conversely.

The opening of the valve is given by Fig. 5f, a max of opening reflects a high $CWSI_T$. The evolution of soil moisture after irrigation is given by Fig. 5g; the rate increases after irrigating and remains constant or decreases slowly during the nights and dawn.
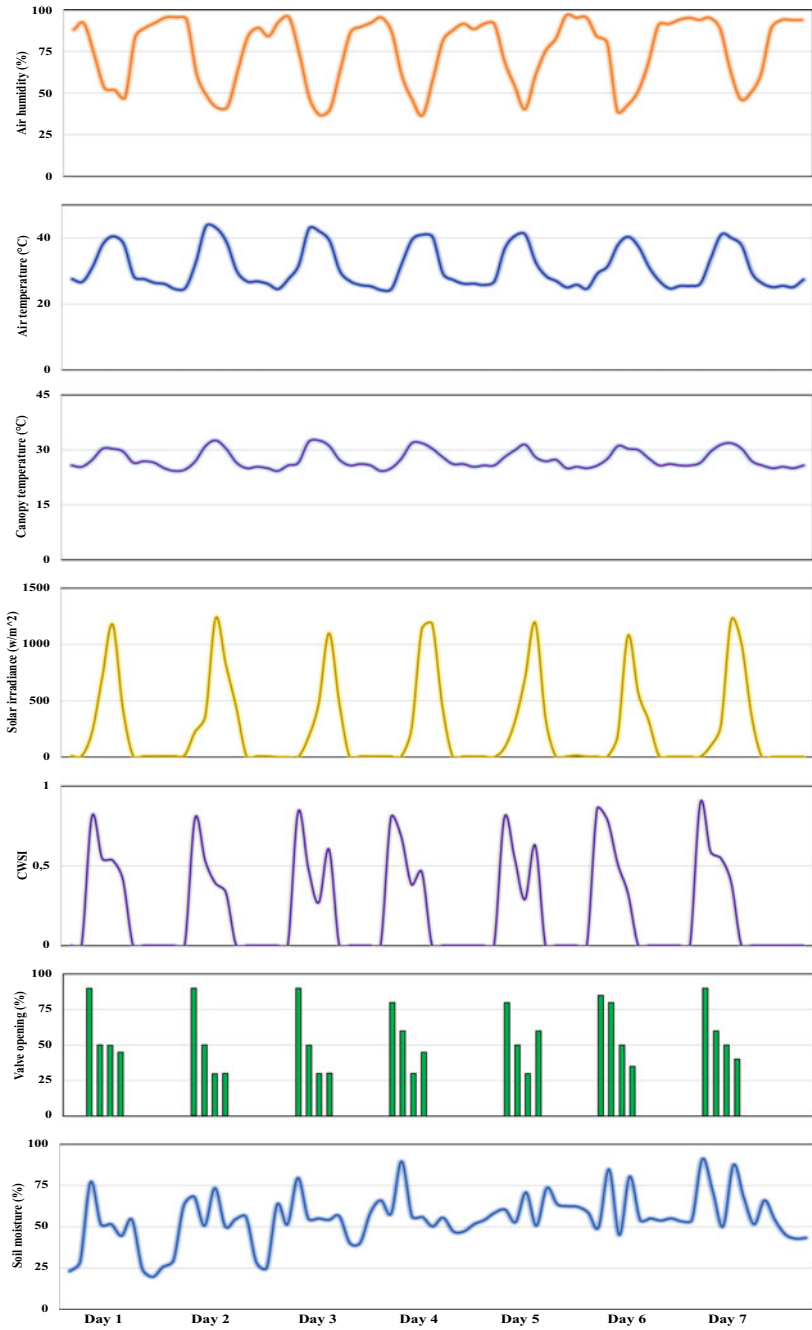
Fig. 5 Evolution of data during the seven days. **a** Air humidity, **b** air temperature, **c** canopy temperature, **d** solar irradiance, **e** CWSI, **f** valve-opening rate, and **g** soil moisture

7 Conclusion

This paper presents the design of an intelligent precision irrigation scheduling system using a strategy that considers both varieties of soil water content and theoretical crop water stress index $CWSI_T$ simultaneously, while using a feed-forward back propagation neural network (PB - ANN). Based on the obtained results, the proposed irrigation scheduling system intends to significantly improve the reliability and accuracy of irrigation, leading to water conservation and rationalization.

References

1. Toumi J, Er-Raki S, Ezzahar J, Khabba S, Jarlan L, Chehbouni A (2016) Performance assessment of AquaCrop model for estimating evapotranspiration, soil water content and grain yield of winter wheat in Tensift Al Haouz (Morocco): application to irrigation management. Agric Water Manag 163:219–235
2. Benzaouia M et al. Energy management strategy for an optimum control of a standalone photovoltaic-batteries water pumping system for agriculture applications. In: International conference on electronic engineering and renewable energy
3. Abioye EA, Abidin MSZ, Mahmud MSA, Buyamin S, Ishak MHI, Abd Rahman MKI, Ramli MSA (2020) A review on monitoring and advanced control strategies for precision irrigation. Comput Electron Agric 173:105441
4. Jaiswal S, Ballal MS (2020) Fuzzy inference based irrigation controller for agricultural demand side management. Comput Electron Agric 175:105537
5. Touati F, Al-Hitmi M, Benhmed K, Tabish R (2013) A fuzzy logic based irrigation system enhanced with wireless data logging applied to the state of Qatar. Comput Electron Agric 98:233–241
6. Benzaouia M, Bouselham L, Hajji B, Dubois AM, Benslimane A, El Ouariachi M (2019, November) Design and performance analysis of a photovoltaic water pumping system based on DC–DC boost converter and BLDC motor. In: 2019 7th International renewable and sustainable energy conference (IRSEC), IEEE, pp 1–6
7. Ma'Mun SR, Loch A, Young MD (2020) Robust irrigation system institutions: a global comparison. Glob Environ Chang 64:102128
8. Romero R, Muriel JL, García I, de la Peña DM (2012) Research on automatic irrigation control: state of the art and recent results. Agric Water Manag 114:59–66

IoTScal-SC2: Two Cloud Computing Systems-Based Collaboration Solution for Scalability Issue in IoT Networks

Abdellah Zyane, Mohamed Nabil Bahiri, and Abdelilah Ghammaz

Abstract Cloud computing and the Internet of things (IoT) are technologies that give services to a wide range of customers by allowing permitted data to be exchanged and consumed in order to make smart and automatic decisions. As a result of the tremendous expansion in the number of connected devices, achieving scalability with acceptable quality of service (QoS) metrics through service-level agreements (SLAs) is a challenge. In this paper, we will begin by outlining our perspective on the scalability issue in IoT networks. Then we will present a new proposed collaboration solution for the scalability challenge, which is based on cloud computing. The proposed solution respects the ETSI architecture standard in IoT networks. The objective is to present a collaboration solution integrating cloud computing, aiming to tackle the scalability issue in IoT networks, by maximizing the number of satisfied requests while maintaining the quality of service at a good level.

1 Introduction

The Internet of things (IoT) concept has recently become more relevant, as has the number of connected objects in every potential industry. Specifically, the industry sector's interest has risen quickly, particularly in the smart city application sector. Furthermore, more research is conducted, resulting in an exponential increase in the number of IoT devices. Furthermore, based on ETSI[1] architecture, IoT devices

---

[1] **ETSI:** European Telecommunications Standards Institute.

A. Zyane (✉) · M. N. Bahiri
S.A.R.S. Team-ENSA Safi, Cadi Ayyad University, Safi, Morocco
e-mail: a.zyane@uca.ma

M. N. Bahiri · A. Ghammaz
L.S.E.E.T, FST-G, Cadi Ayyad University, Marrakech, Morocco

generate streams of data from every imaginable geographical location. As a result, scalability is emerging as one of the most crucial issues [1, 2] for IoT development and deployment, defined as the capacity to automatically maintain an acceptable level of system production performance in the occurrence of a consumer or data overflow [3, 4].

The goal is to increase the number of satisfied sensor layer requests to IoT networks while keeping system performance at acceptable levels in terms of QoS[2] metrics and assuring scalability. The proposed solution will achieve this by collecting events from the managed system, which is the OM2M[3] platform, and then generating symptoms. The activation of dynamic adaption decisions made by the middleware in order to unload the system will be triggered by these symptoms.

Furthermore, cloud computing is the answer to IoT scalability problems. Middleware is also a software layer that resides between connected objects and platforms [4]. It gives developers the tools they need to improve the IoT system's architecture.

The paper is organized as follows: Sect. 2 describes the IoT application domains and architectures as well as our analysis for the IoT networks scalability problem. Section 3 details our proposed collaboration solutions. In Sect. 4; however, we present and analyze numerical results. Finally, Sect. 5 concludes the paper.

2 Problematic

IoT has now become a goal in almost every field. The demand for connected objects is skyrocketing. As a result, Cisco expects exponential growth, from 7.2 billion connected things in 2012 to about 50 billion in 2025. This will make IoT networks incredibly popular and active, driven by the large number of connected objects/users/applications/services, which we call consumers, causing consumer overload. Also, the massive volume of data will cause data overloads [3, 4].

The proposed scalability vision is classified in terms of (objects/users/applications/services) consumers overload at the application layer of the ETSI architecture [3, 4], which will affect the OM2M platform's processing power due to the large number of consumers exchanging data to-and-from different services.

When a consumer overload occurs at the application layer in an IoT system, the OM2M platform resumes its regular function. This normal behavior causes incoming traffic to be queued, increasing response time and perhaps causing a decrease in satisfied traffic. This will slow down responses, reduce the quantity of satisfied requests, and prevent scaling. Traffic will no longer adhere to the ITU-T G.1010 standard of a response time RTT[4] of less than 4 s in this scenario. In addition, we used the highest success rate rates found in the literature for high priority traffic.

---

[2] **QoS:** Quality of service.

[3] **OM2M:** Open-Source platform for M2M communication.

[4] **RTT**: Round Trip Time.

The proposed solution tries to make the system scalable in the situation of an overload in order to successfully address that issue. Scaling will be accomplished by increasing the number of satisfied requests (lowering the loss rate) while maintaining the OM2M platform's QoS metrics (RTT evolution, CPU[5] utilization, and RAM[6] consumption).

The new approach is based on a collaboration architecture that focuses on the monitoring component and scalability issues, taking into account QoS indicators (RTT evolution, CPU utilization, and RAM consumption), and leveraging middleware monitoring capabilities and cloud computing features [5–7].

3 Related Works

According to the literature, the issue of scalability in IoT networks in relation to collaboration has not been adequately stated, discussed, or addressed.

Bahiri et al. proposed a base cloud computing solution improving scalability in IoT networks [8], using autonomic and dynamic scalability capabilities. This solution improved scalability without deteriorating system performance in terms of RTT and resource consumption (RAM, CPU).

4 Proposed Solving Approaches

4.1 Proposed Architectures

In this paper, we intend to present an ETSI-compliant middleware level definition for IoT networks with dynamic and autonomic scalability-oriented capabilities to satisfy a maximum number of requests without degrading the OM2M platform's QoS metrics (RTT). We deployed an autonomous MAPE-K cycle (monitor, analyze, plan, execute, and knowledgebase) of the computing paradigm to achieve our goal [9] (see Fig. 1).

The proposed approach system in general, as shown in Fig. 2, is composed of injectors, middleware, and a platform.

The proposed system is composed of:

---

[5] **CPU:** Central Processing Unit.

[6] **RAM**: Random Access Memory.
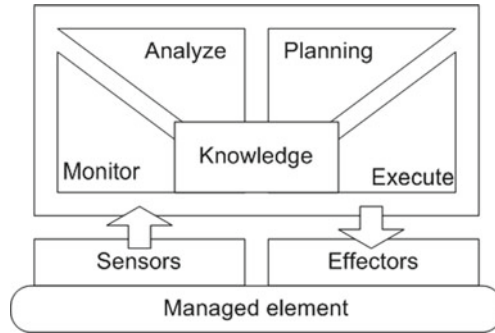
Fig. 1 MAPE-K loop by IBM [9]
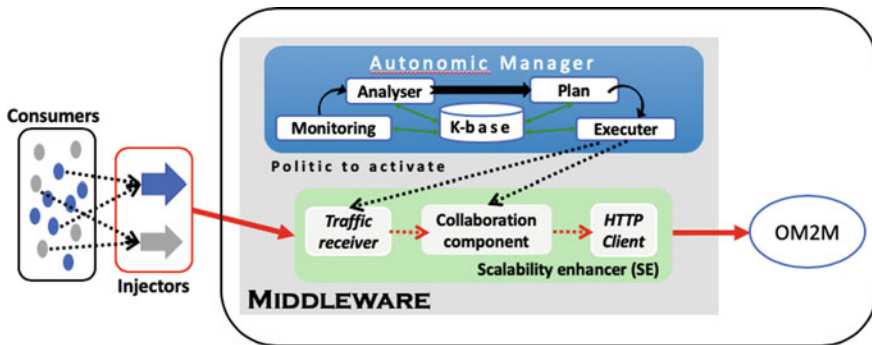
Fig. 2 Our system architecture

4.1.1 Injectors

Injectors simulate real-world consumers by generating traffic. This traffic requires a loss rate under 10% and a response time under 4 s.

Injectors generate requests that are defined by the number of injected requests, their type, their periodicity, and the targeted service on the OM2M platform. Those generated requests will make their way to the platform via the middleware.

4.1.2 The Middleware

The middleware is hosted in a physical machine, which is composed of two main components:

The Autonomic manager assures the following tasks:

(i) Monitor: collects RTT evolution, RAM consumption, and CPU usage of the platform. Then, it sends the collected metrics to the complex event processing (CEP) to generate the adequate symptoms. (ii) Analyzer/Plan: generates plans based on

Fig. 3 Scalability enhancer component
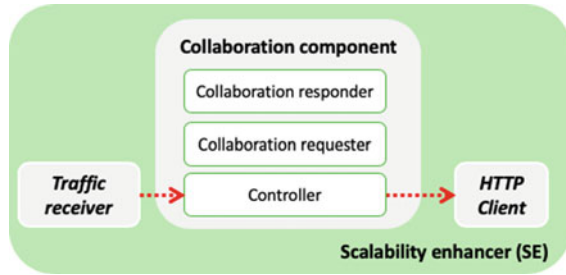
analyzed symptoms. (iii) Executer: sends the policy that will be executed by the receiver, and the collaborative component.

Scalability Enhancer (**SE**): will hold the following components (see Fig. 2):

Traffic receiver: receives requests from the traffic generators:

Collaboration Component (CC): composed of (Fig. 3):

(I)   *controller* based on system states, which takes the decision rather to send the traffic to the OM2M platform or to the Collaborative Cloud,

(II)   *Collaboration requester:* in case of an overload, it sends a request for collaboration to the Collaborative Cloud, and

(III) *Collaboration responder:* returns the systems ID and accepts or rejects the collaboration request based on the system state.

HTTP client: sends traffic to the specified destination.

4.1.3   OM2M Platform

The ETSI-compliant OM2M open-source platform, which is the destination for forwarded requests, is referred to as **OM2M**. This platform hosts a number of services that respond to post and get requests from applications.

4.1.4   IoT-SCal-SC2 High-Level Architecture

The proposed solution is made up of numerous systems that are comparable (three in our case). The operation of collaboration is managed at the same level as other systems. Each system processes traffic in an indecent manner.

An overloaded system and a Collaborative Cloud hosting two systems make up the proposed architecture. Both systems, as indicated in Fig. 4, are holding the identical components.

The studied overloaded system is **System A**, The Collaborative Cloud overloaded system is **System B** and the Collaborative Cloud unloaded system is **System C**.
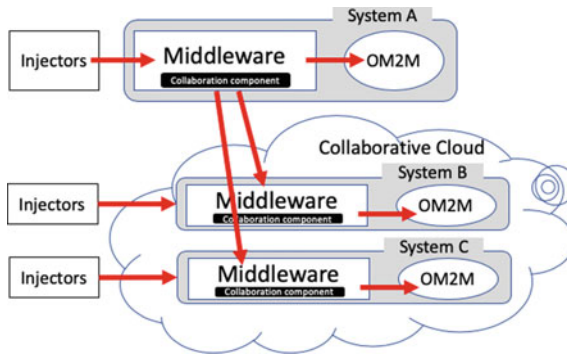
Fig. 4 IoTScal-SC2 high-level architecture of our normal collaborative solution—normal IoTScal-even collaborative

In IoTScal-SC2, in the occurrence of an overload, the collaboration component hosted in the middleware of the overloaded system (System A) will choose a Collaborative Cloud system at random to reroute a portion of its traffic to. In this situation, if the randomly selected system (System B) is unable to manage the collaborative traffic, it (System B) will notify the requesting system (System A) via a **Reject** notification, and the portion of the traffic will not be sent from system A to System B.

If the randomly selected system (System C) is capable of managing the collaboration traffic, it will send the requesting system (System A) an Accept notification. Afterward, System A will send its collaborative traffic to System C, then (system C) will forward the traffic to its OM2M platform.

Each system's collaboration component—hosted in the middleware—makes its own decisions based on each system's state. The collaboration component will use a variety of metrics to determine when and how much traffic can be forwarded to the Collaborative Cloud (System B or C) to help the overloaded system (System A) recover quickly.

4.2   The Proposed Mechanisms

The goal of the proposed approach is to use collaboration in order to maximize the number of satisfied requests in IoT networks. To put it another way, this approach aims to scale the system based on current load and QoS metrics (CPU, RAM, RTT). We will present our proposed mechanisms in this section.

Algorithm 1 Traffic monitor algorithm

```
Begin
    IntLossRate;
    LossRate = GetLossRate();
    If(LossRate < 10%)
    Symptom = 1; // preferable
    Else // LossRate >= 10%
        Symptom = 2; // Critical
    Endif
    End
```

Algorithm 2 Traffic monitor algorithm

```
Begin
    IntRTT[5],Symptom;
    RTT = GetRttFromCEP();
    If(RTT[0] < 4000 & RTT[1] < 4000 & RTT[2] < 4000
    &RTT[3] < 4000 & RTT[4] < 4000)
    Symptom = 1; // Preferable
    Else if (RTT[0] > 4000 & RTT[1] > 4000 &
    RTT[2] < 4000 & RTT[3] > 4000 & RTT[4] < 4000)
        Symptom = 2; // Critical
    Endif
    End
```

Algorithm 1 presents the success rate monitor. When the success rate is over 90% (the loss rate is under 10%), a preferable *(1)* symptom is generated. When the success rate is under 90% (the loss rate is over 10%), a critical *(2)* symptom is generated.

Algorithm 2 presents the traffic monitor. After receiving a window of five consecutive RTTs, when all five RTTs are under 4 s, a preferable symptom is generated; and when all five RTTs are over 4 s, a critical symptom is generated.

We will describe the algorithms for the **collaboration component** (**CC**), which consists of three components: a collaboration requester, a collaboration responder, and a controller.

In the occurrence of an overload in System A, **the collaboration requester** sends a request for collaboration with a system ID to the **Collaborative Cloud.**

The Collaborative Cloud generates for each of the systems a unique ID. This ID will be used in the collaboration process. At first, after receiving a collaboration request, the system selects randomly one of the systems. If the selected system (System B) is uncapable of accepting a collaboration traffic, its **Collaboration Responder** will generate a **Reject** notification. Then it will send the same request (from System A) to the other system (System C) hosted in the Collaborative Cloud. In this case, System C is capable of accepting the collaborative traffic. The latter's **Collaboration Responder** will send an **Accept** notification to System A, then the traffic redirection from System A to System C will start.

Algorithm 3 Controller Algorithm

```
Begin
    Int LossRate, RTTState;
    LossRate = GetLossRate();
    RTTState = GetRTTState();
    // 1 is preferable, 2 is critical
    While(AvailableCollaborativeSystems() = = true)
      SelectNextCollaborativeSystem();
      If(CollaborationConfirmed() = = true)
        If(LossRate = = 2 || RTTState = = 2)
          ForwardToPlatform(25%)
          RedirectToCollCloud(75%);
        Else If(LossRate = = 1 &&RTTState = = 2
              || LossRate = = 2&&RTTState = = 1)
          ForwardToPlatform(50%)
          RedirectToCollCloud(50%);
      EndIf
    EndWhile// No System is available for collaboration
    If(AvailableCollaborativeSystems() = = false)
      Delay(50%, 2000 ms);
    Endif
  End
```

The controller algorithm (algorithm 3) will do the following: After receiving a confirmation of collaboration from another system *(System B or C)* in the Collaborative Cloud, the controller will forward the traffic to the OM2M platform or redirect it to the Collaborative Cloud based on the state of ***the system*** (System A*)*, (without knowing the state of the systems hosted in the Collaborative Cloud). This operation is conducted in order to maximize the number of satisfied requests while respecting the QoS metrics in terms of RTT (ITU-T G.1010 Recommendation). If the system (System A) receives a rejection for its collaboration request, the controller will look for another system to collaborate with. If no system is available for collaboration, half of this traffic will be delayed for 2 s.

Algorithm 3 presents the controller algorithm of the Collaboration Component (CC). After getting the loss rate (loss rate) and the RTT state (RTT state) of the system (System A), **(i) if** both metrics are at a critical state, 25% of the traffic will be processed in system's A OM2M platform, while 75% will be sent to the Collaborative Cloud. **(ii)** If one of the metrics is at a critical state and the other is at a preferable state, half of the traffic will be processed in system's A OM2M platform, while the other half will be redirected to the Collaborative Cloud.

This controller is activated only if one of the metrics is changed from preferable to critical. On the other hand, if no system responds with its ID as a collaboration response, half of the traffic will be delayed by 2 s.

5 Numerical Results

Pointing at the evaluation of our proposed solutions, we compared the number of satisfied requests of the overloaded system. QoS metrics in terms of RTT evolution of all systems' OM2M platform of the overloaded system are discussed later.

In the scenario testbed described in Tables 1, 2, and 3, we consider flows (http requests toward the OM2M platform) coming from different traffic sources simulated by injectors. Each injector is characterized by: the type of traffic, the number of http requests (requests number), the request method (e.g., POST, GET), the destination, the period between two successive requests (periodicity in milliseconds) and finally starting time in seconds.

In order to prove that collaboration can and will provide a solution for the scalability issue in IoT networks, we present our proposed scenario. Table 1 presents the flows injected in system A. Table 2 presents the flow injected in system B, while Table 3 presents the flow injected in system C, as systems B and C are hosted in the Collaborative Cloud (see Figs. 4 and 5).

From Table 1, we can state that System A is injected at start time 0 s of 1000 requests with periodicity of 100 ms. After 10 s, a second injector will overload the System A by injecting 600 requests with periodicity of 100 ms. System A will be overloaded after 10 s and generate a critical state symptom, which will lead to generate a collaborating request.

From Table 2, we can state that System B is injected at start time 0 s of 500 requests with periodicity of 200 ms. This system will be in the preferable state.

From Table 3, we can state that System C is injected at start time 0 s of 1000 requests with periodicity of 100 ms, and at the same time, a second injector will

Table 1 Scenario testbed of system A

Injector	Request number	Sequence (ms)	Start time (s)
Inj1	1000	100	0
Inj2	600	100	10

Table 2 Scenario 1 testbed of system B

Injector	Request number	Sequence (ms)	Start time (s)
Inj1	1000	100	0
Inj 2	400	150	0

Table 3 Scenario 2 testbed of system C

Injector	Request number	Sequence (ms)	Start time (s)
Inj1	500	200	0

Fig. 5 Satisfied requests of
system A
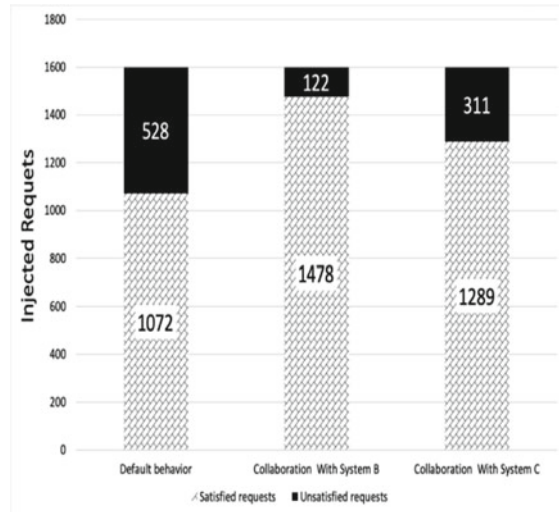
inject 400 requests with a periodicity of 150 ms. This system will be on the edge of
critical state.

5.1 Success Rate

In Figs. 5, 6, and 7, the columns indicate the number of injected requests. We use
different colors to distinguish the number of satisfied and lost requests.

Figure 5 presents the satisfied requests of system A. As shown in Table 1, System
A is injected by 1600 requests and it is generating a critical loss rate symptom. In

Fig. 6 Satisfied requests of
system B
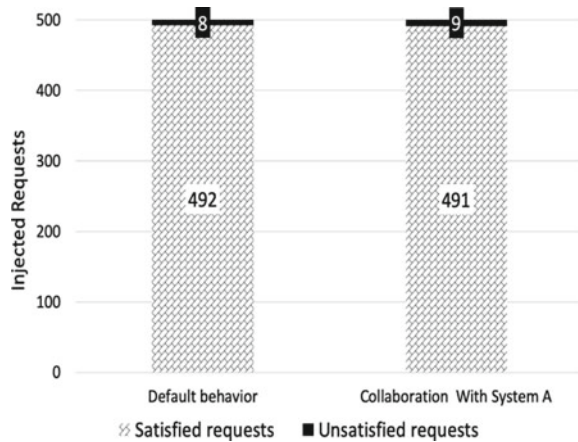
Fig. 7 Satisfied requests of system C
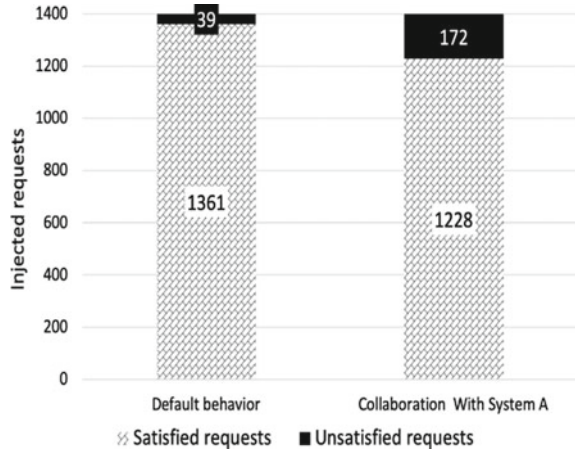
the system's normal behavior, without using our **IoTScal-SC2** solution, we can see that the success rate is at 67% which is below the SLA requirements for high priority traffic mentioned before.

The proposed **IoTScal-SC2** solution mechanisms will activate collaboration with system B or C randomly. If the collaboration is activated with System B which is at preferable state, the success rate will rise to 92%, respecting the SLA requirements for high priority traffic. If the collaboration is activated with System B which is at the edge of critical state, the success rate will go to 81%, improving success rate compared to the normal behavior, but not enough to respect the SLA requirements for high priority traffic.

Figure 6 presents System B success rate, as mentioned in Table 2, system B is at preferable state, so when a collaboration is activated, its success rate stays the same at 98%, respecting the SLA requirements for high priority traffic.

Figure 7 presents System C success rate, as mentioned in Table 3, system B is at the edge of critical state, so when a collaboration is activated, its success rate drops from 97 to 88%, not respecting the SLA requirements for high priority traffic.

We state that our IoTScal-Even Collaboration solution satisfies more traffic requests in both systems for the following reasons:

The decisions based on the system performances help satisfy more requests than its normal behavior. A collaboration with a system at preferable state improved the success rate to a level that respects the SLA requirements for high priority traffic. On the other hand, a collaboration with a system at critical state improved the success rate, but not enough to respect the SLA requirements for high priority traffic.

Overall, our IoT system scalability has been improved using our collaboration solution but not enough if the collaborative system is also overloaded.

5.2 RTT Evolution

In Figs. 8, 9, and 10, the column indicates the number of injected requests. We use different colors to distinguish the RTT evolution in each case. We used the same scale in RTT evolution to show the difference between each system's RTT.

Figure 8 can be described as follows:
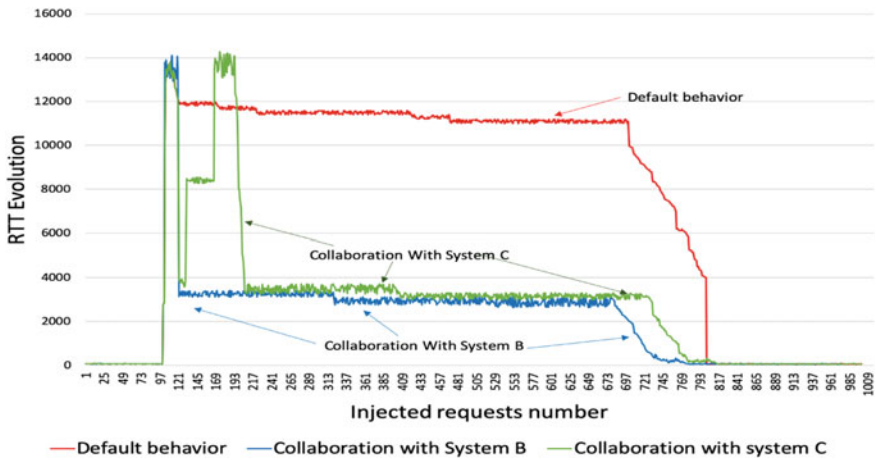
Fig. 8 RTT evolution for system A
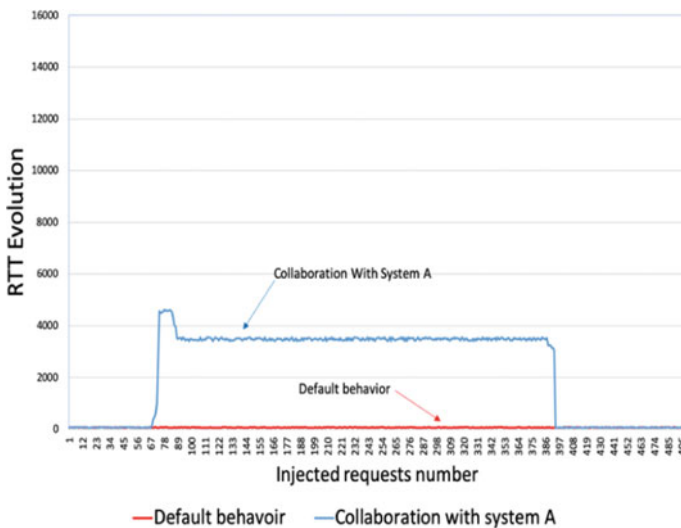
Fig. 9 RTT evolution for system B
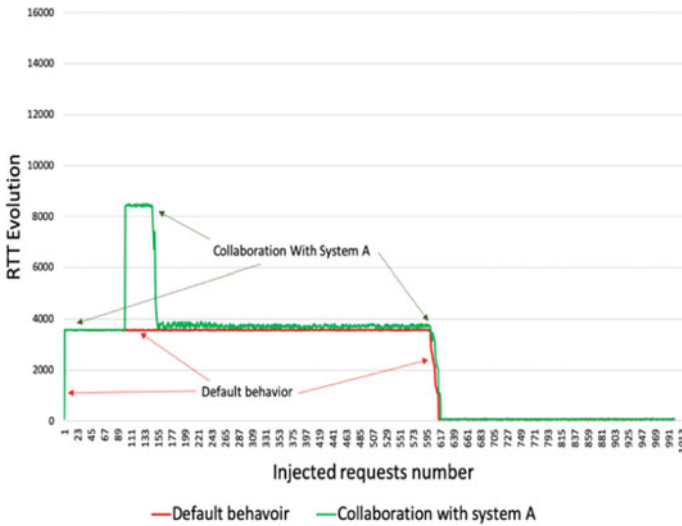
Fig. 10 RTT evolution for system C

In default behavior, after 10 s from the first burst of traffic, the second one will overload the system, which will raise the RTT to over 13,000 ms (13 s), the system can no longer support more traffic due to the huge number of requests received by the system. After that, the OM2M platform starts losing requests which will decrease the RTT to over 11,000 ms (11 s), until after the second traffic stops, then the RTT will drop to under 4 s, stabilizing back the system A.

In this case, neither the loss rate nor the RTT evolution are respecting the recommendation for urgent traffic (loss rate < 10%, RTT < 4000 ms). We proposed the IoTScal-Collaboration solution in order to satisfy more requests, all by respecting QoS metrics (RTT).

The IoTScal-SC2 solution, in case of a detected overload, the system will randomly select a Collaborative Cloud system and send the traffic to it. As mentioned before, there are two systems, one is at preferable state and the other is at the edge of critical state, which leads us to one of two scenarios.

The first scenario is that the selected Collaborative Cloud system is the system B that has a preferable state. In this case, Fig. 8 shows that after detecting an overload, and by activating our Even Collaborative solution, the system will redirect 75% of the traffic (algorithm 3) to System B. This decision drops immediately the RTT evolution from over 13,000 ms (13 s) to under 4000 ms (4 s). This decision will help system A to scale, satisfying 92% of the requests and keeping RTT below 4000 ms (4 s), respecting the recommendation for loss rate and RTT evolution.

The second scenario is that the selected Collaborative Cloud system is the system C that has a near critical state. In this case, Fig. 6 shows that after detecting an overload, and by activating the proposed mechanisms, the system will redirect 75% of the traffic (algorithm 3) to System C. This decision drops immediately the RTT

evolution from over 13,000 ms (13 s) to under 4000 ms (4 s) then rises up to 8000 ms (8 s) goes back to 13000 ms (13 s) then drops back to 4000 ms (4 s). This scenario includes several decisions.

After detecting an overload, the system redirected 75% of the traffic to System C, the RTT goes up to over 13,000 ms (13 s). System C starts processing the received requests, RTT goes to under 4000 ms (4 s). Then, System C got overloaded and RTT went up to over 8000 ms. System C will reject collaboration traffic as mentioned before, in order to maintain its performance at an acceptable state. System A goes back up to over 13,000 ms (13 s) and will select another Collaborative Cloud system to send 75% of the traffic to it (algorithm 3). The new selected system is at preferable state, the RTT evolution will drop down to under 4000 ms (4 s).

The second scenario helps system A scale, satisfying 81% and keeping RTT under 4000 ms (4 s). The loss rate is ameliorated but not enough to respect recommendation, but RTT evolution is respecting recommendation.

Generally, we can state that our *IoTScal-SC2* solution improved Scalability keeping QoS metrics at acceptable levels.

Figure 9 describes RTT evolution for system B. System B is at a preferable state, so it maintains its RTT evolution under 100 ms in default behavior. In case of collaboration with System A, where 75% of the traffic will be sent to system B for collaboration operation, the RTT will go over 4000 ms (4 s) for several instances and then goes back to under 100 ms. System B will keep its success rate at the same level, processing 98% of incoming traffic. In addition, the solution helped stabilize "System A" 10 s quicker than the default behavior.

System B keeps its state at preferable level, helping Overloaded System A scale. Both systems are respecting QoS metrics (RTT).

Figure 10 describes RTT evolution for system C. System C is at the edge of critical state. In case of collaboration with System A, where 75% of the traffic will be sent to system B for collaboration operation, the RTT will go from under 4000 ms (4 s) to over 8000 ms (8 s). Then the system will detect an overload while receiving collaborating traffic and it rejects all collaboration traffic in order to go back to its acceptable state. The RTT evolution of System C goes back from over 8000 ms (8 s) to under 4000 ms (4 s). The success rate of System C drops from 97 to 88% and will no longer respect the recommendation of loss rate. In addition, the solution helped stabilize System A 7.7 s quicker than default behavior.

Figure 10 describes RTT evolution for system C. System C is at the edge of critical state. In case of collaboration with System A, where 75% of the traffic will be sent to system B for collaboration operation, the RTT will go from under 4000 ms (4 s) to over 8000 ms (8 s). Then the system will detect an overload while receiving collaborating traffic and it rejects all collaboration traffic in order to go back to its acceptable state. The RTT evolution of System C goes back from over 8000 ms (8 s) to under 4000 ms (4 s). The success rate of System C drops from 97 to 88% and will no longer respect the recommendation of loss rate. In addition, the solution helped stabilize System A 7.7 s quicker than default behavior.

In this scenario, both systems are no longer respecting loss rate recommendation, but respecting RTT evolution recommendation.

6 Conclusion

We began by debating our vision of scalability in IoT systems in this paper. Following that, we proposed improving our system by adding components and testing it, with the goal of stressing the OM2M platform. The proposed approach is an architecture based on Cloud Computing using collaboration with autonomic scalability traffic-oriented mechanisms. The present IoTScal-SC2 collaboration architecture is presented to improve the IoT system scalability without any QoS symptom degradation.

Simulation results show that our mechanism does not only help improve the success rate (reducing the loss rate), but also stabilizes the IoT system almost instantly, maintaining QoS metrics in terms of RTT evolution by improving it, and without overloading the collaborative system.

The new proposed mechanisms outperform the other mechanisms freshly proposed in IoT system. More complex scenarios can be implemented. Our future work and experiments will focus on the use of a bigger number of injectors, with different SLA requirements, and much more complex decisions.

References

1. Huang PK, Qi E, Park M, Stephens A (2013) Energy efficient and scalable device-to-device discovery protocol with fast discovery. In: IEEE International workshop of internet-of things networking and control (IoT-NC), vol 13(3). pp 1–9
2. Pandey S, Voorsluys W, Niu S, Khandoker A, Buyya R (2012) An autonomic cloud environment for hosting ECG data analysis services. Future Gener Comput Syst 147–154
3. Bahiri MN, Zyane A, Ghammaz A (2018) An enhancement for the autonomic middleware-level scalability management within IoT system using cloud computing, In: Lecture notes in electrical engineering book series (LNEE)–(ICEERE2018), vol 519. Springer, ISBN: 978-981-13-1405-6, pp 80–88
4. Bahiri MN, Zyane A, Ghammaz A, Chassot C (2017) A new monitoring approach with cloud computing for autonomic middleware-level scalability management within IoT systems. In: Advances in intelligent systems and computing book series (AISC)–(ITCS2017), vol 640. Springer, ISBN: 978-3-319-64719-7, pp. 281–296
5. Scalable SQL (2011) Communications of the ACM, 48–53
6. McCabe L, Aggarwal S (2012) La migration vers le Cloud pour les PME, SMB Group, Inc, October 2012
7. Ramasahayam R, Deters R (2011) Is the cloud the answer to scalability of ecologies ?. In: 5th IEEE international conference on digital ecosystems and technologies, pp 317–323
8. Bahiri MN, Zyane A, Ghammaz A (2021) IoTScal-C: a based cloud computing collaboration solution for scalability issue in IoT networks. In: Proceedings of the 2nd international conference on electronic engineering and renewable energy systems—(ICEERE2020), Springer, ISBN: 978-981-15-6259-4, pp 123–133
9. Horn P (2005) An architectural blueprint for autonomic computing, IBM White Paper ed. 3

An Effective Ensemble Learning Method for Fault Diagnosis of Photovoltaic Arrays

Adel Mellit and Sahbi Boubaker

Abstract In this paper, a novel fault diagnosis method for photovoltaic (PV) arrays is proposed. The method combines three machine learning (ML) algorithms: the first one is an unsupervised ML algorithm (principal component analysis, 'PCA') used for features reduction; the second one is a kind of a recurrent neural networks (long short-term memory, 'LSTM') employed to predict the PV power; and the third one is a supervised ensemble learning categorical boosting algorithm (CatBoost), used mainly for fault classification improvement. Four types of faults (dust deposit on PV modules surface, partial shading effect, shunted diode and open-circuit diode) are examined in this study. Results showed the effectiveness of the proposed method in terms of detection and classification accuracy, 98.70% and 98.58% respectively, and demonstrate the superiority of this method over other investigated ML algorithms (e.g. decision tree, support vector machine, *k*-nearest neighbours, and random forest).

Keywords Photovoltaic array · Fault diagnosis · Machine learning · Ensemble learning · Multi-classification

1 Introduction

With reference to the International Energy Agency (IEA), global photovoltaic (PV) capacity installations around the world reached 760 GWp, at the end of 2020 [1]. A huge number of PV plants were installed worldwide, and these plants need to be monitored and supervised carefully [2]. In addition, to avoid power losses and increase the PV plants' reliability, PV plants should be equipped with fault detection and diagnosis systems. Recently, a large number of methods based on machine

A. Mellit (✉)
Department of Electronics, Faculty of Science and Technology, University of Jijel, Jijel, Algeria
e-mail: adel_mellit@univ-jijel.dz

S. Boubaker
Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia
e-mail: sboubaker@uj.edu.sa

learning (ML) algorithms were introduced [3–5] for fault detection and classification of PV systems. The classification accuracy in both detection (binary classification) and classification (multi-classification) depends on many factors, like data quality, (including data preparation, features extraction and important features selection), and the effectiveness of the classification algorithms. Most used supervised ML-based algorithms in this topic are: support vector machine (SVM), k-nearest neighbours (k-NN), decision tree (DT), random forest (RF), neural networks (NNs) and other boosting algorithms [6, 7]. Ensemble learning (EL) algorithms like categorical boosting (CatBoost), extreme gradient boosting (XGBoost) and light gradient boosting machine (LightGBM) algorithms are recently introduced and showed their performance in solving regression and classification problems [8].

To the best of the authors' knowledge, these kinds of EL-based algorithms (CatBoost, XGBoost and LightGBM) are the first time used in fault diagnosis of PV systems based on I–V curves. Thus, the main contribution is to combine one EL algorithm and two ML algorithms. Principal component analysis (PCA) technique is used for features reduction, CatBoost algorithm for the multi-classification purpose and one recurrent neural network (long short-term memory, 'LSTM') for fault detection in PV arrays. The above algorithms are selected due to their superiority over classical ML-based algorithms.

The paper is organized as follows: Materials and methodology are presented in Sect. 2. Section 3 presents and discusses the obtained results. Concluding remarks and future directions are reported in the last Sect. 4.

2 Materials and Methodology

2.1 Datasets Preparation

The datasets used in this study were collected at the Renewable Energy Laboratory, the University of Jijel (Algeria). The first dataset includes 800 I–V curves at different working conditions (normal and abnormal). The second contains measured photovoltaic power (5100 samples) for a time horizon of 1 min, as well the corresponding solar irradiance (G) and air temperature (T). The investigated PV string consists of four modules connected in parallel [9]. The considered faults are respectively, shading effect ($F1$), dust deposit on PV modules ($F2$), open-circuit diode in a PV module ($F3$) and shunted diode in a PV module ($F4$). As example, Fig. 1a shows the collected I–V curves (normal and abnormal) under working conditions. Figure 1b depicts the measured PV power (P) for a short period of one day covering a working time between 7:15 a.m. and 17:30 p.m.
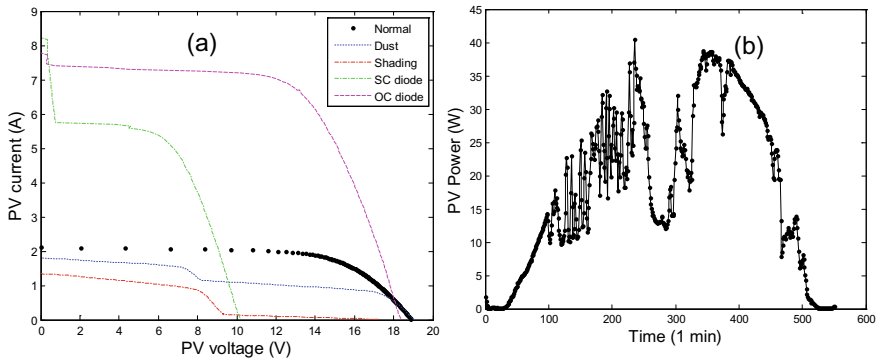
Fig. 1 **a** Measured I–V curves (normal and abnormal), **b** the collected PV power for a period of one day (21st November 2021) for a cloudy day with daily average solar irradiance of 250 W/m^2 and daily average air temperature of 17 °C

2.2 Machine Learning Algorithms (CatBoost, LSTM and PCA)

A brief overview of CatBoost, LSTM and PCA algorithms is given in the following subsections. Note that this paper is not intended to present mathematical development of these ML-based algorithms because of space limitation.

- **CatBoost**

CatBoost was firstly introduced by Yandex in [10], which is an open-sourced ML algorithm for gradient boosting on decision trees. It belongs to ensemble learning-based algorithms and its main features are: (1) Great quality without parameter tuning, (2) categorical features support and (3) improved accuracy with fast prediction.

- **Long-Short-term memory (LSTM)**

LSTM is an advanced configuration of recurrent neural networks (RNNs) able of learning order dependence in time series prediction problems. It was firstly proposed by Hochreiter and Schmidhuber in [11]. LSTM network outperforms other artificial NNs like backpropagation through time, Elmen networks and neural sequence chunking [11]. The LSTM is capable of solving the vanishing gradient problem addressed by traditional RNN.

- **Principal component analysis (PCA)**

PCA is an unsupervised ML-based algorithm, mainly used for dimensionality reduction. In PCA, similar data based on the feature correlation between them without any supervision can be clustered. The main task in the PCA is to select a subset of variables from a larger set, based on which original variables have the highest correlation with the principal amount [12].

2.3 The Proposed Method

The proposed method is presented in the following flowchart (See Fig. 2).

Feature extraction and reduction procedure

After collecting the *I–V* curves, the next step consists to extract different features (I_{mp}, I_{sc}, V_{mp}, V_{oc}, P_{mp}, FF, V_1 and I_1) from the *I–V* curves based on a basic algorithm.

The above-mentioned variables are defined as follows: I_{mp} is the current at maximum power point (MPP), V_{mp} is the voltage at MPP, I_{sc} is the short-circuit current ($V = 0$), V_{oc} is open voltage ($I = 0$), and I_1 and V_1 are the calculated current and voltage ($I_1 = f(\text{V} \times 2/3)$, $V_1 = f(\text{I} \times 2/3)$). In this context, the PCA technique is used for features dimensionally reduction.

Fault detection procedure

To detect the fault, an LSTM network is developed based on historical values of measured PV output power (P^*_t), solar irradiance (G_t) and air temperature (T_t). The model aims to estimate the PV power P^*_t and then compare it with the measured (P_t). The model uses the actual values of G_t and T_t, and the output contains the predicted P^*_t at time t as inputs. A threshold (μ) is estimated based on extensive experiments. If ($\Delta P_t = P_t - P^*_t \cong \mu$) is close to the reference threshold, μ; there is no fault; otherwise, a fault is detected.

Fault classification procedure

Once the fault is detected, the next step aims to identify the type of fault based on a multi-classification algorithm. Here, the CatBoost classifier is used to address this issue. The inputs of the CatBoost-based classifier are the selected features, while the outputs are the labelled faults as: class 1 (*F*1), class 2 (*F*2), class 3 (*F*3) and
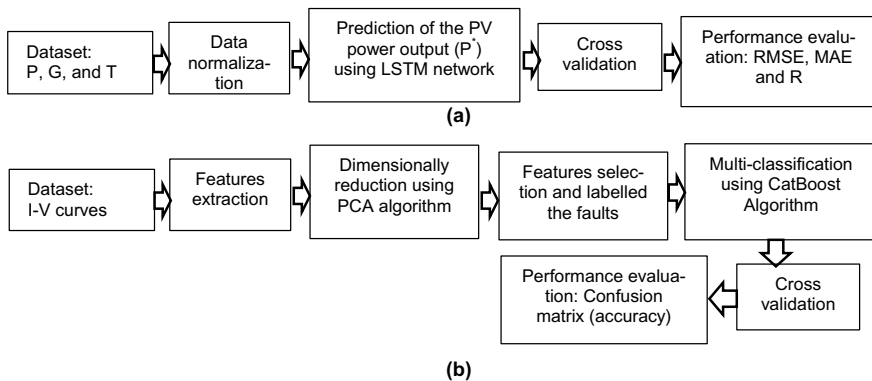
Fig. 2 The proposed method for fault detection and classification with feature selection technique: **a** detection procedure and **b** classification procedure with features reduction

class 4 ($F4$). A set of 80% of observations is used to train the classifier, while the remaining set (20%) is used for testing the classifier. To resample the dataset without replacement, a K-fold cross-validation technique is applied (K is set to 10).

3 Results and Discussion

To evaluate the performance of the developed method, error metrics such as root mean squared error (RMSE), mean absolute error (MAE), correlation coefficient (R), $F1$-score, precision, recall and accuracy are calculated. To obtain a more reliable estimation of the models performance K-fold cross-validation technique is used. The following pseudocode shows how the method operates.

Step 1: Acquisition of measured $I–V$, G and T
Step 2: Estimate the predicted power (P^*) based on the LSTM model: $P^* = f_{LSTM}$ (G, T)
Step 3: **if** $P–P^*>\mu$ **then** move to next step
Step 4: Extract features from the $I–V$ curve
Step 5: Select the important features using PCA algorithm (I_{mp}, V_{mp}, I_{sc}, V_{oc} and FF)
Step 6: Classify the fault based on the CatBoost algorithm: Fault=$f_{CatBoost}(I_{mp}$, V_{mp}, I_{sc}, V_{oc} and FF)
Step 7: Display the type of fault and return to the first step

3.1 Features Selection and Reduction

The features have been take out from the $I–V$ characteristics. Table 1 shows an example of the extracted features with the corresponding G and T. To reduce the number of features, we set the principal component (PC), of the PCA algorithm, to 5 and the results are reported in Fig. 3. The important features in this dataset are: V_{mp}, V_{oc}, P_{mp}, I_{sc} and FF.

3.2 Fault Detection

Figure 4a shows the loss function of the developed LSTM predictor, the MSE is close to 0.0006, which indicates the ability of the model to predict the power with good accuracy. Figure 4b depicts a comparison between measured and predicted PV power. Good correlation is observed between measured and predicted power as

Table 1 Extracted features from the I–V curves

G (W/m²)	T (°C)	I_{sc} (A)	V_{oc} (V)	I_{mp} (A)	V_{mp} (V)	P_{mp} (W)	FF	I_1 (A)	V_1 (V)
215	19	4.13	19.11	3.67	13.66	50.54	0.64	2.06	16.91
250	20	4.85	18.81	4.35	13.36	58.16	0.63	2.42	16.61
300	21	6.60	18.61	5.44	13.16	71.61	0.63	3.03	16.41
310	25	6.54	18.41	5.87	12.96	76.13	0.63	3.27	16.21
315	22	7.27	18.21	6.39	13.03	83.34	0.62	3.63	16.01
320	25	7.75	18.01	6.82	12.83	87.53	0.63	3.88	15.81
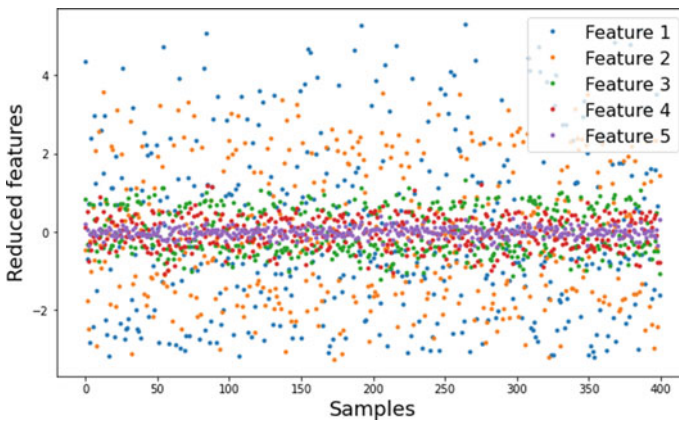
Fig. 3 Reduced features using PCA algorithm (example of PC $= 5$)

shown in Fig. 4c. With reference to Table 2, the correlation coefficient is 98.70%, the RMSE and the MAE are very low (less than 0.02 W). These results demonstrate the ability of the LSTM algorithm to estimate the PV power based on solar irradiance and air temperature.

3.3 Fault Classification

Classification procedure aims to classify the four faults ($F1$, $F2$, $F3$ and $F4$) afore-mentioned. The calculated recall, precision, accuracy and $F1$ are reported in Table 3. The precision, recall and $F1$-score are ranged between 95 and 100%, and the accuracy is 98.58%.
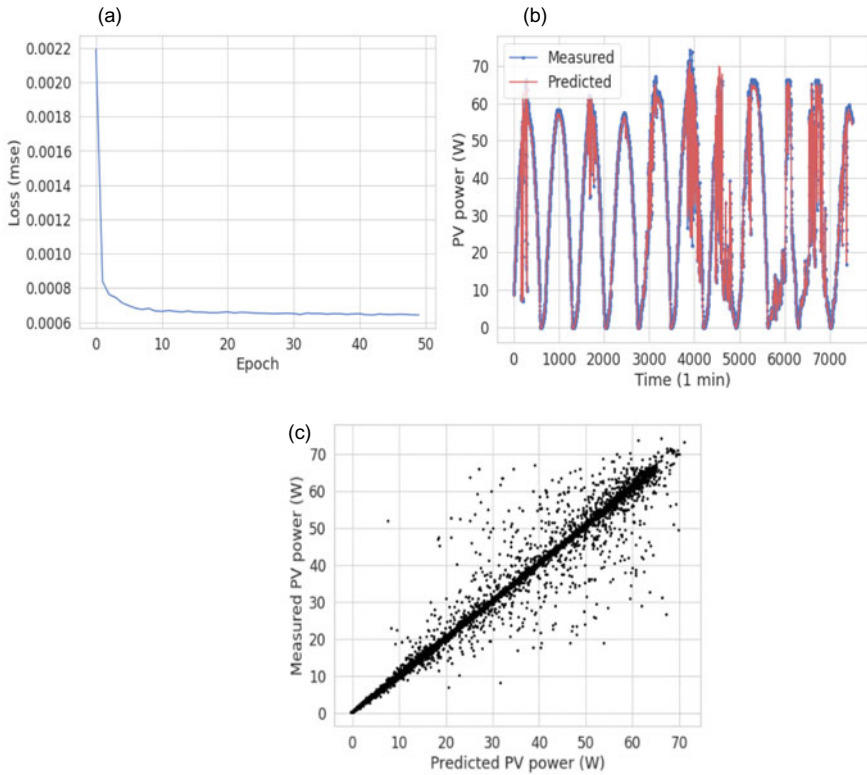
Fig. 4 a Loss function evolution, **b** measured versus predicted PV power and **c** scatted plot of measured and predicted PV power

Table 2 Error metrics (MAE, RMSE and R) of the LSTM model used for fault detection

LSTM model	MAE (W)	RMSE (W)	R (%)
Epoch $= 20$, Batch-size $= 32$, Optimiser $=$ Adam, Loss $=$ mse, Number of unit $= 100$	0.018	0.045	98.70

Table 3 Calculated error metrics of the fault classification method (CatBoost classifier)

CatBoost classifier (iteration $=$ 5000, learning_rate $= 0.10$)	Precision (%)	Recall (%)	$F1$-score (%)	Accuracy (%)
$F1$ (class #1)	95	100	98	98.58
$F2$ (class #2)	100	97	100	–
$F3$ (class #3)	100	98	96	–
$F4$ (class #4)	99	100	100	–

Table 4 Accuracy of various classification models (DT, SVM, k-NN, RF and CatBoost)

Classifier	Accuracy (%)
DT (nmax_depth = 5, random_state = 42)	95.68
SVM (kernel = 'linear', C = 1E10)	97.80
k-NN (n_neighbours = 30)	96.37
RF (n_estimators = 300, max_depth = 5)	98.00
CatBoost (iteration = 5000, learning_rate = 0.10)	98.58

3.4 Comparative Study

In order to show the efficacy of the proposed method in terms of classification accuracy, we made a comparative study with other ML-based algorithms (e.g. DT, SVM, k-NN, and RF). Table 4 depicts the obtained results and from this table, it can be seen that CatBoost outperforms other examined ML algorithms and achieve good accuracy. The lower accuracy is obtained by the DT algorithm, SVM and k-NN have relatively close accuracy and RF algorithm presents a closer result to CatBoost algorithm. It should be pointed that these algorithms are not optimized, and the results could be improved by tuning their hyper-parameters.

4 Conclusion and Future Directions

A method for fault diagnosis of PV systems is introduced. The accuracy of the fault detection method depends on the dataset, which should be updated periodically to avoid false alarm (due to the ageing of PV modules over time). The classification accuracy could be improved by using a large dataset, good quality and an appropriate features selection. It should be emphasized that the classified faults, investigated in this work, are not very complex (i.e. only single faults are considered in this work), but in some situation dust distribution on the PV surface has similar I–V curve as partial shading, which makes the classification procedure difficult (i.e. to distinguish between both faults).

The method could be further improved by optimizing the hyper-parameters and verifying it experimentally to show its capability for real-world applications. Our future works will devote to the design of an embedded system for fault diagnosis of PV systems, which is still an open challenge in this field.

References

1. Snapshot of Global PV Markets (2021) Report IEA-PVPS T1–39. 2021. Available online: https://iea-pvps.org/snapshot-reports/snapshot-2021/ (Accessed 25 Apr 2021)
2. Mellit A, Kalogirou S (2021) Artificial intelligence and internet of things to improve efficacy of diagnosis and remote sensing of solar photovoltaic systems: challenges, recommendations and future directions. Renew Sustain Energy Rev 143:110889
3. Kapucu C, Cubukcu M (2021) A supervised ensemble learning method for fault diagnosis in photovoltaic strings. Energy 227:120463
4. Tina GM, Ventura C, Ferlito S, De Vito S (2021) A state-of-art-review on machine-learning based methods for PV. Appl Sci 11(16):7550
5. Li B, Delpha C, Diallo D, Migan-Dubois A (2021) Application of artificial neural networks to photovoltaic fault detection and diagnosis: a review. Renew Sustain Energy Rev 138:110512
6. Mandal RK, Kale PG (2021) Assessment of different multiclass SVM strategies for fault classification in a PV system. In: Proceedings of the 7th international conference on advances in energy research, pp 47–756
7. Mellit A, Kalogirou S (2022) Assessment of machine learning and ensemble methods for fault diagnosis of photovoltaic systems. Renew Energy 184:1074–1090
8. Anghel A, Papandreou N, Parnell T, De Palma A, Pozidis H (2018) Benchmarking and optimization of gradient boosting decision tree algorithms. arXiv preprint arXiv:1809.04559
9. Mellit A, Herrak O, Rus Casas C, Massi Pavan A (2021) a machine learning and internet of things-based online fault diagnosis method for photovoltaic arrays. Sustainability 13(23):13203
10. CatBoost. https://catboost.ai/. Assessed 18 July 2017]
11. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
12. Subramanian SK (2011) Hybrid methods in feature selection: a data classification perspective: hybrid feature selection methods are the proven methods for large scale feature selection. LAP LAMBERT Academic Publishing

A Lightweight Deep Learning Model for Forest Fires Detection and Monitoring

Mimoun Yandouzi, Idriss Idrissi, Mohammed Boukabous, Mounir Grari, Omar Moussaoui, Mostafa Azizi, Kamal Ghoumid, and Aissa Kerkour Elmiad

Abstract Forests play a critical role in our everyday lives, they provide us with fresh air and clean water and are home to a variety of plants and animals. Forests are also an important source of economic security, jobs, and recreation. However, fires are one of the most disastrous natural catastrophes that may strike these forests; they can swiftly burn significant tracts of forests and are extremely difficult to manage. People and animals can be seriously injured, killed, or even constrained to move far from forest fires, homes and properties can be destroyed. Fires move swiftly across the forest, burning everything in their path. In this paper, we proposed a lightweight deep learning (DL) convolutional neural network (CNN) model (with around one million parameters) for wildfires' aerial reconnaissance, intended to be deployed on unmanned aerial vehicle (UAV) or drones. The obtained results are promising in multiple metrics such as accuracy, recall, and precision (with more than 98%), and a loss of 9% demonstrating that using drones can lead firefighting efforts by allowing firefighters to detect and monitor wildfires.

M. Yandouzi (✉) · K. Ghoumid · A. Kerkour Elmiad
Lab. LARI, FSO, Mohammed First University, Oujda, Morocco
e-mail: m.yandouzi@ump.ac.ma

K. Ghoumid
e-mail: k.ghoumid@ump.ac.ma

I. Idrissi · M. Boukabous · M. Grari · O. Moussaoui · M. Azizi
Lab. MATSI, ESTO, Mohammed First University, Oujda, Morocco
e-mail: idrissi@ump.ac.ma

M. Boukabous
e-mail: m.boukabous@ump.ca.ma

M. Grari
e-mail: m.grari@ump.ac.ma

O. Moussaoui
e-mail: o.moussaoui@ump.ac.ma

M. Azizi
e-mail: azizi.mos@ump.ac.ma

697

1 Introduction

Forests are one of the most precious natural resources on Earth, and they are vitally important for both the health of the planet and the people who live on it. Forests cover about one-third of the Earth's land area, and they play a critical role in the global ecosystem. They are homes to millions of plant and animal species, and they help regulate the Earth's climate. They also provide many economic benefits to people around the world.

Forest fires are one of the most devastating disasters that can occur in a forest. They can quickly destroy large areas of forest and can be very difficult to control. Forest fires can also cause serious injuries or death to people and animals and can damage homes and properties [1].

Several things can cause a forest fire. Most forest fires are started by people, often inadvertently, such as by campers or hunters who fail to properly extinguish their campfires or dispose of their cigarettes. Lightning strikes can also start forest fires. In some cases, forest fires can be started by natural causes, such as a spark from a volcanic eruption or a wildfire that has jumped from another part of the forest [2].

No matter what the cause, a forest fire can be very dangerous. The flames can quickly spread through the forest, destroying everything in their path. The heat from the fire can also damage trees and other plants and cause serious burns to people and animals. The smoke and air pollution from a forest fire can also be a health hazard.

Forest fires detection techniques are used to identify and map wildfires to provide early warning to firefighters and the public. These techniques include ground and aerial reconnaissance, remote sensing, and satellite imagery. In this paper, we developed a lightweight deep learning CNN model for wildfires aerial reconnaissance, intended to be deployed on a UAV or a drone.

The rest of this paper is organized as follows: the second section gives some backgrounds on the issue of detecting forest fires and machine learning. The third section provides an overview of the related works, while the fourth section presents our proposed methods. Before concluding, we discuss our study's findings in the fifth section.

2 Backgrounds

2.1 Forest Fires Detection Techniques

Forest fires detection techniques allow to identify and map wildfires to provide early warning to firefighters and the public. Detection techniques include ground and aerial reconnaissance, remote sensing, and satellite imagery.

- Ground reconnaissance involves human observers walking through forests to look for fires. This is a time-consuming and labor-intensive process, but it can be effective in detecting small fires.
- Aerial reconnaissance is a more efficient way to detect fires and involves using aircraft and drones to fly over forests and look for signs of fire. Aerial reconnaissance can map large areas quickly and can to detect fires that are not visible from the ground.
- Remote sensing is a method of detecting fires using sensors that are not in contact with the ground. Remote sensing can help to detect fires in areas with difficult access and can be used to identify fires that are hidden by smoke and haze.
- Satellite imagery is a method of detecting fires using data collected from satellites orbiting Earth. Satellite imagery allows to map large areas quickly and identify fires that are not visible from the ground.

2.2 Computer Vision (CV) and Deep Learning (DL)

Computer vision is a field of study that deals with the algorithms that help computers "see." The problem is vastly different from the one solved by cameras and the human visual system, yet the goal is the same: extract from an image a representation that can be used to recognize objects and scenes, and to understand the contents of an image [3].

It is a field of study that is currently being revolutionized by the application of deep learning algorithms [4]. These algorithms can learn representations of data that are much more complex than can be captured by handcrafted features. This is allowing computer vision systems to achieve state-of-the-art results on a wide range of tasks, including object detection, scene recognition, and tracking [5]. Deep learning has emerged as a powerful tool for computer vision. In particular, convolutional neural networks (CNNs), which are composed of a large number of layers of neurons, have been shown to be very effective in modeling complex data structures [6].

A CNN is a type of neural network, specifically designed to process input data ranged in a grid, such as an image. The layers of neurons in a CNN are organized into a stack, each layer is dedicated to processing a small region of the input grid [7]. The input of a CNN is typically a matrix of numbers, where each number corresponds to a pixel in the image (see Fig. 1). A CNN typically contains a large number of layers,
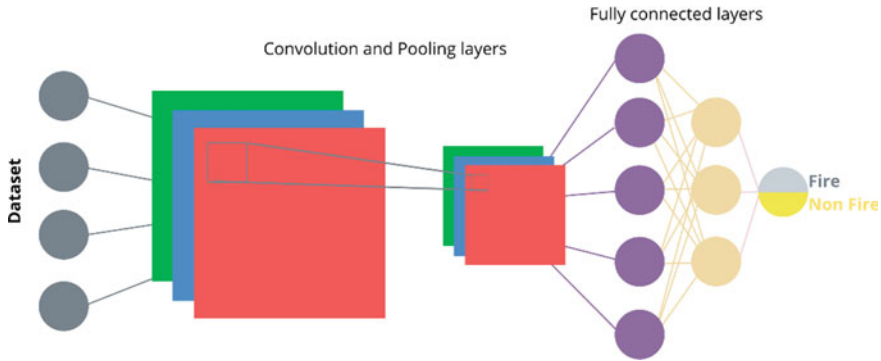
Fig. 1 CNN layers architecture

and each layer is composed of many neurons. The number of layers and the number of neurons in each layer can be tuned to achieve the desired level of accuracy [8].

CNNs have been found to be very effective in computer vision applications. In particular, they have been shown to be very effective in recognizing objects and scenes in images. They have also been shown to be effective in identifying the contents of an image [9]. CNNs have been found to be effective in identifying the contents of an image. In this application, the CNN is trained to identify wildfire in images.

3   Related Works

Zheng et al. [10] proposed a forest fire risk monitoring system that includes the hardware consisting of a UAV and an image acquisition system, as well as the corresponding software. In forest fire recognition, after image pre-processing, region segmentation, and feature extraction, different classifiers fit to recognize the fire risk images. The comparison with the general algorithm shows that their proposed algorithm can recognize the forest fire risk with better accuracy while meeting the requirements of real-time data processing of scene recognition. The dataset used is composed of images collected from a drone during a fire simulation. They obtained an accuracy of 81.97%.

Yanık et al. [11] introduced a new low-cost drone equipped with image processing and object detection abilities for smoke and fire recognition tasks in forests. For this purpose, they used a drone equipped with a Raspberry that has been programmed by training with deep learning methods based on the lightweight DL MobileNet. For real-time object recognition, the speed of the model is the most important parameter. In this work, the dataset used for pre-training is COCO, and their proposed model "ssdlite MobileNet" is tested on four parameter variants related to the number of images in the training and testing phases. The study also addresses the issue of battery consumption in order to increase the number of flight hours of the UAV.

Khennou et al. [12] proposed a model "FU-Netcast" to predict the spread of forest fires. Their proposed model is able to predict the geographical area that will be burnt in the next 24 h. The model is based on Unet and uses Landsat images from bands 2 to 7, digital elevation model, aspect, slope, and weather data. They obtained an accuracy of 92.73% and an AUC of 80%. The work is well suited to help firefighters be more accurate, locate future burn areas, and act quickly to stop the spread of fire.

Mohammed et al. [13] showed a deep learning technology using transfer learning to extract forest fire and smoke characteristics. The DL model used is Inception-ResNet-v2 pre-trained on the ImageNet dataset to be trained on a collected dataset of 1102 images for each fire and smoke class. The accuracy, precision, recall, $F1$ score, and specificity of the classification were 99.09%, 100%, 98.08%, 99.09%, and 98.30%, respectively. They deployed this model on a Raspberry Pi device equipped with a camera.

4 Our Proposed Method

The proposed solution is to use drones and deep learning algorithms to detect wildfires before they become out of control. Drones can fly over forests and identify hot spots that could spark a wildfire, once a hot spot has been identified, the deep learning algorithm can be used to determine whether it is a wildfire. This scheme has several advantages over traditional forest wildfire detection methods. For starters, drones can fly over large areas much faster than ground crews. Second, the deep learning algorithm can identify wildfires much more accurately than human observers can. Third, the use of drones and deep learning algorithms can aid in the safety of firefighters and other emergency personnel by alerting them to wildfires before they become a problem.

Our proposed model is a CNN model trained using a dataset that contains a huge number of photographs of past wildfires recorded in various areas across the world, as well as images of forest landscapes without fire, in different environments and types of terrain, including forests, hills, mountains, and grasslands. It was created by combining and integrating numerous smaller datasets from search engines and Kaggle [14], producing 2525 photos labeled "no fire" and 2136 images labeled "fire" after cleaning some corrupted images. Afterward, we used data augmentation to considerably expand the amount of our training dataset, and as a result, the quality of the trained models [15]. Through data augmentation, we added additional data to the dataset that is close to the original data with some minor changes, which can increase the performance and accuracy of our model from data [16].

The first layer in our CNN model is called the input layer with the size of 224 × 224, and it accepts the input matrix and passes it on to the next layer in the stack. The next layers are a stack of convolutional and max-pooling layers (3 stacks). The convolutional layer is responsible for computing the features of the image; these features are a set of numbers that describe the image in some way, for example, the features might capture the intensity, color, and location of each pixel in the
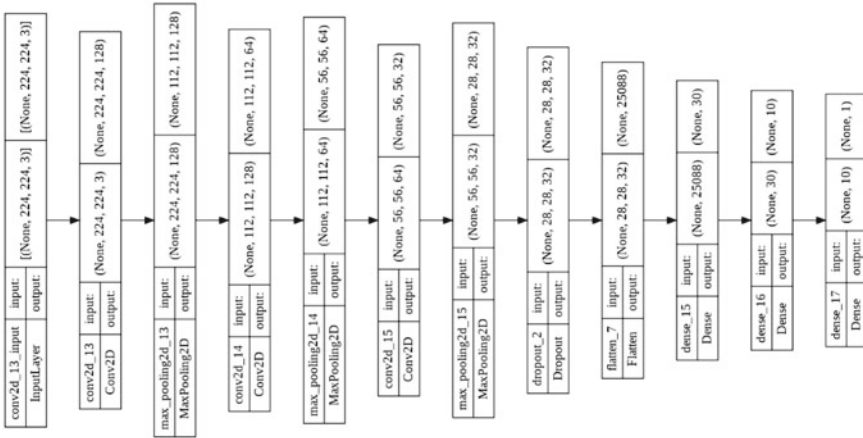
Fig. 2 Proposed CNN model architecture

image. Then, the max-pooling layer, this layer down samples the input by taking the maximum value from a set of adjacent pixels, which helps to reduce the number of parameters in the network and makes the network more efficient [17].

The next layers are fully connected, and they are responsible for classifying the input. These layers contain many neurons, one for each class that the network is trying to learn. The output of the fully connected layer is passed on to the final layer in this stack, which is called the output layer, it contains one neuron which is in our case a binary classification (Fire or Not), with the sigmoid activation layer (see Fig. 2).

5 Results and Discussions

5.1 Hardware Characteristics

In order to get our results, we used TensorFlow on an HPC system with the following hardware specifications:

- $2 \times$ Intel Gold 6148 (2.4 GHz/20 cores) CPU's
- $2 \times$ NVIDIA Tesla V100 graphics cards, each having 32 GB of RAM.

We used TensorFlow v2.7.0 in our experiments, it is an open-source data analysis and machine learning software library [18].

5.2 Evaluating the Results

The findings demonstrate that our proposed deep learning CNN model is efficient, with an accuracy of 98.03%, a precision of 96.95%, a recall of 98.96%, and a loss of 9.18% (see Figs. 3 and 4). This demonstrates that our deep learning CNN model is capable of properly predicting future forest fires. Our proposed model can potentially be used to anticipate wildfire spread and assist firefighters in planning their response to the fire and in evacuating those who are in the fire's path.
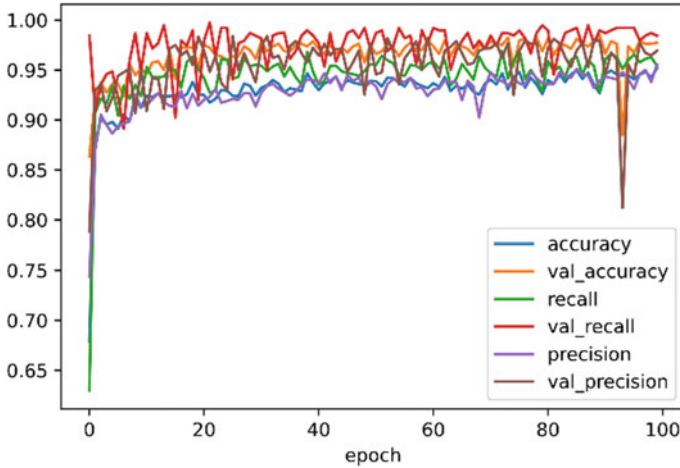
Fig. 3 Accuracy, precision, and recall training and validation results
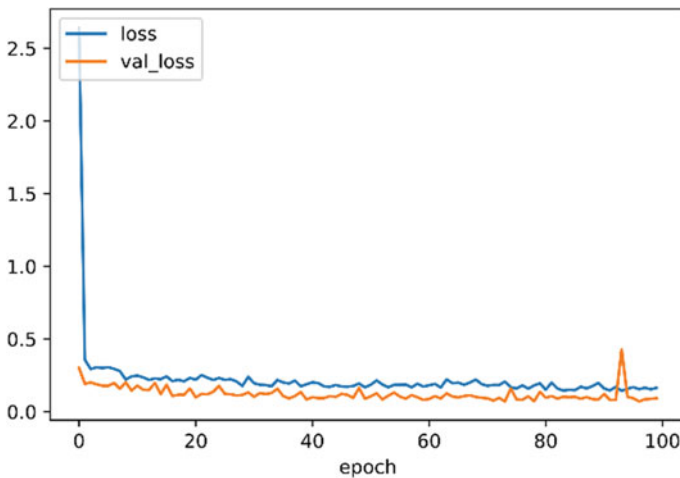
Fig. 4 Loss training and validation results

With a number of parameters of around just one million, our CNN model is one of the lightweight models compared to other most known computer vision models such Xception (22.9 M), VGG16 (138.4 M), ResNet50 (25.6 M), InceptionV3 (23.9 M), MobileNet (4.3 M), MobileNetV2(3.5 M), DenseNet121 (8.1 M), NASNetMobile (5.3 M), or EfficientNetB0 (5.3 M) [19]. It could be an excellent model to deploy on Internet of Things (IoT) devices such as drones regarding the capability of this including processed resources and batteries.

The use of drones for wildfire detection is a promising new field that has the potential to help reduce the damage caused by wildfires. A CNN model for wildfire detection using drones can be trained using data from past wildfires and tuned to take into account the specific conditions in a particular area. The model can be used to generate real-time alerts when a wildfire is detected, which can include information about the location and the size of the fire, and the weather conditions at that time. We can also use this model to generate maps of the extent of a wildfire.

6 Conclusion

The use of drones for wildfire detection is a promising new field that has a relevant potential to help reducing the damage caused by wildfires. Our proposed lightweight CNN model for wildfire detection using drones was trained using data from previous wildfires, and it can be tuned to fit specific conditions in a given area. The obtained results, more than 98% in multiple metrics of accuracy, recall, and precision, and as lowest as 9% in the loss and a low number of parameters (around 1 million), are relevant when compared to other computer vision models, for providing earlier detection and monitoring wildfires, and for mapping the extent of a wildfire once it is detected.

As future works, we intend to investigate further methods that use transfer learning to gain from previously learned models such as VGG, ResNet, MobilNET, and other well-known deep learning models for computer vision, as well as other computer vision techniques such as object detection and segmentation, to precisely determine the location of fires.

References

1. World Health Organization Wildfires. https://www.who.int/health-topics/wildfires. Accessed 3 Oct 2020
2. Alkhatib AAA (2014) A review on forest fire detection techniques:https://doi.org/10.1155/2014/597368
3. Kherraki A, El OR (2022) Deep convolutional neural networks architecture for an efficient emergency vehicle classification in real-time traffic monitoring. IAES Int J Artif Intell 11:110–120
4. Boukabous M, Azizi M (2022) Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers. Indones J Electr Eng Comput Sci 25:1131–1139. https://doi.org/10.11591/IJEECS.V25.I2.PP1131-1139
5. Kherraki A, Maqbool M, El Ouazzani R (2021) Traffic scene semantic segmentation by using several deep convolutional neural networks. In: 2021 3rd IEEE Middle East North Africa communications conference, pp 1–6. https://doi.org/10.1109/MENACOMM50742.2021.9678270
6. Idrissi I, Boukabous M, Azizi M et al (2021) Toward a deep learning-based intrusion detection system for IoT against botnet attacks. IAES Int J Artif Intell 10:110–120
7. Idrissi I, Azizi M, Moussaoui O (2020) IoT security with deep learning-based intrusion detection systems: a systematic literature review. In: 4th International conference on intelligent computing in data sciences, ICDS 2020, Institute of Electrical and Electronics Engineers (IEEE), pp 1–10
8. Boukabous M, Azizi M (2021) Review of learning-based techniques of sentiment analysis for security purposes. Innovations in Smart Cities Applications, vol 4. Springer, Cham, pp 96–109
9. Idrissi I, Azizi M, Moussaoui O (2022) An unsupervised generative adversarial network based-host intrusion detection system for internet of things devices. Indones J Electr Eng Comput Sci 25:1140–1150. https://doi.org/10.11591/IJEECS.V25.I2.PP1140-1150
10. Zheng S, Wang W, Liu Z, Wu Z (2021) Forest farm fire drone monitoring system based on deep learning and unmanned aerial vehicle imagery. Math Probl Eng 2021:1–13. https://doi.org/10.1155/2021/3224164
11. Yanık A, Serdar Güzel M, Yanık M, Bostancı E Machine learning based early fire detection system using a low-cost drone
12. Khennou F, Ghaoui J, Akhloufi MA (2021) Forest fire spread prediction using deep learning. 16. https://doi.org/10.1117/12.2585997
13. Mohammed RK (2022) A real-time forest fire and smoke detection system using deep learning. Int J Nonlinear Anal Appl 13:2008–6822. https://doi.org/10.22075/ijnaa.2022.5899
14. Forest Fire Images|Kaggle. https://www.kaggle.com/mohnishsaiprasad/forest-fire-images. Accessed 27 Jan 2022
15. Berrahal M, Azizi M (2022) Optimal text-to-image synthesis model for generating portrait images using generative adversarial network techniques. Indones J Electr Eng Comput Sci 25. https://doi.org/10.11591/IJEECS.V25.I2.PP
16. Berrahal M, Azizi M (2020) Review of DL-based generation techniques of augmented images using portraits specification. In: 4th International conference on intelligent computing in data sciences, ICDS 2020, Institute of Electrical and Electronics Engineers (IEEE), pp 1–8
17. Idrissi I, Azizi M, Moussaoui O (2021) Accelerating the update of a DL-based IDS for IoT using deep transfer learning. Indones J Electr Eng Comput Sci 23:1059–1067
18. API Documentation|TensorFlow Core v2.7.0. https://www.tensorflow.org/api_docs. Accessed 27 Jan 2022
19. Keras Applications. https://keras.io/api/applications/. Accessed 30 Jan 2022

A Review on Precision Irrigation Techniques and Controls for Better Water Use Efficiency in Agriculture

Hassan Mokhtari, Mohammed Benzaouia, Hajji Bekkay, and Khalid Chaabane

Abstract The rapid increase in demand for freshwater due to population growth and the emergence of different industrial sectors, specifically the agricultural sector, consumes huge quantities of water, while the effects of global warming pose a serious threat to food security and water availability. To achieve this, irrigation system development is needed to reduce and streamline both water and energy use. In addition, the automation orientation of these systems can improve productivity, increase irrigation efficiency, and reduce labor costs. This paper presents in-depth research on traditional irrigation techniques and techniques recently developed by researchers over the last few years. This paper is expected to be used to improve researchers' knowledge and provide in-depth knowledge related to irrigated agriculture.

Keywords Precision irrigation · Irrigation systems · Advanced intelligent control · Water-saving

1 Introduction

Water is a key element of all agricultural production and its availability for the crop considerably affects its productivity. Rainfed agriculture, which depends on meteorological precipitation, is dependent on the amount of water received and its distribution in space and time. However, its importance is limited to regions with favorable climates [1]. In regions with semi-arid to arid climates or low rainfall, irrigation, which is defined as an artificial supply of water to the plant, is practiced in large agricultural areas to fill any partial or general water deficit [2]. The recourse to irrigation is increasingly manifested given a growing population and climate change accentuating the unavailability of water [3]. It is also necessary to know that the

H. Mokhtari (✉) · K. Chaabane
Faculty of Sciences Oujda, Mohamed First University, Oujda, Morocco
e-mail: gr_hassan@yahoo.fr

M. Benzaouia · H. Bekkay
Renewable Energy, Embedded System, and Data Processing Laboratory, National School of Applied Sciences, Mohamed First University, Oujda, Morocco

community must make good use of this scarce water resource to obtain the maximum production yield and the maximum financial profitability while giving priority to its economy [4]. The rational provision of irrigation water depends mainly on the concept of uniformity of distribution which reflects the equity of water sharing between plants but also on the performance of the distribution network and the management of water supplies to meet the water needs of crops.

In this chapter, a presentation of the irrigation systems, especially those that are more adapted to precision irrigation and the main management methods, especially those based on automated systems with the latest technological inventions, but especially those with the lowest costs. Irrigation needs are also presented in the broadest sense, namely the plant's own needs, the needs of the required washing, and the conservation of the wetting bulb near the roots.

Irrigation can be defined as a process of bringing water in an artificial way to the plant to satisfy in part or totality its needs in the water. The objective of irrigation is mainly to satisfy the water needs of crops to achieve yields similar to those of rainfed agriculture with well-distributed rainfall in space and time but also to other objectives including; the reduction of the risk of frost, the washing required for the leaching of the soil and the conservation of the salinity profile in the bulb of humectation in the vicinity of the roots in a drip irrigation system.

Irrigation uniformity gives an idea of the fairness of water distribution among plants in the same plot [5]. The uniformity coefficient proposed by Benzaouia et al. [6], whose rule is named according to him, is equal to the ratio of the variation in water flow to plants in the same irrigation station (Eq. 1).

$$CU = \frac{\Delta Q}{Q} \tag{1}$$

where CU is the uniformity coefficient expressed in %, $\Delta Q = Q \max - Q \min$ is the maximum variation between the flow rate received by the most favored plant and the most disadvantaged plant, Q is the average flow to be distributed to all plants.

For uniform irrigation, Christiansen sets a minimum CU limit of 90% meaning a maximum allowable flow variation of 10% between the most and least watered plants on the same irrigation [7]. The design and sizing of an irrigation project are essentially based on this rule, which must be satisfied [8].

Another very important factor in the evaluation of irrigation efficiency is the yield of the irrigation system which is measured as the ratio of the amount of irrigation water to the amount of water mobilized (Eq. 2) [9].

$$E = \frac{Va}{Vt} \tag{2}$$

where Va is the watering volume, Vt is the total mobilized volume.

This factor is of considerable importance since it gives an idea of the volume of water lost but not used by the plant [10].
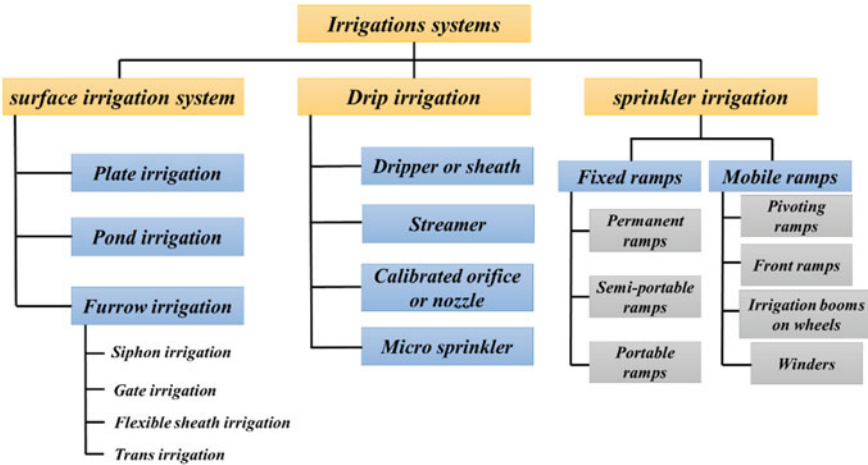
Fig. 1 Classification of irrigation methods

2 Classification of Irrigation Methods and Performances

Several classifications are cited in the scientific literature. In this paper, we are interested in the classification based on irrigation technology.

In a global way, the irrigation systems are classified into three, namely surface or gravity irrigation, sprinkler irrigation, and micro-irrigation or localized irrigation (Fig. 1).

Surface irrigation is simply the distribution of water over the plot by gravity runoff from the highest point of the land. The water is then distributed either by submersion (pond irrigation), in furrows in the ground (furrow or furrow irrigation), or by runoff on the surface of a watering board (board irrigation).

Sprinkler irrigation is an imitation of rain by distributing water as artificial rain through sprinklers.

Micro-irrigation or localized irrigation is a punctual contribution of water in the vicinity of the plants. It is particularly well known for its higher distribution efficiency. Table 1 gives an overview of the efficiency of each system.

3 Design of an Irrigation System

To succeed in an irrigation project is to be able to answer the questions, of when, how, and how much water to mobilize for a crop. If in the previous paragraph we have pertinently noted the interest of the choice of the localized drip irrigation system in water-saving and therefore answer the question of how it remains to answer the questions of when and how much by an adequate management processor. Answering

Table 1 Efficiency of irrigation systems [10]

Irrigation system or method		E (%)
Surface Irrigation	Earth canal network	40–50
	Lined canal network	50–60
	Pressure pipe network	65–75
	Hose irrigation systems	70–80
Sprinkler irrigation	Low and medium pressure sprinkler systems	70–75
	Micro-sprinklers, micro-jets, mini-sprinklers	75–85
Localized irrigation	Drip irrigation	80–90

these last two questions is to identify the water needs of the plant to try to meet them at the right time and in sufficient quantity.

3.1 Crop Water Requirements

The water requirement of crops corresponds to the reaction of the soil–plant complex to the climatic demand. It is assessed by the evapotranspiration of the crop, which is the sum of evaporation from the soil and transpiration by the plant of a quantity of water under given climatic conditions. This evapotranspiration is evaluated either by in situ measurements (lysimeter, class A tank) or by empirical calculation based on climatic parameters. It can also be estimated by processing satellite or aerial imagery (plane or drone).

3.2 Precision Irrigation

Precision irrigation refers to modern techniques to control and manage irrigation to automate the water supply operation and meet the needs of the crop optimally. Several techniques are studied and used. They are based mainly on the manipulation of agro-pedo-climatic data which are then processed by a rational system to decide in an automatic way to bring a volume of water to the crop in sufficient quantity and at the right time.

4 Control Techniques for Precision Irrigation

Irrigation systems require the adoption and implementation of different advanced control techniques, to apply the necessary amount of water to a plant, which will improve the precision of irrigation, thus rationalizing the use of water and energy.
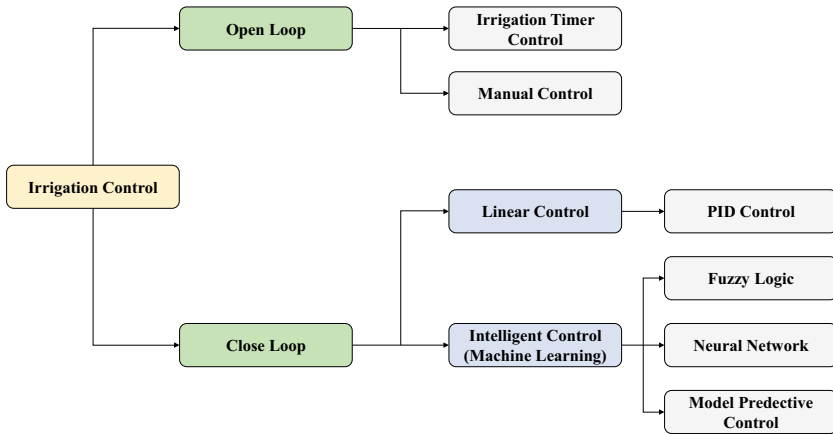
Fig. 2 Classification of different control techniques for precision irrigation system

Methods based on quantifying the crop's water requirements either by evaluating the weather conditions that mainly influence irrigation or approaches based on the evaluation of soil or plant-related parameters such as canopy temperature have been proposed with different types of control either by an open-loop or closed-loop controls or by hybridizing both. Figure 2 presents the different control techniques and their approaches.

4.1　Open-Loop Control

The first type of control (open loop) is widely used by farmers, either by using mechanical or electromechanical irrigation timers or by direct water applications where agriculture relies on its prior knowledge of crop response to irrigation. This control does not require the use of sensors to measure the different parameters that affect the irrigation process, which makes the system easy to implement and less expensive, but of low reliability and efficiency. In [11], the authors proposed an open-loop control as shown in Fig. 3, the time was programmed using a commercial clock connected to an Arduino-type development board, which allows the operation of two alternative and continuous pumps for sprinkler and drip irrigation purposes. The use of this type of control has also been used in greenhouses. The problem of this type of control is its low sensitivity to the conditions and disturbances either meteorological or land, as well as the stages of development of crops. So generally, the open-loop irrigation technique by irrigation timers, by volumetric approach, or by conventional approach represents a simple control composed of timing units allowing activating one or more units such as solenoid valves at a precise moment without taking into account the influence of the applied quantity of water on the plant response. This control is less effective knowing that the application of excessive amounts of water
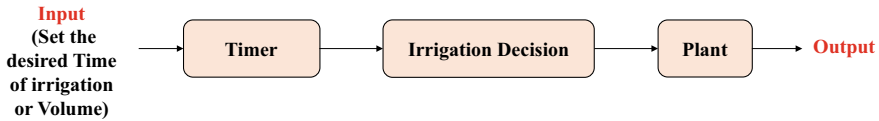
Fig. 3 Open-loop irrigation system
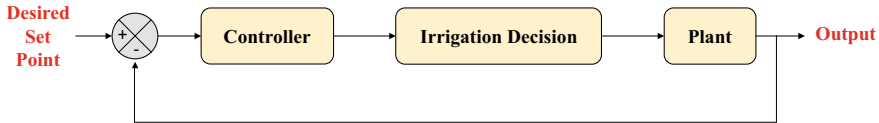
Fig. 4 Closed-loop irrigation system

can cause serious damage to the plant and allows the appearance and spread of diseases and weeds.

4.2 Closed-Loop Control

For the second control (closed loop) (Fig. 4), the system works by feedback. A feedback loop is a control that links the effect to its cause, with or without delay, either to maintain the desired output or to make changes. The application of this control in irrigation consists in either deciding the irrigation times or the opening rate of the solenoid valves while evaluating different parameters. In [12], the authors proposed a fully automated irrigation system based on a closed-loop control, while relying on the calculation of the plant's water needs. The system processes the information from sensors to estimate the time and frequency of water application and the place to irrigate. The sensors are therefore important devices to generate information and data such as air humidity, soil humidity, temperature, solar radiation, and NDVI to guide the irrigation process and to initiate the activation of the actuators (motor, pump, solenoid valve, etc.).

The closed-loop control has known important improvements through the use of artificial intelligence (AI) algorithms such as neural networks, prediction and forecasting algorithms, fuzzy logic, and PID controllers. This results in the development of a system that is both intelligent and accurate.

4.2.1 Proportional–Integral–Derivative (PID) Controller

Proportional–integral–derivative systems, known as PID, are widely used in feedback control systems due to their simple structure, and the reliability of tracking a desired reference output value while minimizing the error between the setpoint and the

Fig. 5 Fuzzy logic structure
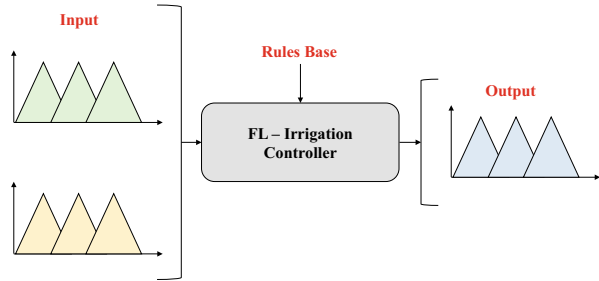
resulting values. However, conventional PID controllers may diverge when the three parameters Ki, KP, and Kd are not optimized, which decreases the control performance. Several methods are used to adjust the three parameters such as genetic algorithm and Ziegler–Nichols. In [13], and improved PID controller has been proposed, to ensure the regulation of water quantities to meet the needs of the plant. The environmental conditions including rainfall, plant stress are taken into account, the controller reacts quickly to the action result.

4.2.2 Fuzzy Logic Controller

The nonlinearity and the difficulty to describe a mathematical model to describe the behavior of an irrigation system, pushed several researchers to develop controllers based on fuzzy logic. The input and output variables are real; in this sense, it extends the classical Boolean logic with partial truth values (Fig. 5). The control consists in taking into account the different numerical factors to reach acceptable decisions. In [14], a fuzzy feedback control has been used to develop an irrigation approach that results in water conservation in hyper-arid lands like Qatar. Triangular and trapezoidal membership functions were used to efficiently generate irrigation times for a given crop. The control also allowed to maintain a user-defined soil moisture content. A fuzzy inference-based irrigation controller is designed and developed in [15] and implemented using LabVIEW and the GPRS/GSM platform while providing an innovative solution for efficient irrigation scheduling. The fuzzy logic control decision is the percentage of valve opening based on the predefined soil moisture value. The system also compensates for the water loss due to evapotranspiration during the whole plant growth period. The proposed technique has shown that it is possible to save water at 2.85%.

4.2.3 Neural-Network Controller

Neural networks are an information processing tool inspired by the functioning of the human biological nervous system. It is widely used in the fields of industry and automation because of its aspect of learning and adapting to the dynamic variables
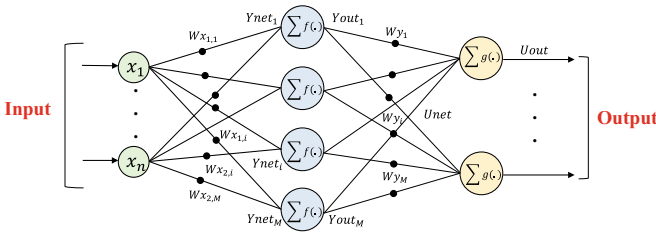
Fig. 6 ANN structure

that affect the systems. This controller is used to develop irrigation systems; in [16], the authors proposed an on–off system based on a neural controller and a strategy of quantification of water requirements by evapotranspiration calculated by the formula of Penman–Monteith. The results obtained show that the error of the ON–OFF controller can be adjusted by setting the sampling time of the dead zone discretization. Other research focused on image processing through a standard video camera to capture and analyze the different colors of the ground at different distances, at different times, and at different illumination levels. The proposed system used the color information as input to an artificial neural network (ANN) system to decide to irrigate or not the soil. The proposed system was found to be highly accurate, making it a promising technology to support precision agriculture. Figure 6 gives the structure of an ANN-based controller.

4.2.4 Model Predictive-Based Irrigation Controller

The prediction algorithms have the advantage of predicting a variable with high accuracy; moreover, their ability to use simple models with better closed-loop stability and their robustness against parametric uncertainties make them one of the most widely used multivariate control algorithms in the irrigation field. In [17], the authors proposed an improved weather forecasting method where the predicted weather data influences soil moisture. Machine learning is developed to predict soil moisture. The algorithm uses sensor data from the recent past and forecasted weather data to predict soil moisture for the next few days. The predicted soil moisture value is better in terms of accuracy and error rate. For example, the long short-term memory (LSTM) (Fig. 7) algorithm is used to predict weather conditions or a parameter such as soil moisture in order to decide the irrigation process.

5 Conclusion

The review of the techniques mentioned confirms the interest of AI and precision controllers in the management of irrigation in the world as a concern of water-saving

Fig. 7 LSTM structure
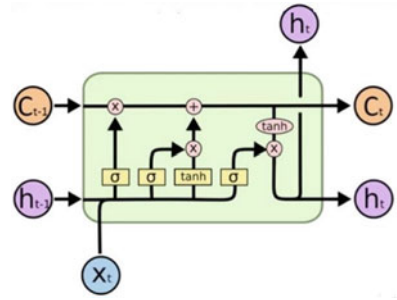
given the effects of climate change on territories with low rainfall and growing population. A lot of research work is being done to rationalize the use of irrigation water and to considerably reduce water losses by improving irrigation systems or by good control of irrigation management. Precision irrigation provides a window of optimism for humanity to face the challenge of increasing agricultural demand and limited or declining water resources. The present work is interested in a presentation of these techniques by specifying the advantage of taking into consideration in the estimation of the irrigation need; the water need of the crop with the fraction of the required wash while preserving, for the case of drip irrigation, the volume of the wetting bulb to avoid the phenomenon of osmotic shock in case of withdrawal of this bulb and invasion of the roots by the salt front. The comparison of these techniques leads us to think about the simultaneous use of climatic data for the calculation of reference evapotranspiration according to the modified Penman formula and the measurement of soil moisture with remote handling of the reception and transmission of data and the operation of valves. The combination of these techniques with the use of low-cost and low-energy consumption electronic and electromechanical components will undoubtedly allow a democratization of the exploitation of these techniques.

References

1. Abioye EA, Abidin MSZ, Mahmud MSA, Buyamin S, Ishak MHI, Abd Rahman MKI, Ramli MSA (2020) A review on monitoring and advanced control strategies for precision irrigation. Comput Electron Agric 173:105441
2. Jaiswal S, Ballal MS (2020) Fuzzy inference based irrigation controller for agricultural demand side management. Comput Electron Agric 175:105537
3. Li M, Sui R, Meng Y, Yan H (2019) A real-time fuzzy decision support system for Alfalfa irrigation. Comput Electron Agric 163:104870
4. Touati F, Al-Hitmi M, Benhmed K, Tabish R (2013) A fuzzy logic based irrigation system enhanced with wireless data logging applied to the state of Qatar. Comput Electron Agric 98:233–241
5. Benzaouia M, Bouselham L, Hajji B, Dubois AM, Benslimane A, El Ouariachi M (2019) Design and performance analysis of a photovoltaic water pumping system based on DC-DC boost converter and BLDC motor. In: 2019 7th International renewable and sustainable energy conference (IRSEC). IEEE, pp 1–6

6. Benzaouia M et al (2020) Energy management strategy for an optimum control of a standalone photovoltaic-batteries water pumping system for agriculture applications. In: International conference on electronic engineering and renewable energy. Springer, Singapore

7. Ma'Mun SR, Loch A, Young MD (2020) Robust irrigation system institutions: a global comparison. Glob Environ Chang 64:102128

8. Romero R, Muriel JL, García I, de la Peña DM (2012) Research on automatic irrigation control: state of the art and recent results. Agric Water Manage 114:59–66

9. Goodchild MS, Kühn KD, Jenkins MD, Burek KJ, Button AJ (2015) A method for precision closed-loop irrigation using a modified PID control algorithm. Sens Transducers 188(5):61

10. Sammis TW (1980) Comparison of sprinkler, trickle, subsurface, and furrow irrigation methods for row crops 1. Agron J 72(5):701–704

11. Capraro F, Patino D, Tosetti S, Schugurensky C (2008) Neural network-based irrigation control for precision agriculture. In: 2008 IEEE International conference on networking, sensing and control. IEEE, pp 357–362

12. Krishnan RS, Julie EG, Robinson YH, Raja S, Kumar R, Thong PH (2020) Fuzzy logic based smart irrigation system using internet of things. J Clean Prod 252:119902

13. Aquino-Jr PT, Maia RF, Kamienski C, Soininen JP. A fuzzy irrigation control system

14. Jamroen C, Komkum P, Fongkerd C, Krongpha W (2020) An intelligent irrigation scheduling system using low-cost wireless sensor network toward sustainable and precision agriculture. IEEE Access 8:172756–172769

15. Zerouali M, Zouirech S, El Ougli A, Tidhaf B, Zrouri H (2019) Improvement of conventional MPPT techniques P&O and INC by integration of fuzzy logic. In: 2019 7th International renewable and sustainable energy conference (IRSEC). IEEE, pp 1–6

16. Zerouali M, El Ougli A, Tidhaf B, Zrouri H (2020) Fuzzy logic MPPT and battery charging control for photovoltaic system under real weather conditions. In: 2020 IEEE 2nd International conference on electronics, control, optimization and computer science (ICECOCS). IEEE, pp 1–5

17. Soulis KX, Elmaloglou S (2018) Optimum soil water content sensors placement for surface drip irrigation scheduling in layered soils. Comput Electron Agric 152:1–8

Renewable Energy

Assessing the Thermal Performance of Traditional and Modern Building Materials for Hot and Arid Climate. Case Study: Er-Rachidia, Morocco

Ali Lamrani Alaoui, Abdel-illah Amrani, Ahmed Alami Merrouni, Abdelkarim Daoudia, Youssef El Hassouani, Elmiloud Chaabelasri, and Mohammed Halimi

Abstract The current study aims the comparison of the energy performance for two building techniques—traditional and modern—located in Er-Rachidia, southeast of Morocco. To do so, a comparison of different building materials and their impact on the thermal load and heating and cooling has been carried out. The energy performance of the building is determined through a numerical study carried out by means of a dynamic multi-zone modeling of several configurations of the restaurant using TRNSYS 16 software. The simulations were performed with meteorological data given by Meteonorm7.3. The results show that clay construction reduces the temperature variation during the day to 3 °C, while in double-walled brick this variation can reach over 6 °C. Moreover, since both types of construction have the same energy performance, the difference in total consumption does not exceed 5%.

Keywords Energy efficiency · Building · Clay · Double brick · South-east Morocco

A. Lamrani Alaoui (✉) · A. Amrani · Y. El Hassouani
Equipe des Energies Nouvelles et Ingénierie des Matériaux (ENIM), Laboratoire de Sciences et Techniques de l'Ingénieur (LSTI), Faculté des Sciences et Techniques, Moulay Ismail University of Meknès, Boutalamine BP 509, 52000 Errachidia, Morocco
e-mail: alilamranialaoui@gmail.com

A. Daoudia
Equipe de Physique Théorique et Modélisation (PTM), Laboratoire de Sciences et Techniques de l'Ingénieur (LSTI), Faculté des Sciences et Techniques, Moulay Ismail University of Meknès, Boutalamine BP 509, 52000 Errachidia, Morocco

A. Alami Merrouni · E. Chaabelasri
Materials Science, New Energies and Applications Research Group, LPTPME Laboratory, Department of Physics, Faculty of Sciences, Mohammed First University, 60000 Oujda, Morocco

M. Halimi
Department of Physics, Faculté des Sciences et Techniques, Moulay Ismail University of Meknès, Boutalamine BP 509, 52000 Errachidia, Morocco
e-mail: m.halimi@edu.umi.ac.ma

1   Introduction

The energy consumption in the Moroccan building sector in Morocco is considered as high. This consumption is reaching up to 33% from the total consumption distributed as follows: 6% for residential buildings and 7% for tertiary buildings. The main cause for this consumption is the demographic growth, the creation of new cities and the increase of the live quality, mainly the use of air conditioning and heating systems [1]. On the other hand, Morocco is a country lacking from fossil energy sources since it imports 94% of its energy needs from abroad [2, 3]. This demonstrates the importance of improving the thermal performance of buildings by the integration of renewable energies [4–6] and the improvement of the constriction materials. In this context, several studies have been carried out on traditional constructions based on raw clay to improve their thermal comfort. Aboubou et al. [7] and Grigoletto et al. [8] show that the use of clay in construction can achieve specific properties of mechanical strength and thermal inertia. They also studied the effect of adding different percentages by weight of rice husks to clay collected in the Bamako region. This additive reduced the thermal conductivity of this building material. Lamrani et al. [9] reported that the addition of hemp fibbers has a positive impact on the mechanical strength and thermal conductivity of fired clay bricks. This clay is collected from the city of Erfoud [10–12]. In the same context, several studies are carried out on modern construction techniques in Morocco. The types of windows in the building can improve indoor thermal comfort. Lachheb et al. [13] demonstrate that with a xenon triple glazing, energy savings can reach more than 90% with a window to wall ratio of 90% in all climate zones. Lamrani et al. [14] show that clay and traditional building materials provide better thermal comfort than brick construction.

The present research consists in studying the effect of building materials on the thermal comfort of buildings, through the dynamic thermal simulation of two restaurants built by two different types of construction in the region of Er-Rachidia, to evaluate their thermal comfort and its environmental benefits to use the proposed material in a residential building located in southern Morocco.

2   Method and Materials

2.1   Building Description

The model of the studied building is a room of a Moroccan restaurant of a surface of 80 m^2 oriented toward the south, built by two different techniques. In the first case, the room is built by walls based on clay, while in the second case, the walls are double partition of bricks. The representative plan of this room is illustrated in Fig. 1.

The thermal properties of the building materials used in this study are presented in Table 1. Most of the windows in Moroccan buildings today are made of single

Fig. 1 Plans of the room
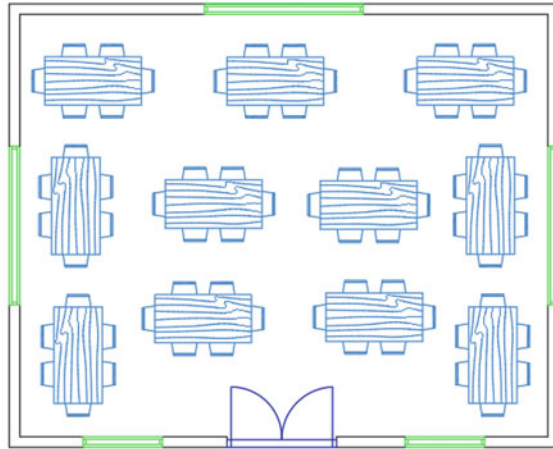
glazing with a U-value of 5.74 w/m²k. The window surfaces are 4 m² on the south front and 3 m² on the other front of the room. The exterior door is 2 m wide and 2 m high, and it is selected to be facing south in this study. In addition, the door is made of heavy wood, which is a typical construction material for a door in Moroccan buildings. The U-value of the door is 2.95 W/m²K.

Table 1 Characteristics of the building construction elements for two cases

	Building components	Material (layers)	Thickness (cm)	Specific heat (kJ/kg.K)	Thermal conductivity (W/m.K)	Density (kg/m³)	U-value (W/m²K)
Case 1	Exterior wall	Clay	40	1.042	0.7	1835	1.131
	Roof	Clay Wood	15 5	1.042 1.8	0.7 0.25	1835 580	1.77
Case 2	Exterior wall	Mortar Brick Air gap Brick Mortar	1.5 10 10 10 1.5	0.84 0.794 1.23 0.794 0.84	1.15 0.47 0.1 0.47 1.15	2000 720 1 720 2000	0.515
	Roof	Mosaic Cement Concrete hourdis Mortar	1.5 7 15 1.5	0.84 0.92 0.65 0.84	2.42 1.755 1.23 1.15	3200 2300 1300 2000	2.535
both cases	Floor	Tile Mortar Cement	1.5 5 20	0.7 0.84 0.92	1.7 1.15 1.755	2300 2000 2300	2.975

(a) Exterior wall

(b) Roof

(c) Floor

Fig. 2 Building constituents of case 1

(a) Exterior wall

(b) Roof

Fig. 3 Building constituents of case 2

The detailed composition of the walls, floor and roof construction of the room for the first case is presented in Table 1 and Fig. 2, and for the room of the 2nd case is presented in Table 1 and Fig. 3.

2.2 Weather Data

The city of Er-Rachidia is located in the southeast of Morocco at a latitude of 31°48' North, a longitude of 4°4' West and at an altitude of 1009 m. It is classified among the hot and arid zones with relatively high daily thermal amplitudes. According to the meteorological data given by metronome, the average annual temperature is 20.6 °C.

To run the dynamic thermal simulations, we used a Typical Meteorological Year (TMY) weather file of the city Er-Rachidia provided by Meteonorm [Meteonorm7.3]. The TMY dataset is more representative than those measured over one year since it presents an average of 10 years' measurements—from 2000 to 2009—which

overlap the inter-annual variance of the meteorological parameters. The used TMY file contains the following climate data for Er-Rachidia: ambient air temperature, relative humidity, direct and diffuse solar radiation and wind speed.

3 Results and Discussion

3.1 *Indoor Temperatures Evaluation*

The hourly indoor simulation of the air temperature variations among the year is presented in Fig. 4. As illustrated, the ambient temperature of Er-Rachidia city has a minimum value equal to −0.5 °C, while the maximum value reaches 42 °C.

Regarding the building's temperature, it is observed that the double brick wall is warmer than the room built by clay despite they have the same orientation, the same glass surface and the same volume. This difference is mainly due to the types of the construction materials. This difference varies between the climatic seasons of the year.

The following Figs. 5 and 6 compare the evolution of the indoor temperatures in the two rooms during the coldest and the warmest week of the year, respectively. According to Fig. 5, the temperature inside the room built by clay varies in the third week of January between 7 °C and 12.37 °C, while it varies between 6 °C and 14.15 °C for the room built in brick. This reflects a temperature difference of 0.1 and 3.9 °C between the cold weeks, respectively. Figure 6 shows that the clay room temperature in summer varies between 24.22 °C and 27.66 °C, while the variation
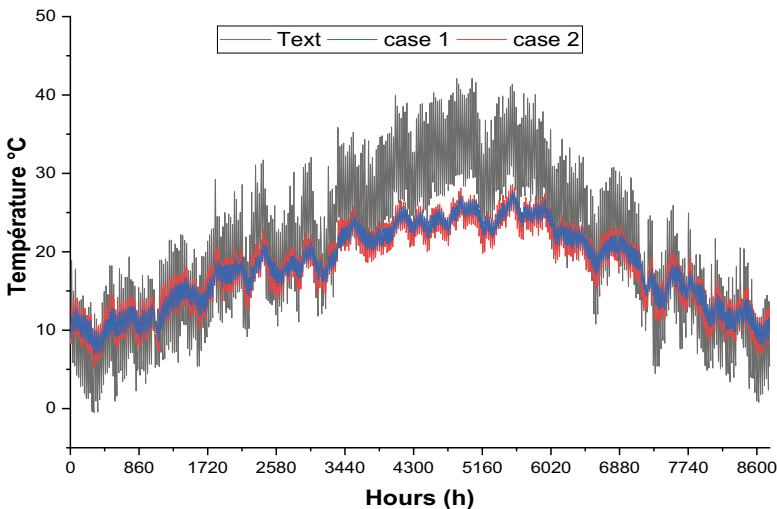
Fig. 4 Evolution of the indoor temperature for both cases throughout the year
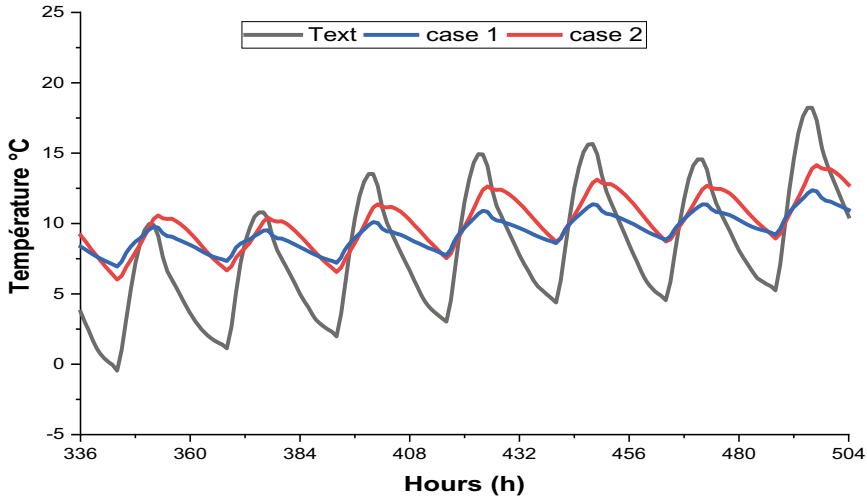
Fig. 5 Variation of the temperature of the two cases, winter period

is between 23.11 °C and 28.6 °C for the double brick construction. This mean that during the summer period, the difference between the two cases/building materials varies between 3.04 and 6.19 °C. This value can be considered as high in comparison to the winter, and they can lead to a great drop in the energy consumption, knowing the fact that Er-Rachidia has a hot summer which requires the use of air-conditioning devices for long periods. Thus, the presence of a material with a high thermal capacity such as clay reduces the temperature variations. Indeed, throughout the year, we see that the interior temperature increases and decreases less quickly (lower slope) in the room built in clay than in the room built in double brick wall. Furthermore, clay has a high density, which gives it valuable thermal inertia qualities for heat storage and regulation of temperature changes between day and night. Therefore, the double brick wall construction technique is a little more efficient in winter than the clay technique, but tends to overheat in the hot season. Consequently, for more complete and fair comparison of the overall materials efficiency, we need to evaluate the monthly energy consumption for both rooms. The results of the simulations will be presented in the next section.

3.2 Analysis of the Monthly Energy Needs

For our simulations, the cooling and heating setpoint temperatures were set at 26 °C and 22 °C, respectively, for the calculation of energy demand. We need to mention that for the energy consumption calculations, we take into consideration several parameters like the presence of people in the building, the existence and the operation of materials, etc. Figure 7 presents a comparison of the monthly energy consumption
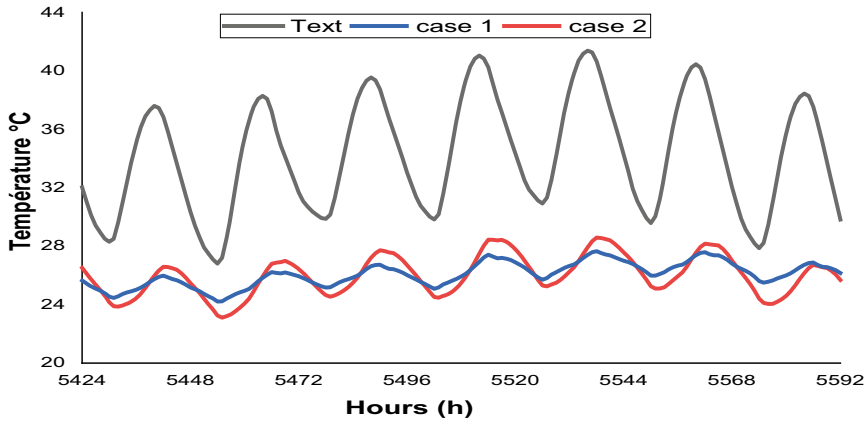
Fig. 6 Variation of the temperature of the two cases, summer period

Fig. 7 Comparison of energy consumption in kWh between the two cases
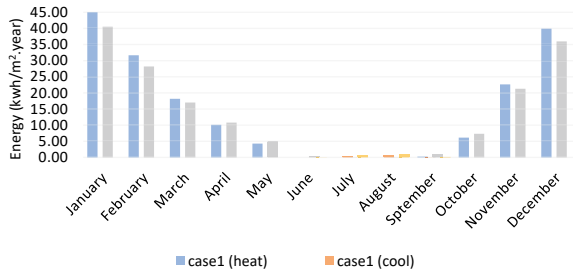
between the two cases. As it can be noticed, the building of case 1 consumes more energy in comparison to the other case in heating for the cold months by a difference of 914 kWh. On the other hand, in the summer the energy consumption in air-conditioning for the room built in clay is less than that of the other room by a difference of 72 kWh. The total energy consumption of air conditioning and heating for the whole year in case 1 is 179.75 kWh/m^2.year, and for the second case is 169.22 kWh/m^2.year. It is observed that the annual energy consumption of room 1 is 10.53 KWh greater than room 1.

4   Conclusion

In this paper, we have presented the results of mono-zone thermal simulations; using TRNSYS16 software; of a restaurant type building located in Er-Rachidia. For this a comparison is done between two construction techniques: (i) traditional using clay walls and (ii) modern using double brick walls. The present study focuses on the air temperature inside the buildings and the calculation of energy requirements. To wrap

up, both types of construction have virtually the same energy performance, but clay construction has many advantages beyond reducing temperature variations during the day. Clay is an abundant and renewable material and has no negative impact on the health of the occupants. Also, clay construction can produce no waste, and its recycling is done in the form of a reuse or a simple restitution to the local natural environment. Finally, for hot arid climates, clay represents the optimal material for building construction as it is more adapted to the hot climate. In the future, we are aiming to evaluate different mixture of different local materials together with clay and to assess their performances on the energy efficiency and the thermal comfort.

References

1. AMEE. https://www.amee.ma/fr/expertise/batiment. Last accessed 10 Jul 2021
2. Programme National de l'Efficacité Energétique dans le Bâtiment, Agence de Développement des Energies Renouvelables et de l'Efficacité Energétique, ADEREE (www.ceeb.ma). Last accessed 11 Jul 2021
3. Touili S, Merrouni AA, El Hassouani Y, Amrani A, Rachidi S (2020) Analysis of the yield and production cost of large-scale electrolytic hydrogen from different solar technologies and under several Moroccan climate zones. Int J Hydrogen Energy 45(51)
4. Merrouni A, Ait Lahoussine Ouali H, Moussaoui M, Mezrhab A (2016) Integration of PV in the Moroccan buildings: simulation of a small roof system installed in Eastern Morocco. Int J Renew Energy Res (IJRER) 6(1):306–314
5. Merrouni AA, Ghennioui A, Wolfertstetter F, Mezrhab A (2017) The uncertainty of the HelioClim-3 DNI data under Moroccan climate. AIP Publishing LLC, p 140002
6. Mouaky A, Merrouni AA, Laadel NE (2019) Simulation and experimental validation of a parabolic trough plant for solar thermal applications under the semi-arid climate conditions. Sol Energy 194:969–985
7. Aboubou I, Diabaté L, Mengy G (2004) Détermination de la conductivité thermique de matériaux de construction à bamako (balle de riz). J Sci Pour l'Ingénieur 4:1–6
8. Grigoletto S, Lebeau F, Courard L, Paul J, Moutschen P (2015) Applications de l'argile crue en construction. In: Conférence Internationale Francophone NoMaD 2015. Douai, France, p 10
9. Lamrani M, Mansour M, Laaroussi N, Khalfaoui M (2019) Thermal study of clay bricks reinforced by three ecological materials in south of morocco. Energy Procedia 156:273–277
10. Laaroussi N, Cherki A, Garoum M, Khabbazi A, Feiz A (2013) Thermal properties of a sample prepared using mixtures of clay bricks. Energy Procedia 42:337–346
11. Laaroussi N, Lauriat G, Garoum M, Cherki A, Jannot Y (2014) Measurement of thermal properties of brick materials based on clay mixtures. Constr Build Mater 70:351–361
12. Laaroussia N, Ouakarroucha M, El Azharya K, Mansourb M, Garouma M (2019) Thermal study of clay bricks reinforced by sisal-fibers used in construction in south of Morocco. In: 6th International conference on power and energy systems engineering (CPESE 2019), Okinawa, Japan
13. Lachheb A, Allouhi A, Saadani R, Jamil A, Rahmoune M (2020) Thermal and economic analyses of different glazing systems for a commercial building in various Moroccan climates. Int J Energy Clean Environ 15–41
14. Lamrani AA, Amrani A, Merrouni AA, Daoudia A, El Hassouani Y (2021) Optimization of thermal efficiency in traditional clay-based buildings in hot-dry locations. Case study: the South-eastern region of Morocco. Int J Energy Environ Eng

Mathematical Modeling of the Serpentine Configuration Evaporator for the Construction of Mini Solar Tower Project in Oujda, Morocco

Firyal Latrache, Zakia Hammouch, Benaissa Bellach, and Mohammed Ghammouri

Abstract Morocco is called "The Kingdom of the Sun" because of the higher potential solar concentrated on the whole area of the country. By the way, the construction of a solar plant is motivated, and the application of the national strategy of the sustainable development by using a new configuration and a developed concept allows to a good result that can be used to make a technical dimensioning of the solar plant. The present work is about the using of the serpentine configuration as an evaporator in the process of the production of the vapor that be used after in the process of the production of the electrical energy. The concept of the mini solar tower is the basis of our work by using a configuration: the serpentine configuration that not used in its large term in the process of the evaporation.

Keywords Solar · Thermodynamics · Modelization

1 Introduction to the Topic and an Overview of the Literature

Renewable energy is, in fact, regarded as one of the most important keys to achieving a sustainable world and a clean environment [1]. By the way, renewable energy is becoming more competitive and efficient than fossil energy resources [2]. Solar energy, hydraulic energy, wind energy, geothermal energy, and other sources of renewable energy exist, and solar energy is separated into thermal and photovoltaic energy [3]. Because of their low-temperature processes, solar thermal energy is used in many industrial industries [4]. Sun thermal energy focuses solar irradiation using a variety of technologies, including cylinder-parabolic concentrators, parabolic concentrators, tower concentrators, and Fresnel lenses, and enables for power generation by thermoelectric conversion [5]. The goal of this research is to present a

F. Latrache (✉) · Z. Hammouch · B. Bellach · M. Ghammouri
Team Mechanics, Energetics, Signals and Systems, Laboratory of Laboratory of Mechanics and Scientific Calculation, National, School of Applied Sciences, University of Mohammed First, Oujda, Morocco
e-mail: firyalfaridlatrache@gmail.com

mathematical model of the evaporator of a mini solar power plant that would be suitable for use in remote areas in Morocco's eastern region. According to the literature, serpentine tubes are frequently employed as evaporators, but they are more commonly utilized as condensers and have not yet been developed for evaporator technologies [6]. The nodal approach, which tries to discretize the serpentine geometry of the evaporator into nodes [7], is used to make the mathematical modelization in this article. Indeed, each node determines the heat and mass transfer equations, which are then resolved to predict the output parameters such as the temperature and pressure of the heat transfer fluid.

2 Configuration of Prototype

Figure 1 illustrates the heat exchanger configuration, and in this figure, the configuration is shown in three dimensions, the mass and heat transfers are shown, and the heat transfer fluid is water. The configuration is made up of the serpentine tube (the outer tube) and the cylinder tube (the inner tube). The length of the configuration is 200 cm, the serpentine tube diameter is 8 mm, and the cylinder tube diameter is 11.4 cm. In order to model the heat, the nodal method is used in each node of the inner and outer tubes to determine the outlet temperature and pressure of the vapor. In order to model the heat and mass transfer through the prototype, the nodal method is used in each node of the inner and outer tubes to determine the outlet temperature and pressure of the vapor.

3 Mathematical Modelization

The nodal method is used to determine the heat and mass transfer equations for each component of the heat exchanger: the serpentine tube, the heat transfer fluid, and the cylinder tube.

(a) For the serpentine tube:

$$
\frac{\partial T_{ts}}{\partial t} = \frac{1}{\rho_{ts} \pi D_{ts} C_{ts}} [R_d(\beta) \mathrm{SF}(\cos\theta) \rho \alpha \tau - h_{\mathrm{wind}} A_{ts}(T_{ts} - T_a) \\
+ \in \sigma A_{ts}(T_{ts} - T_a)(T_{ts} + T_a)(T_{ts}^2 + T_a^2)] \tag{1}
$$

With:
T_{ts} is the serpentine tube temperature.
ρ_{ts} is the mass volumic of serpentine tube.
D_{ts} is the diameter of serpentine tube.
C_{ts} is the heat capacity of serpentine tube.
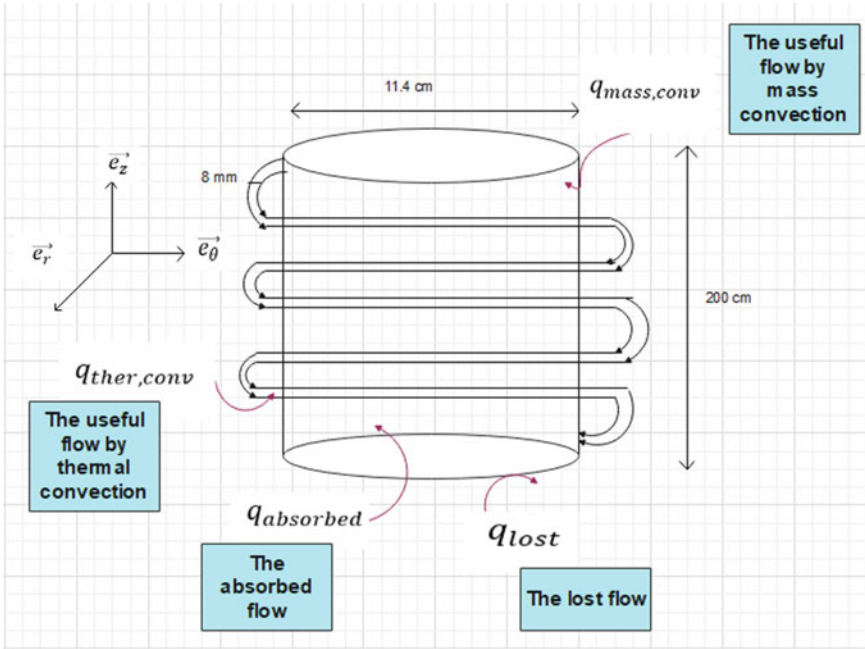
Fig. 1 Representation of thermal and mass transfers through the heat exchanger

$R_d(\beta)$ is the direct radiation component estimated by the Liu and Jordan model:

$$R_d(\beta) = I_h R_b \tag{2}$$

I_h is the direct radiation component received on a horizontal surface given by:

$$I_h = Ae^{\left(\frac{-1}{B\sin(h+C)}\right)} \tag{3}$$

A, B, and C are constants that depend on climatic conditions.
R_b is the inclination factor.
S is the surface of solar concentrator.
$F(\cos\theta)$ is the correction of angle of inclination.
$\rho\alpha\tau$ is the optical efficiency of the system.
h_{wind} is the correlation of the wind convection.
A_{ts} is the surface of the serpentine tube.
T_a is the ambient temperature.
σ is the Boltzmann constant.
$\in$ is the emissivity of the absorber tube.

(b) For the heat transfer fluid:

$$\frac{\partial T_f}{\partial t} = \frac{\dot{V}_f}{A_f}\frac{\partial T_f}{\partial z} + \frac{h_f}{\rho_f C_f}\left(T_f - T_{ts}\right) \tag{4}$$

With:

T_f is the heat transfer fluid temperature.
$\dot{V}_f$ is the volume flow rate of the heat transfer fluid.
A_f is the effective cross section of the heat transfer fluid.
h_f is the thermal convection coefficient of heat transfer fluid.
ρ_f is the volumic mass of heat transfer fluid.
C_f is the heat capacity of heat transfer fluid.

(c) For the cylinder tube:

$$\frac{\partial T_c}{\partial t} = \frac{k}{\rho_c A_c C_c \Delta z}\left(\rho_c - \rho_f\right) \tag{5}$$

With:

T_c is the temperature of cylinder tube.
k is the diffusion speed of the vapor.
ρ_c is the volumic mass of cylinder tube.
C_c is the heat capacity of cylinder tube.
A_c is the surface of cylinder tube.

4   Results and Discussions

The objective of this research work is to build a mini thermodynamic solar power plant in Oujda, Morocco, and the principle of operation is described by the Organic Rankine cycle, which works in the low power range by using organic fluids [8]. Water is used as a heat transfer fluid due to its optimized thermodynamic properties in the high-power range when compared to organic fluids [9]. In practice, the numerical estimation of the direct normal irradiance can reach up to 5 KWh/m^2 on a sunny day in 2019 in Oujda, Morocco, while the measured value is found by "Atlas des Ressources Solaires au Maroc" [10], which can reach up to 5.7 KWh/m^2 in the same period in Oujda. The numerical resolution of the equations describing the thermal and mass transfers through the heat exchanger, on the other hand, allows for the estimation of the outlet temperature and pressure of water vapor, which are 180 °C and 10 bars, respectively. The following results demonstrate the performance of the geometric form of absorber tube placed on the outside of the heat exchanger: when comparing serpentine geometry to cylinder geometry [11]. To construct this mini solar power plant, the length and diameter of the cylinder tube (the interior tube) must be 2 m and 11.43 cm, respectively, and the serpentine tube must be 31.4 m long. In terms of coil count, 83 coils support a volume of water of 0.9 litters, while

250 coils support a volume of water of 2.72 L. According to [12], the lower volume of water leads to a higher outlet temperature of water vapor, which demonstrates the use of 83 spires coiled on the cylindrical tube containing 0.9 L of water (Figs. 2 and 3).
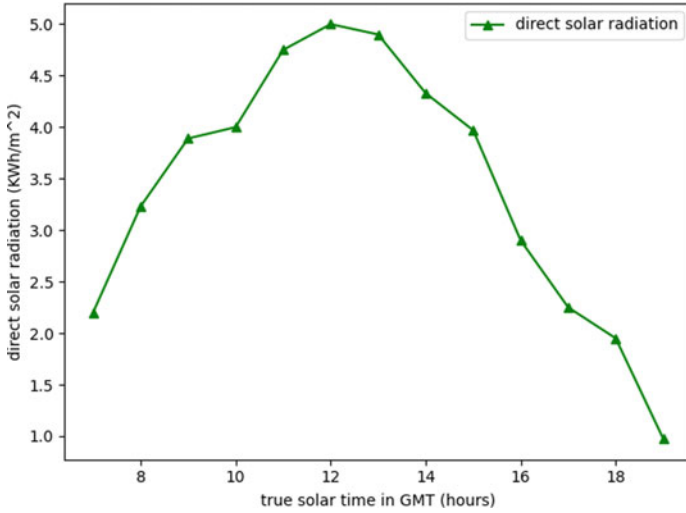
Fig. 2 Representation of the direct solar radiation on Oujda for a typical day: July 25, 2019
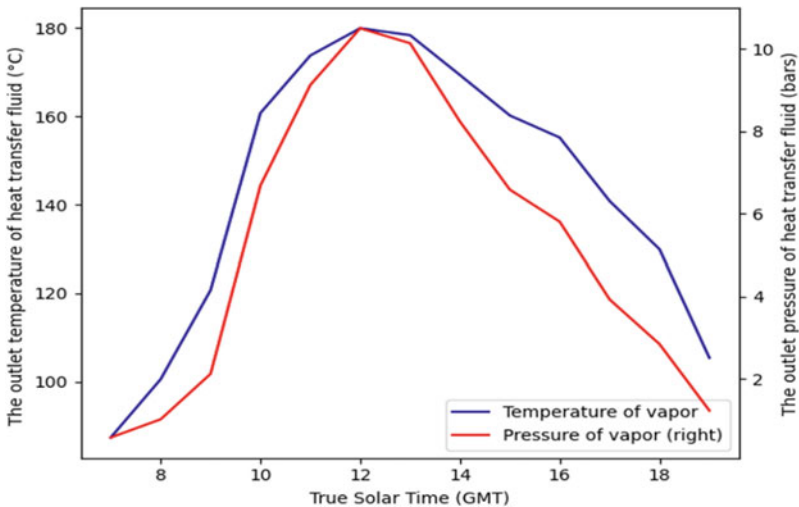
Fig. 3 Representation of the outlet temperature and pressure of water vapor for a typical day in 2019

5 Conclusion and Perspectives

In this study, the outlet parameters such as pressure and temperature were estimated using mathematical modeling of heat and mass transfer through the heat exchanger. In effect, the mathematical modeling enabled the heat exchanger to be dimensioned in order to build the mini solar thermodynamic plant in real plane. Mathematical modeling is not the only method for optimizing and dimensioning the solar thermodynamic mini power plant; our research team is also working on numerical simulation of the thermal process. This work compares the efficiency of a serpentine tube as an absorber to that of a cylindrical tube, as shown in the following work [12]. The fracture mechanics could leave its mark in this work by initiating the cracking of the absorber tube to study its tenacity and the industrial maintenance of the prototype [13].

References

1. Lan H, Gou Z, Xie X (2021) A simplified evaluation method of rooftop solar energy potential based on image semantic segmentation of urban streetscapes. Sol Energy 230:912–924
2. Li J, Huang J (2020) The expansion of China's solar energy: challenges and policy options. Renew Sustain Energy Rev 132:110002
3. Tong G, Chen Q, Xu H (2022) Passive solar energy utilization: a review of envelope material selection for Chinese solar greenhouses. Sustain Energy Technol Assessments 50:101833
4. Nilson RS, Stedman RC (2022) Are big and small solar separate things? The importance of scale in public support for solar energy development in upstate New York. Energy Res Soc Sci 86:102449
5. Güney T (2022) Solar energy, governance and CO_2 emissions. Renew Energy 184:791–798
6. Khoshvaght-Aliabadi M, Feizabadi A (2020) Employing wavy structure to enhance thermal efficiency of spiral-coil utilized in solar ponds. Sol Energy 199:552–569
7. Dumont O, Georges E, Declaye S, Lemort V, Quoilin S Dimensionnement et modélisation d'une microcentrale solaire, Université de Liège – Laboratoire de thermodynamique
8. Dione KhR, Louahlia H, Marion M, Berçaits JL (2018) Evaporation heat transfer and pressure drop for geothermal heat pumps working with refrigerants R134a and R407C. Int Commun Heat Mass Transfer 93:1–10
9. Rovense F, Reyes-Belmonte MÁ, Romero M, González-Aguilar J (2022) Thermo-economic analysis of a particle-based multi-tower solar power plant using unfired combined cycle for evening peak power generation. Energy 240:122798
10. Cantoni R, Rignall K (2019) Kingdom of the Sun: a critical, multiscalar analysis of Morocco's solar energy strategy. Energy Res Soc Sci 51:20–31
11. Li W, Tian X, Li X, Liu J, Li C, Feng X, Shub C, Yu Z-Z (2022) An environmental energy-enhanced solar steam evaporator derived from MXene-decorated cellulose acetate cigarette filter with ultrahigh solar steam generation efficiency. J Colloid Interface Sci 606(Part 1):748–757
12. Latrache F, Hammouch Z, Bellach B, Ghammouri M, Jbira N, Salhi O (2022) Modelling and dimensioning of the heat exchanger for the construction of a solar thermodynamic micro plant in the region of Oujda-Morocco. In: 2022 2nd International conference on innovative research in applied science, engineering and technology, IRASET 2022
13. Benhaddou M, Ghammouri M, Latrache F, Hammouch Z (2020) Study of cleavage in a rectangular plate by the XFEM method and the integral contour J method. Mater Today Proc 27(Part 4):2993–2998

Energy Performances of a Photovoltaic Thermal Plant Using Different Coolant Nanofluid

Stefano Aneli, Antonio Gagliano, Hajji Bekkay, Giovanni Mannino, and Giuseppe M. Tina

Abstract This paper presents a numerical study of a photovoltaic thermal plant that uses coolant fluid water nanofluids. The thermal properties of several investigated water nanofluids are derived from literature data. In particular, Al_2O_3–water (2–4% particle volume fraction) and TiO_2–water (2–6% particle volume fraction) are investigated in this study. The effect of different working fluids on the performances of real PVT plants is simulated through "TRNSYS" software. First, several limitations and constraints needed to be resolved to develop a trustworthy simulation environment within the TRNSYS framework. The great sensitivity to particle volume fraction and temperature of the thermal conductivity of the nanofluid has to be taken into account. The performance in terms of thermal energy produced increases using nanofluids by up to 11%, making the use of nanofluids a promising technology. However, the annual energy benefit in a complete PVT system serving a home is only 2.5% greater than that of using water alone.

Keywords Nanofluid · Numerical model · Thermal energy · Electrical energy · WISC PVT collector

1 Introduction

The use of PVT systems for the simultaneous production of electrical and thermal energy is now well known [1, 2], as PVT technology allows more energy to be produced for the same surface area compared to conventional solar systems [3].

S. Aneli · A. Gagliano (✉) · G. M. Tina
University of Catania, Viale Andrea Doria 6, 95125 Catania, Italy
e-mail: antonio.gagliano@unict.it

H. Bekkay
Mohammed First University, BP 669, 60000 Oujda, Morocco

G. Mannino
CNR–IMM, Strada VIII n. 5 Zona Industriale, Catania, CT, Italy

In PVT systems, the heat transfer fluid plays a fundamental role. As an alternative to water, several studies have been conducted in recent years using nanofluids, i.e., fluids consisting of solid nanoparticles or nanofibers with dimensions typically of 1–100 nm suspended in the liquid. Various studies have shown that nanofluids have substantially higher thermal conductivities than base fluids [4, 5], and therefore, a higher heat removal rate is possible using nanofluids [6].

Several reviews [7, 8] have listed research works using nanofluids as working fluid in PVT systems and discussed the performance obtained with the use of nanofluids as refrigerant in PVT systems. The use of nanofluids in PVT collectors is interesting because any improvement in heat extraction can be accompanied by a secondary benefit in electrical efficiency [9]. However, other studies [10] have shown that the use of nanofluids instead of water improves thermal efficiency but causes a reduction in electrical efficiency due to the higher temperature of the PVT collector. Michael and Iniyan [10] suggest that electrical efficiency could still be improved with better heat exchanger design with high efficiency taking into account the volume concentration of nanofluids. In [11], the authors observed that as the temperature of the inlet fluid varies, the use of nanofluids allows an increase in the thermal level of 2%. In [12], the authors presented a comprehensive review of the technologies used in PVT systems and outlined the challenges and opportunities for design considerations.

This study, unlike those in the literature, analyzes the effect on the performance of the use of nanofluids for a real-size PVT system serving a home. The performance of a PVT system was carried out considering different heat transfer fluids: pure water, nanofluid composed of water and Al_2O_3 nanoparticles, and nanofluid composed of water and TiO_2 nanoparticles. The mathematical model was initially calibrated with data from a real PVT system installed at the University of Catania (IT) [13]. Subsequently, the performance at the level of a single panel is analyzed as the heat transfer fluid varies. Finally, annual analyzes were conducted, where the thermal and electrical energy produced by the plant is used to meet the demands of a home located in Catania.

2 Methodology

This study aims to evaluate the performance of a real PVT system serving a home by considering different heat transfer fluids: pure water and nanofluids.

The analysis is performed using the TRNSYS software and considering the climatic data of the city of Catania, located on the eastern coast of Sicily (IT).

2.1 TRNSYS Model

The numerical model includes uninsulated and unglazed PVT panels, thermal storage, regulation devices and various TRNSYS types used to implement the thermal

and electrical demand [3]. In this TRNSYS project, the PVT panels were simulated using type 942, which permit modifying the characteristic of working fluid (conductivity, density, specific heat and viscosity). This numerical model is based on the PVT system installed at the University of Catania (IT), which is equipped with commercially WISC panels, as presented in [13]. The model was previously calibrated using the experimental analysis [14] and was subsequently enlarged reaching an electrical peak power of 3.0 kW, which is the average electrical power installed in individual homes in Italy.

The simulations are carried out for the whole year, using a timestep of 10 min. The numerical model allows calculating the temperature of the working fluid, the storage's temperature, the electricity produced "E_{el}", the thermal energy used to satisfy the DHW demand "E_{th}", the net electricity "$E_{el,net}$", the coverage factors of demands "f", and the realised self-consumption rate "Rsc". The net electricity is calculated by subtracting from the electricity produced by PVT the electric consumption of the pump. The self-consumption rate indicates how much of the energy produced is self-consumed, considering a consumption for DHW and electricity respectively of 200 l/day and 6000 kWh/y, and the standard hourly profile shown in [15].

$$f = \frac{f_{\text{produced}}}{f_{\text{demand}}} \times 100[\%] \tag{1}$$

$$\text{Rsc} = \frac{E_{el,\text{sefconsumed}}}{E_{el}} \tag{2}$$

2.2 Working Fluids

The study on the performance of a PVT system was carried out considering different heat transfer fluids: pure water, nanofluid composed by water and Al_2O_3 nanoparticles, called $Al_2O_3\_2\%$ and $Al_2O_3\_4\%$, and nanofluid composed by water and TiO_2 nanoparticles, called $TiO_2\_2\%$ and $TiO_2\_6\%$, where the number indicates the percentage in volume of nanoparticles.

The main thermos-physical properties of the nanofluid have been calculated using the equations proposed by [16], but due to the limitations of Maxwell method, the conductivity was estimated using the experimental data shown by [17].

3   Result and Discussions

3.1   Thermal Comparison During Day-Type

In this section, the comparison of performances of the PVT panel using different working fluids is discussed. The simulations were carried out considering a sunny day, flow rate of 60 kg/(h·m²) and constant inlet temperature of 20 °C.

Figure 1 shows the comparison of the outlet temperatures, thermal, and electrical power for the several fluids analyzed.

It can be observed that the outlet temperature is greater using the nanofluid, especially with the high rate of nanoparticles. This temperature increase translates into an increase in the thermal power produced, with differences of more than 5% in the case of using low rates of nanoparticles and differences of about 11% for high flow rates.

It is important to note that, even the thermal performances (thermal level and power) increase with the use of nanofluids, this does not affect the production of electricity; observing Fig. 1 on the right, the differences of power with the variation of the working fluid are negligible. Therefore, the use of nanofluids in PVT panels can be considered promising because it improves their thermal performance without compromising electrical performance.
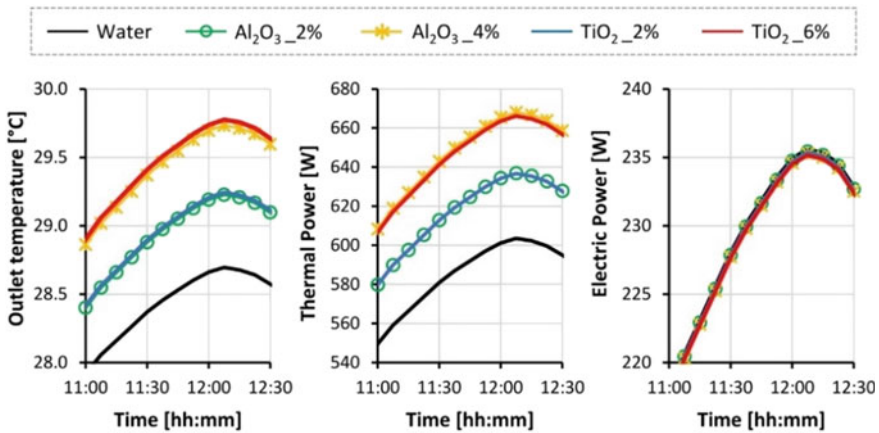
Fig. 1 Outlet temperatures (at left), thermal power (at center) and electric power (at right)

3.2 Impact of Nanofluid on the Performances of PV/T System During the Whole year

In this section, the performances of various coolant fluids in a real plant are analyzed considering the whole year.

Figure 2 shows the weather data (ambient temperature and irradiance on the panel) and the comparison of average storage tank temperature using several working fluids: pure water (black), nanofluid composed by water and 4% by weight of Al_2O_3 (yellow) and nanofluid composed by water and 6% by weight of TiO_2 (red) for a summer day. It is important to point out that the tank temperature is an indicator of the fulfilment of the thermal demand.

The use of nanofluids as cooling fluid allows for achieving temperatures in the tank higher than using pure water. Consequently, it is possible to fulfil a greater share of the thermal demand through the PVT plant. The performance increase occurs both during summer and winter.

Table 1 shows the annual performances for the several working fluids studied.
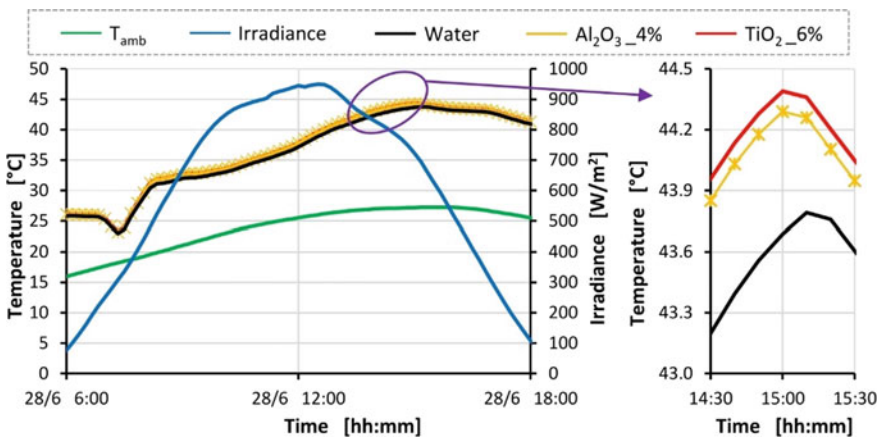
Fig. 2 Environment conditions and average tank temperature of the several working fluids during a summer day

Table 1 Annual results

Working fluid	E_{el} (kWh)	$E_{el,net}$ (kWh)	E_{DHW} (kWh)	f_{el} (%)	Rsc (%)	f_{DHW} (%)	$f_{overall}$ (%)
Water	5250.3	4995.3	1127.9	83.45	52.25	53.23	75.55
Al_2O_3_2%	5247.5	4996.5	1139.4	83.47	52.24	53.77	75.70
Al_2O_3_4%	5245.5	4998.3	1151.2	83.50	52.22	54.33	75.87
TiO_2_2%	5247.2	4995.0	1137.5	83.44	52.26	53.68	75.66
TiO_2_6%	5242.7	4995.5	1156.0	83.45	52.25	54.55	75.90

Observing the electricity produced, it can be noted that the use of nanofluid decrease the annual production, especially for large amounts of nanoparticles. This is because the increase in temperatures reduces, albeit slightly, the electrical productivity. Differently, the net electrical energy produced using the nanofluid is greater respect the use of pure water, because the better thermal performance of nanofluid allows to reduce the working time of pump and consequently the reduction of the electrical consumption of the pump. The little increase of net energy production using nanofluid generates almost negligible increases in the coverage factor of electricity demand. The same increase in electricity produced does not lead to an improvement in the RSC because it does not improve the self-consumed portion. The use of nanofluids, on the other hand, generates an increase in the thermal energy produced, which can be used to satisfy the demand for DHW, with differences that reach a maximum of 2.5%. Therefore, unlike what we saw in Sect. 3.1 for a single panel, where the increases in terms of thermal power produced reached 11%, the effect of the nanofluids on the whole system is greatly reduced, reaching a maximum of 2.5%. This is due both to the increase in system losses (panels, tank, etc.) and also to the need to modify other aspects of the system, such as flow rates or accumulation volumes. Finally, the total demand coverage factor improves using nanofluids, but the improvement is almost negligible.

4 Conclusions

The paper shows the effects of changing the coolant fluid from pure water to a nanofluid composed of water and aluminum oxide or water and titanium oxide, in a complete PVT system serving a home.

The simulations show that the use of nanofluids in only PVT panels can be considered promising because it improves their thermal performance without compromising electrical performance. On the other hand, the use of nanofluids in a complete PVT system generates a negligible effect in terms of electricity produced and an increase in the DHW thermal energy of a maximum of 2.5%. Therefore, unlike what we highlighted in the analysis for a single PVT panel, where the increases in terms of thermal power produced reached 11%, the effect of the nanofluids on the whole system is greatly reduced. This is due to the increase in system losses (panels, tank, etc.) and highlights that the use of nanofluids needs to modify other aspects of the system, such as flow rates or accumulation volumes.

References

1. Chandrasekar M, Senthilkumar T (2021) Five decades of evolution of solar photovoltaic thermal (PVT) technology—a critical insight on review articles. J Clean Prod 322:128997
2. Joshi SS, Dhoble AS (2018) Photovoltaic-thermal systems (PVT): technology review and future trends. Renew Sustain Energy Rev 92:848–882
3. Gagliano A, Tina GM, Aneli S, Nižetić S (2019) Comparative assessments of the performances of PV/T and conventional solar plants. J Clean Prod 219:304–315
4. Lenin R, Joy PA, Bera C (2021) A review of the recent progress on thermal conductivity of nanofluid. J Mol Liquids 338:116929
5. Coccia G, Tomassetti S, Di Nicola G (2021) Thermal conductivity of nanofluids: a review of the existing correlations and a scaled semi-empirical equation. Renew Sustain Energy Rev 151:111573
6. Mikkola V, Puupponen S, Granbohm H, Saari K, Ala-Nissila T, Seppälä A (2018) Influence of particle properties on convective heat transfer of nanofluids. Int J Therm Sci 124:187–195
7. Diwania S, Agrawal S, Siddiqui AS, Singh S (2020) Photovoltaic–thermal (PV/T) technology: a comprehensive review on applications and its advancement. Int J Energy Environ Eng 11:33–54
8. Abbas N, Awan MB, Amer M, Ammar SM, Sajjad U, Ali HM, Zahra N, Hussain M, Badshah MA, Jafry AT (2019) Applications of nanofluids in photovoltaic thermal systems: a review of recent advances. Phys A: Stat Mech Appl 536:122513
9. Salari A, Taheri A, Farzanehnia A, Passandideh-fard M, Sardarabadi M (2021) An updated review of the performance of nanofluid-based photovoltaic thermal systems from energy, exergy, economic, and environmental (4E) approaches. J Clean Prod 282:124318
10. Michael JJ, Iniyan S (2015) Performance analysis of a copper sheet laminated photovoltaic thermal collector using copper oxide–water nanofluid. Sol Energy 119:439–451
11. Aneli S, Gagliano A, Tina GM, Hajji B (2021) Analysis of the energy produced and energy quality of nanofluid impact on photovoltaic-thermal systems. In: International conference on electronic engineering and renewable energy, pp 739–745
12. Huang G, Curt SR, Wang K, Markides CN (2020) Challenges and opportunities for nanomaterials in spectral splitting for high-performance hybrid solar photovoltaic-thermal applications: a review. Nano Mater Sci 2:183–203
13. Gagliano A, Tina GM, Nocera F, Grasso AD, Aneli S (2019) Description and performance analysis of a flexible photovoltaic/thermal (PV/T) solar system. Renew Energy 137:144–156
14. Gagliano A, Aneli A (2021) Chapter 2—Energy analysis of hybrid solar thermal plants (PV/T). Recent advances in renewable energy technologies, vol 1, pp 45–90
15. Gagliano A, Tina GM, Aneli S, Chemisana D (2021) Analysis of the performances of a building-integrated PV/Thermal system. J Clean Prod 320:128876
16. Avsec J, Oblak M (2007) The calculation of thermal conductivity, viscosity and thermodynamic properties for nanofluids on the basis of statistical nanomechanics. Int J Heat Mass Transf 50:4331–4341
17. Longo GA, Zilio C (2011) Experimental measurement of thermophysical properties of oxide–water nano-fluids down to ice-point. Exp Thermal Fluid Sci 35:1313–1324

Simulation and Comparative Study of the Effect of the Wet and Dry Cooling Modes on the Performance of Parabolic Trough Solar Thermal Power Plants in the Arid Zone of Morocco

Hanane Ait Lahoussine Ouali, Mujeeb Iqbal Soomro, Samir Touili, and Ahmed Alami Merrouni

Abstract The primary goal of this research is to evaluate the competitiveness of technology in concentrated solar power (CSP) plant under the weather conditions of Ouarzazate city located in the south of Morocco. Thus, a techno-economic comparison of wet and dry cooling modes in a parabolic trough plant (PTC) using molten salt as a thermal storage medium was performed using the System Advisor Model (SAM) software. This research is based on an evaluation of thermodynamic yield, water cooling demand, and the levelized cost of electricity (LCOE). The results demonstrate that the dry cooling option can make a significant reduction in water consumption by almost 94% in the PTC plant. However, the annual energy and capacity factor have been dropped by around 7.1% when compared to the PTC plant with a wet (evaporative) cooling mode. Because of the increase in equipment costs for dry a cooling system, the actual and nominal LCOE increased by 7.3%. Based on these findings, we can infer that dry cooled solar PTC plants are technically and economically feasible for Ouarzazate city, which has an arid climate and a shortage of water, and we recommend that this technology be implemented in similar places.

Keywords Dry cooling · Wet cooling · Concentrating solar power · Morocco · SAM software

H. Ait Lahoussine Ouali (✉) · S. Touili · A. Alami Merrouni
Materials Science, New Energies and Application Research Group, LPTPME Laboratory, Department of Physics, Faculty of Science, University Mohammed First, Oujda, Morocco
e-mail: hananeaitlahoussine@gmail.com

M. Iqbal Soomro
Department of Mechanical Engineering, Mehran University of Engineering and Technology, SZAB Campus, Khairpur Mir's 66020, Sindh, Pakistan

1   Introduction

Energy is the backbone of the economic development of countries where the continuous increase in the energy demand in several sectors (like, industry, transportation, and building) presents a real issue for countries with short fossil sources. In addition, for a global sustainable development, the world is currently shifting to the use of renewable energy sources for achieving sustainable development. In this context, Morocco is highly investing in the integration of renewable sources in its energy mix, taking the advantage of its high potential in solar and wind. For this, Morocco has launched an ambitious initiative to boost renewable energies, supporting energy efficiency and achieving a renewable energy mix of at least 52% by 2030. Among the most prominent of these projects, is the Noor Ouarzazate Solar Complex with a capacity of 580MWe [1] from a mix between 500 MWe through concentrated solar power (CSP) technology and 70 MWe from PV technology. Generally, the CSP plants consist of four technologies: parabolic trough collector [2], solar power tower [3], linear Fresnel reflector [4], and parabolic dish [5].

Several studies in the literature compared various CSP configurations with various heat transfer fluids (HTFs). For example, Turchi et al. [6] have used the solar advisor model program to evaluate the influence of dry and wet cooling systems in parabolic trough power facilities. The results reveal that the cost of energy and investment rises by 3–8% when switching from wet to dry cooling. Nevertheless, it reduces the water consumption by more than 90%. Boukelia et al. [7] examined the competitiveness of deploying the dry cooling mode in parabolic trough power plants using thermic oil and molten salt as HTFs. It was found that using the dry cooling decreases the yields for oil and salt configurations down to 8.7% and 9.3%, respectively. However, the dry cooling mode is reducing water usage by more than 94% for both configurations of the plants. In the Ma'an area of southern Jordan, Liqreina et al. [8] explored a parabolic trough plant using dry cooling technology. The obtained results show that the dry cooled plant in Ma'an's overall efficiency has decreased by 3.1%, the energy output has increased by 21.8%, the LCOE has decreased by 18.8%, and there is a possibility of saving roughly 92% of water demand. In another study, Martín [9] optimized the operation of CSP plant using dry cooling option using the climatic data of Almería city (Spain). As a result, a multi-period mixed-integer nonlinear mathematical formulation with several periods was adopted. The findings show that the average energy output is 17.2 MW, the cooling system consumes 5% of the energy, and the production cost is 0.16 €/kWh. Under the Tunisian weather conditions, Trabelsi et al. [10] conducted a study on PTC plants with dry cooling systems using two types of heat transfer fluids. The research is carried out by adjusting the size of the solar field and the thermal energy storage system to find the ideal combination for the lowest power production cost. They found that the cost of electricity for a plant adopting molten salt as HTF is 13.8 c€/kWhe, compared to 17.24 c€/kWhe for a plant using Therminol VP-1 as HTF.

It's worth noting that PTC technology has been widely commercialized and is the most reliable and bankable configuration for CSP facilities across the world. As a

result, this article explores the impacts of employing the two cooling configurations (wet and dry) in a PTC plant in terms of thermodynamic yield, water consumption, and economic indicators under the meteorological conditions of Ouarzazate, Morocco.

2 Methodology and Simulation Inputs

2.1 Description of the PTC System

On account of its maturity and industrial development, a parabolic trough technology (PTC) employing synthetic oil as an HTF was chosen for this investigation. The proposed PTC system is organized into three primary subsystems: (i) a solar field consisting of collector loops; (ii) a thermal energy storage (TES) consisting of two tanks of molten salt storage (60% $NaNO_3$ and 40% KNO_3); and (iii) a power block employing the Rankine power cycle. Figure 1 displays a flow schematic diagram of the PTC system under investigation.
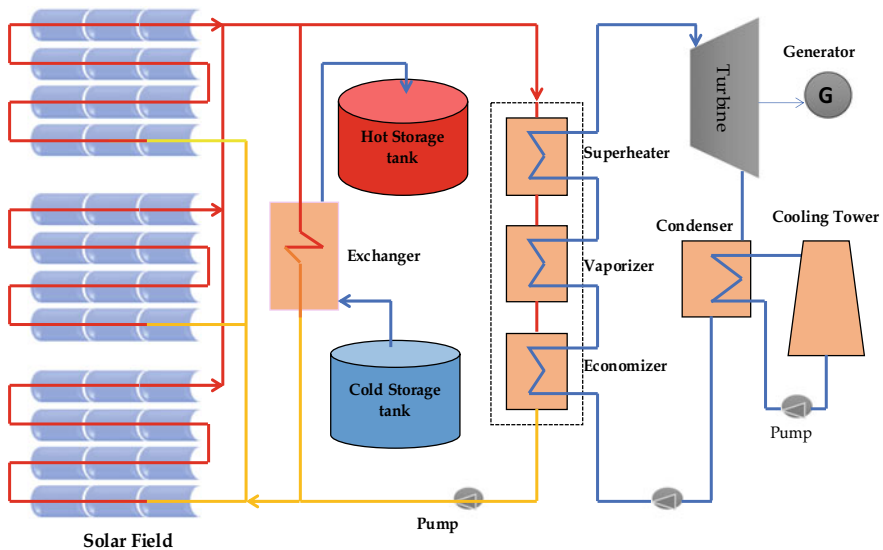
Fig. 1 Diagram of the PTC facility under investigation

2.2 PTC Model

The PTC plant's solar field is made up of a number of collectors that reflect incoming solar irradiation to a focal line and then to a receiver where the HTF (Therminol VP1) is heated. The reflected solar power may be estimated as the following formula (1) [11].

$$Q_{fd} \quad DNI \quad \cos\theta_i \quad A_{field} \tag{1}$$

where DNI: direct normal irradiation, A_{field}: effective aperture area, and θ_i: angle of incidence of collector. The absorbed power by the receiver tubes can be evaluated by the following Eq. (2) [12]:

$$Q_{rec} = Q_{fd} * \eta_{col} * \eta_{rec} * k_\theta * \eta_{shadow} * \eta_{end-losses} * \eta_{track} \tag{2}$$

where $\boldsymbol{\eta}_{col}$ is the optical efficiency, $\boldsymbol{\eta}_{rec}$ is the nominal receiver efficiency, $k_{\boldsymbol{\theta}}$ is the incidence angle modifier, $\boldsymbol{\eta}_{end-losses}$: is the end losses factor, $\boldsymbol{\eta}_{shadow}$ is the row shadowing element, and $\boldsymbol{\eta}_{track}$ is the field tracking factor for solar collectors.

The power collected by the HTF over the field may be calculated using Eq. (3).

$$Q_{collected} = Q_{receiver} - Q_{receiver-losses} - Q_{pipe-losses} \cdot \tag{3}$$

where $Q_{receiver}$ is the absorbed power by the receiver, $Q_{receiver-losses}$: Receivers' heat loss, and $Q_{pipe-losses}$: Piping heat loss.

- LCOE

The levelized cost of electricity (LCOE) is the key metric for investors enabling the assessment of the project's profitability, as well as, the feasibility of a plant configuration/technology in a specific site. The utility-scale LCOE in this investigation is estimated based on the required revenues over the project life and is named as "nominal" and "real" giving by the following equations [13]:

$$\text{nominal LCOE} = \frac{\sum_{n=1}^{N} \frac{R_n}{(1+DR_{nominal})^n}}{\sum_{n=1}^{N} \frac{Q_n}{(1+DR_{nominal})^n}} \tag{4}$$

$$\text{real LCOE} = \frac{\sum_{n=1}^{N} \frac{R_n}{(1+DR_{nominal})^n}}{\sum_{n=1}^{N} \frac{Q_n}{(1+DR_{nominal})^n}} \tag{5}$$

where R_n is the necessary income from power sales in year n, Q_n is the amount of power generated in year n, $DR_{nominal}$ is the nominal discount rate, and DR_{real} is the real discount rate. The nominal LCOE, as can be seen in Eq. (4) and (5), is a current dollar value that is better suited to short-term assessments. The real LCOE, on the

other hand, is a constant dollar (inflation-adjusted) value that is better for long-term assessments. In this study, the plant's lifespan is estimated to be 25 years.

2.3 Solar Plant Design and Simulation Inputs

The System Advisor Model (SAM) program [14] was used in this research to evaluate the techno-economics competitiveness of dry and wet cooling modes for a 50 MWe solar plant based on PTC technology. For pre-feasibility studies, this software is extensively used. It predicts performance and estimates energy costs for grid-connected electrical projects using installation and operating costs, as well as system design characteristics that we provide as inputs. Table 1 displays the design parameters of the suggested PTC facility.

3 Field of Study and Solar Resource Assessment

Because solar resource data has a substantial influence on the potential of concentrated solar power systems[17], it is required to present and debate the quantity of solar resource obtained by our field of research. Accordingly, we used a Typical Meteorological Year (TMY) data with one-hour interval dataset of Ouarzazate city (latitude 30.91°N, longitude −6.89°E, altitude 1113 m). Ouarzazate is located in the south of Morocco which is characterized by an arid climate and receives a high solar radiation. Figure 2 depicts a heat map of direct normal irradiation (DNI) at the chosen site. As can be observed, the highest DNI value is always around noon. Besides, it is noticeable that the city of Ouarzazate receives a significant amount of solar radiation for the majority of the day, during the year, with an average DNI value of 6.70 kWh/m^2/day. Furthermore, the temperature distribution may be shown in Fig. 3 histogram. It is apparent that around 6323 h with temperature between 10 and 30 °C around 290 h occur with temperature over 35 °C, and only 330 h out of 8760 h of the year had temperatures below 5 °C. The use of dry cooling mode in Ouarzazate city is a technically possible alternative, according to meteorological data.

4 Results and Discussion

4.1 Technical Comparison

Figures 4 and 5 illustrate the system performance parameters for each month of the year, including field incident thermal power, cycle thermal power input, and gross

Table 1 Main input data and assumption using for the investigated solar plant [15, 16]

Parameter	Value
Technical data	
Design irradiation	950 W/m^2
Solar collector type	Euro trough ET150
Aperture of the field	510 120m^2
Number of loops	78
Row spacing	15 m
Thermal receiver type	Schott PTR80
Absorber tube inner/outer diameter	0.076 m/0.08 m
HTF type	Therminol VP1
Field HTF minimum operating	12 °C
Field HTF maximum temperature	400 °C
Design inlet temperature	391 °C
Design outlet temperature	293 °C
Amount of water used per wash	0.7 L/m^2 aperture
Number of washes	63/year
Cycle type	Rankine cycle
Design gross output	50MWe
Thermal storage fluid	60% NaNO$_3$ and 40% KNO$_3$
Storage capacity	7.5 h
Economic data	
Site	25 \$/m^2
Solar field	150 \$/m^2
HTF	60 \$/m^2
Power block	910 \$/kWh$_{el}$
TES system	62 \$/kWh$_{th}$
Contingency	7 \$/kWh$_{el}$
Fixed cost by capacity	66\$/kWh_yr
Variable cost by generation	4\$/MWh

electric power. According to the figures, the month of May had the greatest field incident thermal power of 436.03 MW$_{th}$ for both cooling systems, and the highest gross electric power output from the power block for wet and dry systems are 56.01 and 53.42 MWe, respectively, which represents a decreasing about 4.8%. Likewise, Fig. 6 illustrates the monthly energy generation from the investigated parabolic trough collector plants in Ouarzazate city. As can be observed from the figure, the monthly energy output by the PTC plant peaks in May, reaching 23.81 GWhe and 22.03 GWhe for wet and dry configurations, respectively. While December produced the least amount of energy, with 7.74 GWhe and 7.46 GWhe for wet and dry options,
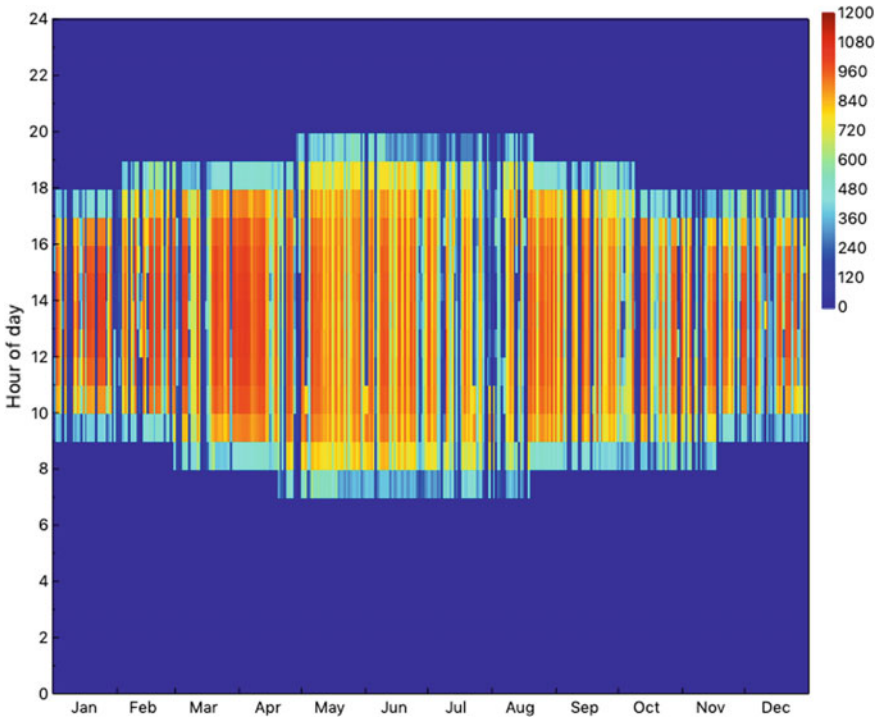
Fig. 2 2D heat map of DNI in the studied site
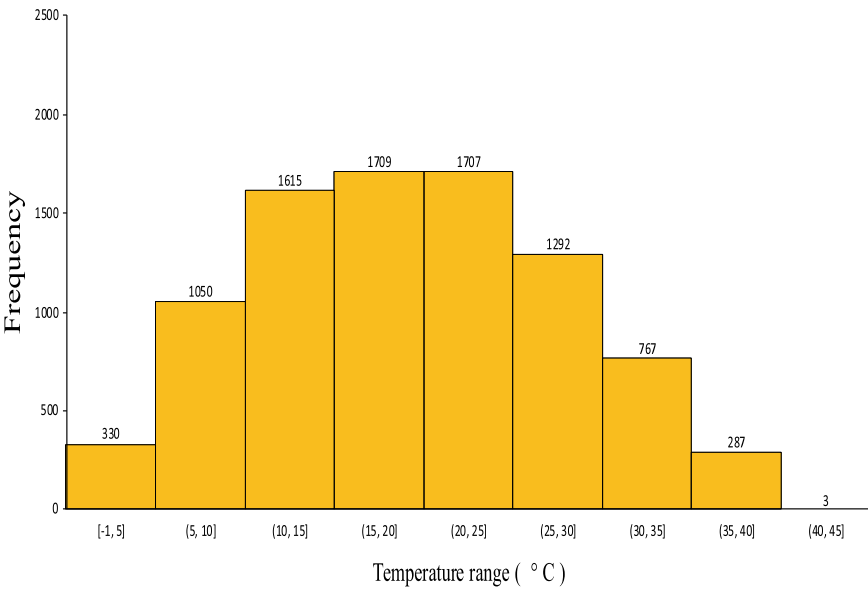
Fig. 3 Histogram of ambient temperature in the studied site
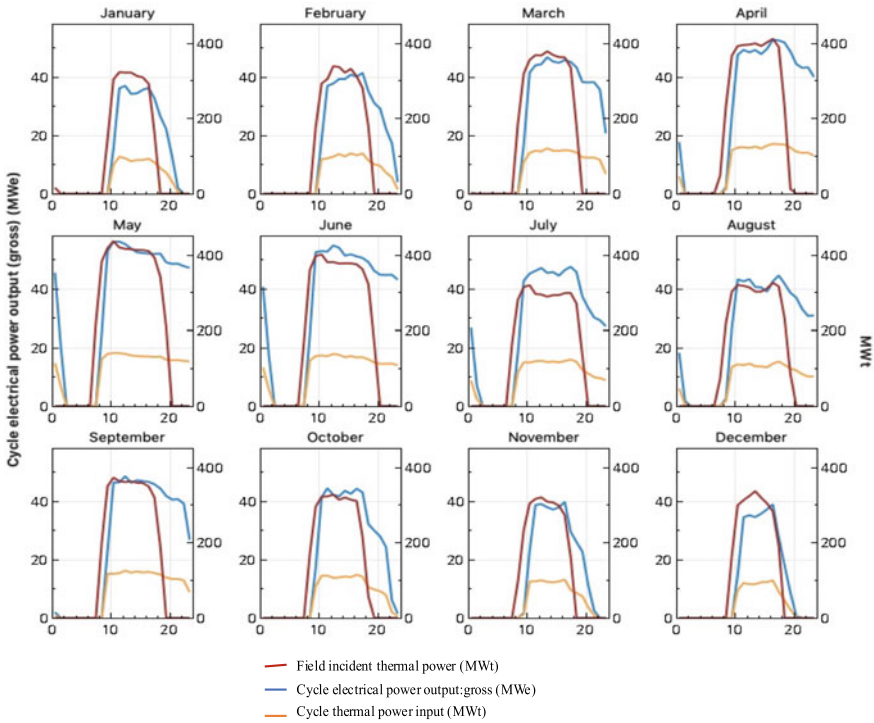
Fig. 4 Simulation results for wet a cooling system

respectively. Concerning the yearly energy output by PTC facilities are estimated to be 168.47 GWh and 181.20 GWh for dry and wet cooling options, respectively, which represents an increasing about 7.3%. Likewise, the annual capacity factor is 46% for the wet cooling option instead of 42.7% for dry cooling options.

4.2 *Water Consumption*

Figure 7 depicts the wet and dry cycle cooling water mass flow rate as a function of time over a day during the May and December months. The figure clearly shows that cycle water demand is higher in May than in December, which is dependent on climatic conditions such as DNI. Likewise, the cooling water is very important during the period between 11 AM and 16 PM which corresponds to the strong period of sunshine. Furthermore, Table 2 lists the monthly expected water consumption included cycle cooling and water washing mirrors in the investigated solar facility for the two cooling options. As can be shown, the yearly water consumption employing dry cooling option is 39,539 m^3 as opposed to 655,813 m^3 with a similar PTC facility working with wet cooling mode. Dry cooling mode saves around 93.8% of water,
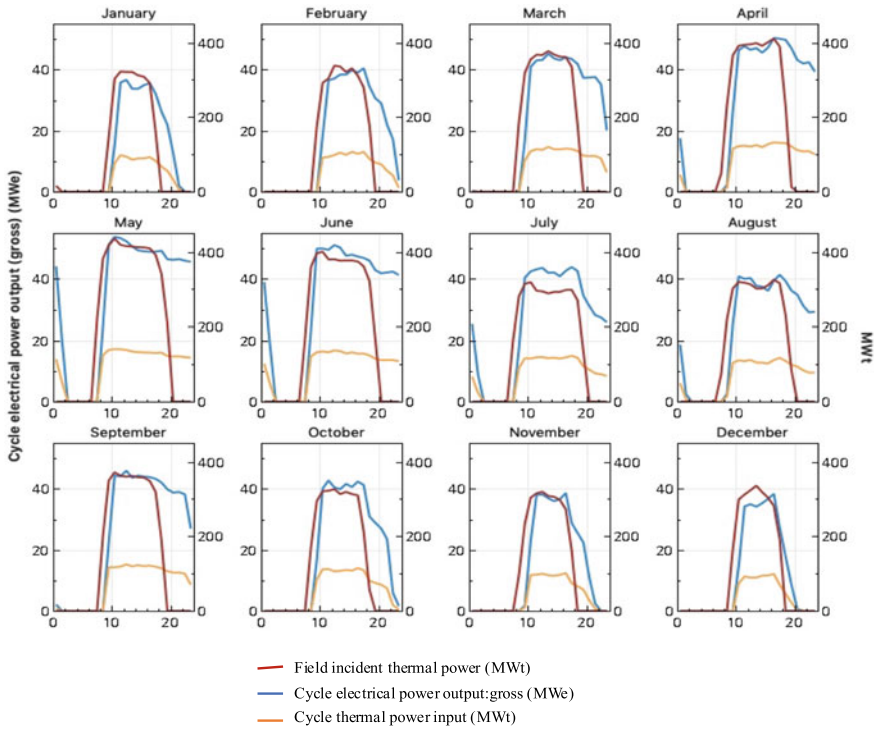
Fig. 5 Simulation results for dry a cooling system
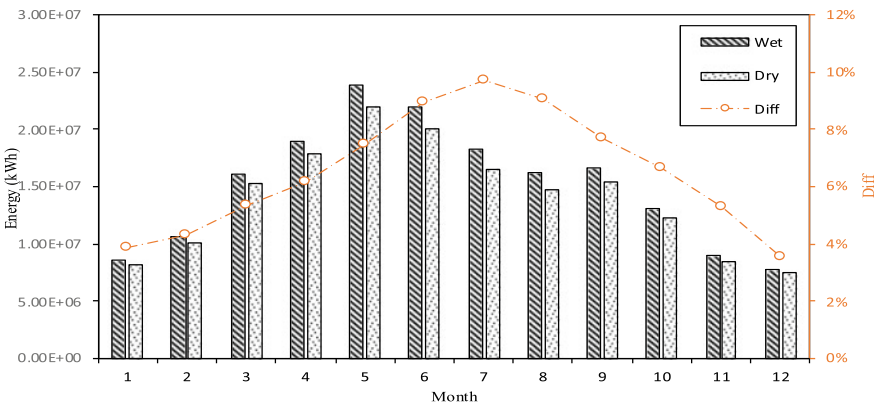
Fig. 6 Monthly energy generation for two cooling systems
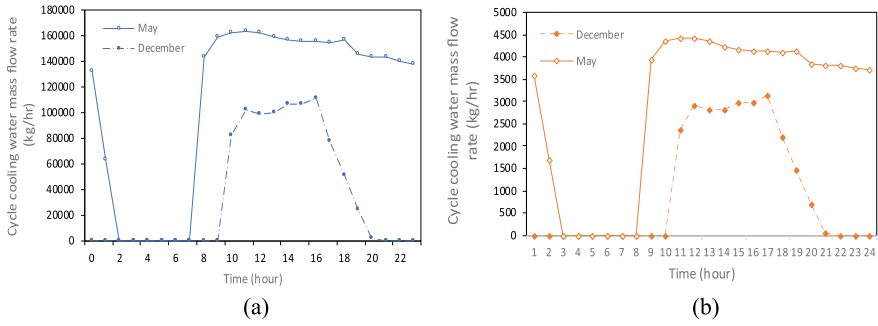
Fig. 7 Cycle cooling water mass flow rate: **a** wet option and **b** dry option

Table 2 Predicted water usage for the two cooled systems

Month	Wet cooled (m^3)	Dry cooled (m^3)	Reduction (%)
January	31,088.87	1937.08	93.77
February	38,363.44	2384.68	93.78
March	58,269.00	3580.53	93.86
April	68,764.66	4189.72	93.91
May	86,136.79	5171.95	94.00
June	79,622.24	4694.73	94.10
July	65,869.31	3866.12	94.13
August	58,992.84	3472.55	94.11
September	60,440.52	3612.77	94.02
August	47,411.42	2879.85	93.93
November	32,500.35	1996.14	93.86
December	28,048.74	1752.96	93.75

which might be an appealing choice in desert places with limited water supply, such as our case study (Ouarzazate city). It should be noted that the DNI parameters for the location stated in Fig. 3 have a significant impact on the yearly energy and water usage.

4.3 Economic Feasibility Analysis

The economic evaluation of the investigated PTC plant will be based on the real and nominal LCOE as shown in the figure below. As shown in the figure, the real and nominal LCOE have increased from 11.77 c$/kWh and 14.83 c$/kWh for the PTC facility with wet cooling mode, respectively, to 12.63 c$/kWh and 15.91 c$/kWh for the same facility with dry cooling mode, representing a 7.3% increase. This is due
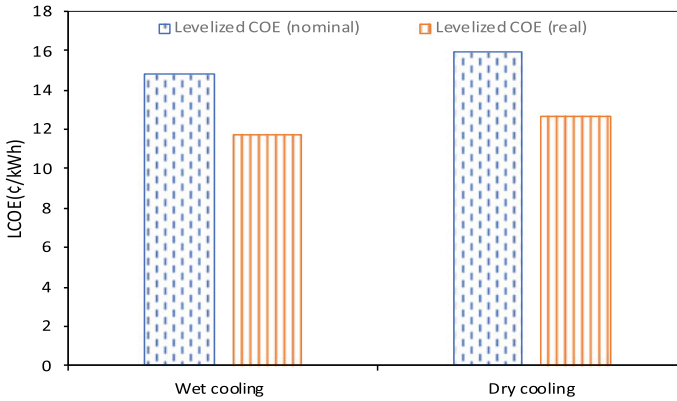
Fig. 8 Real and nominal LCOE for the studied PTC facility

to a decline in yearly electricity generation (168.47 GWhe for the dry cooling rather than 181.20 GWhe for wet cooling) and the other hand the increase of the equipment cost of this solar facility with dry cooling as compared with the same facility with wet equipment (Fig. 8).

5 Conclusion

The study provided a preliminary analysis for comparing the techno-economic of the two cooling configurations based on parabolic trough facility with thermal energy storage. Using SAM software, we will use Ouarzazate, a city in Morocco's southern area, as a case study. The key conclusions may be summed up as follows:

- The annual energy production yield of 168.47 GWh for dry cooling option instead of 181.20 GWh wet cooling option.
- The real and nominal LCOE of dry cooled PTC plant in Ouarzazate site is found to be economically attractive and does not exceed 12.63 and 15.91 c$//kWh, respectively.
- The dry cooling system can make a significant reduction in water consumption by almost 94% in the PTC plant in the arid climate, which is quite advantageous in terms of water conservation.

References

1. Masen [Internet] [cited 2022 Jan 5]. Available from https://www.masen.ma/en
2. Ait Lahoussine Ouali H, Guechchati R, Moussaoui MA, Mezrhab A (2017) Performance of parabolic through solar power plant under weather conditions of the Oujda city in Morocco. Appl Sol Energy (English Transl Geliotekhnika) 53(1):45–52
3. Ouali HAL, Moussaoui MA, Mezrhab A, Naji H (2020) Comparative study between direct steam generation and molten salt solar tower plants in the climatic conditions of the eastern Moroccan region. Int J Renew Energy Dev. 9(2):287–294
4. Alami Merrouni A, Amrani AI, Lahoussine Ouali HA, Moussaoui MA, Mezrhab A (2017) Numerical simulation of Linear Fresnel solar power plants performance under Moroccan climate. J Mater Environ Sci 8(12):4226–4233
5. Ait Lahoussine Ouali H, Raillani B, Amraqui S, Moussaoui MA, Mezrhab A (2020) Performance assessment of solar dish-stirling system for electricity generation in eastern Morocco. Lect Notes Electr Eng. 684:244–252
6. Turchi CS MJW and CFK. Water use in parabolic trough power plants: summary results from Worley Parsons' analyses
7. Boukelia T, Mecibah M, Laouafi A, Mekroud A (2017) Potential assessment of using dry cooling mode in two different solar thermal power plants. Int J Energy 2(2) :18–24. [Internet]. 2017 Dec 31 [cited 2022 Jan 4]. Available from https://www.ijeca.info/index.php/IJECA/article/view/38
8. Liqreina A, Qoaider L (2014) Dry cooling of concentrating solar power (CSP) plants, an economic competitive option for the desert regions of the MENA region. Sol Energy 103:417–424
9. Martín M, Martín M (2017Sep) Cooling limitations in power plants: Optimal multiperiod design of natural draft cooling towers. Energy 135:625–636
10. Trabelsi SE, Qoaider L, Guizani A (2018Jan) Investigation of using molten salt as heat transfer fluid for dry cooled solar parabolic trough power plants under desert conditions. Energy Convers Manage 15(156):253–263
11. Wang J, Wang J, Bi X, Wang X (2016) Performance simulation comparison for parabolic trough solar collectors in China. Int J Photoenergy 2016
12. Llorente García I, Álvarez JL, Blanco D (2011) Performance model for parabolic trough solar thermal power plants with thermal storage: comparison to operating plant data. Sol Energy 85(10):2443–2460
13. Darling SB, You F, Veselka T, Velosa A (2011) Enhanced oxygen reduction activity on surface-decorated perovskite thin films for solid oxide fuel cells. Title: Light scattering by nanostructured anti-reflection coatings Assumptions and the levelized cost of energy for photovoltaics. Energy Environ Sci [Internet]. [cited 2021 Dec 23]; 4(9). Available from www.rsc.org/ees
14. Home-System Advisor Model (SAM) [Internet]. [cited 2022 Jan 5]. Available from: https://sam.nrel.gov/
15. Azouzoute A, Alami Merrouni A, Touili S (2020) Overview of the integration of CSP as an alternative energy source in the MENA region. Energy Strateg Rev 29:100493
16. El Boujdaini L, Ait Lahoussine Ouali H, Mezrhab A, Moussaoui MA (2019) Techno-economic investigation of parabolic trough solar power plant with indirect molten salt storage. Proceedings of 2019 International Conference on Computing Science Renewable Energies, ICCSRE 2019, pp 1–7
17. Merrouni AA, Ghennioui A, Wolfertstetter F, Mezrhab A (2017) The uncertainty of the HelioClim-3 DNI data under Moroccan climate. AIP Conf Proc 27:1850

CFD and Wake Analysis of the Wind Flow Through Two Wind Turbines

Diogo Silva ⓘ**, João Silva** ⓘ**, Paulo Pinto** ⓘ**, Senhorinha Teixeira** ⓘ**, and José Teixeira** ⓘ

Abstract In a wind park, the wake effects cause a decrease in the power output and an increase in the loads of the downstream turbines in comparison to those facing freestream flows. Hence, it is very important to understand and analyse the wake characteristics of a wind turbine and the effects on how a wake impacts the performance of a downstream turbine to rearrange the multiple wind turbines in a reasonable and efficient way. In this way, the objective of this study is to analyse the effects of placing a turbine in front of another at six times the rotor's diameter and evaluate its energy production. This was accomplished with a CFD procedure developed in ANSYS Fluent. The results were promising, in which a 28% loss in power in the downstream turbine due to the wake effects produced by the upstream turbine was detected.

Keywords ANSYS fluent · CFD · Horizontal axis wind turbine · Wake effects · Wind energy

1 Introduction

A wind farm is a group of wind turbines placed in the same area with the purpose of producing electricity. These parks can contain up to several hundred turbines or be simpler and include ten or fifty turbines. Regardless of the number of turbines, the

D. Silva (✉)
Department of Mechanical Engineering, University of Minho, 4800-058 Guimaraes, Portugal
e-mail: a80914@alunos.uminho.pt

J. Silva · J. Teixeira
MEtRICs Research Centre, University of Minho, 4800-058 Guimaraes, Portugal

P. Pinto
MEGAJOULE, 4465-171 Sao Mamede de Infesta, Portugal

S. Teixeira
ALGORITMI Research Centre, University of Minho, 4800-058 Guimaraes, Portugal

placement of each of these structures is a crucial aspect to maximize the output of the wind farm.

To maximize the amount of energy generated by a wind farm, the wind turbines are spaced between each other in the calculated method. If the spacing between the turbines is too high, the wind farm will occupy a larger area and the cable lengths connecting each structure to the electric grid will increase. If the spacing is too low, the downstream turbines will drastically be affected by the upstream ones causing a serious reduction in efficiency. As such, the spacing between turbines must be optimized in terms of compactness of the wind park and in terms of minimizing the energy loss due to wind shadowing from the upstream turbines [1]. As such, the wake effects produced by wind turbines are a very important topic to study when trying to maximize the production of a wind farm. Due to this fact, many researchers and engineers utilize Computational Fluid Dynamics (CFD) techniques to analyse the best wind turbine arrangements for specific wind flows.

Sawant et al. [2], O'Brien et al. [3] and Miller et al. [4] performed an overall and broad review on various wind turbines and wind flows topics, in which some special attention was given to issues such as CFD approaches and wake effects. Chang et al. [5] produced an approach to model the Atmospheric Boundary Layer (ABL) with reasonable results by implementing a RANS turbulence model onto a flat terrain. Richards and Hoxey [6] performed a crucial study which allowed to produce suitable boundary conditions for wind engineering models using the RANS equations as the method to model turbulence, which were later revisited by Richards and Norris [7]. These modelling proposals were also used by Hargreaves and Wright [8] to create a neutral ABL flow over a flat terrain, producing adequate results. Tang et al. [9] performed various tests to analyse wake effects produced by wind turbines, allowing to verify the decrease of the wind velocity and increase of turbulence upon contact with the rotor. Uchida et al. [10] used a Large Eddy Simulation (LES) CFD approach coupled with a porous disc representation of the rotor as to study the variation of the downstream velocity of the wind, creating therefore a method capable of reproducing accurately the flow characteristics. Barthelmie et al. [11] assessed the state of the art in wake and flow modelling for offshore wind farms and performed a comparison study between CFD results and real measurements from offshore wind farms. Porté-Agel et al. [12] carried out a review of various researches of ABL flows with wind farms and turbines, displaying results of turbulence effects generated by wakes produced by wind turbines.

This study is centred around the analyses of various parameters, namely the velocity variations of the wind, pressure drop at the rotor and turbulence produced by wake shadowing. This was accomplished by means of a CFD approach that considers a neutral ABL and simplified wind turbine rotors modelled as actuator discs. For this purpose, two wind turbines were placed in the same oncoming wind direction, distanced six times the rotor's diameter (6D).

2 Numerical Model

To develop a CFD procedure capable of modelling the wind flow through two horizontal axis wind turbines distanced six times their rotor diameter in the prevailing wind direction (6D), many different parameters must first be stated. These involve topics such as turbulence modelling, computational domain, boundary conditions, rotor modelling, solver configuration and discretization scheme.

For the simulation, a 3D steady-state incompressible and isothermal flow was assumed, and the RANS equations along with the standard k-ε model with some modifications in the turbulence constants were solved [13].

The computational domain consists in a rectangular prism shape, with a length of 2000 m, a width of 300 m and a height of 458 m, containing a mesh with roughly 6 million elements, as shown in Fig. 1. The first wind turbine is placed at $x = 800$ m, and the second turbine is located at $x = 1052$ m. As for the boundary conditions, the inlet is represented with a velocity-inlet conditions with Eq. (1) characterizing the velocity and with Eqs. (2) and (3) represented the turbulence of the flow, assuming the flow as a neutral atmospheric wind flow [8, 8]. The bottom of the domain was characterized with a standard wall condition, and both sides and top of the geometry were represented with a symmetry condition. To the outlet was given a pressure-outlet condition and the rotor of the wind turbine was modelled as an actuator disc with zero thickness and characterized by a porous jump boundary condition [14].

$$u_x = \frac{u_*}{\kappa} \ln\left(\frac{y + y_0}{y_0} \right) \tag{1}$$

$$k = \frac{u_*^2}{\sqrt{C_\mu}} \tag{2}$$

$$\varepsilon = \frac{u_*^3}{\kappa (y + y_0)} \tag{3}$$

where u_x is the velocity in the x-direction, k is the turbulent kinetic energy, ε is the turbulent dissipation rate, u_* is the friction velocity, κ is the von Kármán constant
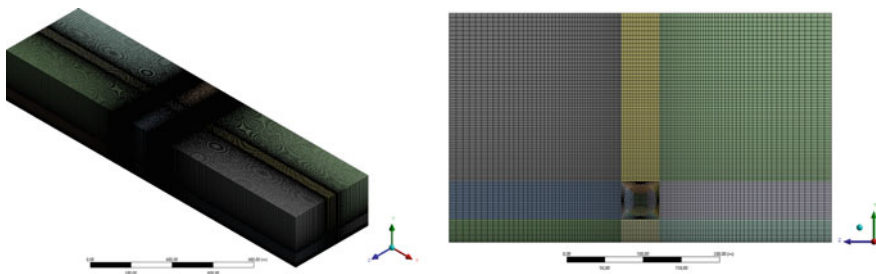
Fig. 1 Mesh of the entire domain and its respective inlet view

($\kappa = 0.4$), y is the height, y_0 is the roughness length, and C_μ is a turbulence model constant.

The CFD analysis was performed with the use of the CFD software ANSYS Fluent 2020 R2. The pressure-coupling was made with the coupled method, and the discretization of the gradients was accomplished with the Green-Gauss node-based approach [15]. The PRESTO! scheme was chosen for the discretization of the pressure parameter, and regarding the momentum, turbulent kinetic energy and turbulent dissipation rate, a Second Order Upwind scheme was applied in order to achieve a high-order accuracy. The monitoring was set at 1×10^{-5} in all residuals.

For the set up data, many parameters were considered. It was assumed only one direction for the wind and a fixed intensity, assuming a velocity of 8.5 m/s at a height of 55 m. The terrain was assumed as a flat grassy plain [16], assuming the absence of a canopy model. The wind turbine selected for the study is a V42-600 kW with a hub height of 55 m from Vestas and was modelled with its standard specifications [17, 18]. It is important to note that the tower structure of the turbine was neglected to simplify the simulation. The validation of the CFD model was performed by comparing the results with the Jensen and Larsen wake models, as described in [19].

3   Results and Discussion

An analysis to the velocity profiles, to the pressure drop at the rotor and to the turbulence in the flow was performed. This allows to study the performance of the turbine due to its positioning since the upstream turbine impacts the efficiency of the downstream turbine.

As observed in Fig. 2, the oncoming wind speed of the upstream turbine assumes a value of 8.5 m/s but, because of the wake effects, the oncoming wind speed of the downstream turbine is reduced to 7.63 m/s. This will automatically influence in the power output of the second wind turbine since the power varies with the cube of the wind velocity. This reduction in wind velocity translates to a power reduction of roughly 28%, which is a similar result as obtained by Tang et al. [9] in an experimental work. Figure 3 represented the velocity contours, where it can be clearly seen the velocity deficit produced by the flow.
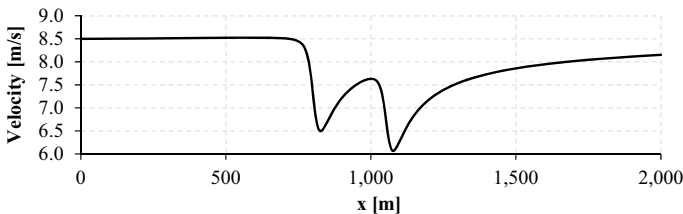
Fig. 2   Velocity along the centre line of both wind turbines
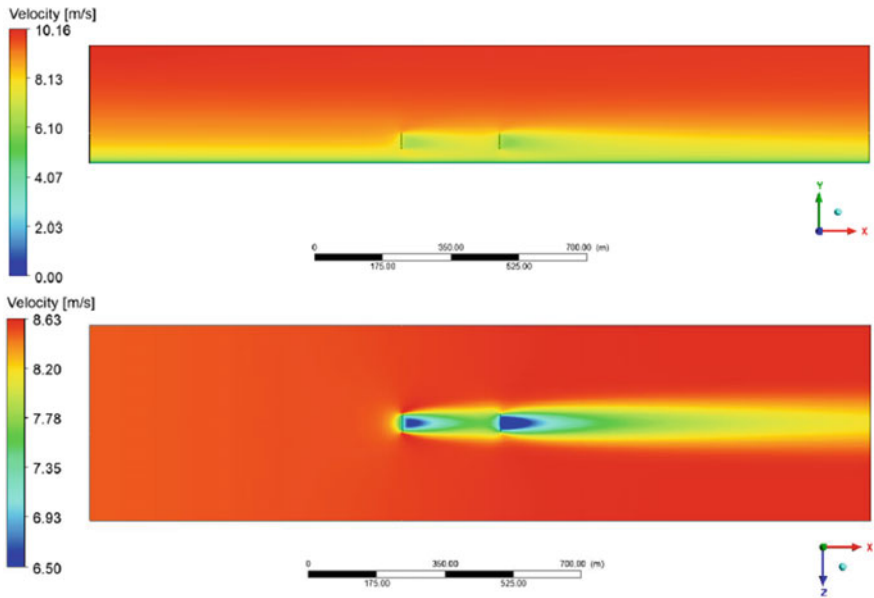
Fig. 3 Velocity contours of the in-line wind turbines simulation along the side plane and top plane, respectively

In terms of pressure drop, as shown in Fig. 4, in the first actuator disc, there is a 22 Pa pressure drop, whereas due to the energy loss because of the impact of the wind with the first turbine, in the second disc, the pressure drop is only of 19 Pa.

Regarding the turbulent kinetic energy, it is possible to observe in Fig. 5 a sudden second increase in turbulence due to the obstacle in the flow induced by the second wind turbine. The impact of the wake effect on the downstream turbine can be easily understood with the use of the turbulent kinetic energy contours demonstrated in Fig. 6.
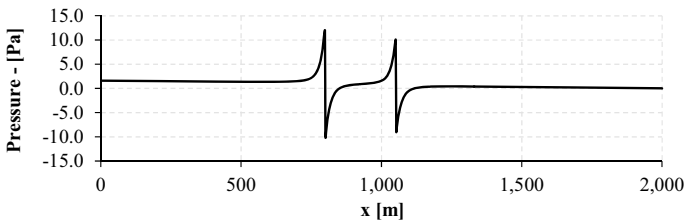
Fig. 4 Pressure along the centre line of both wind turbines
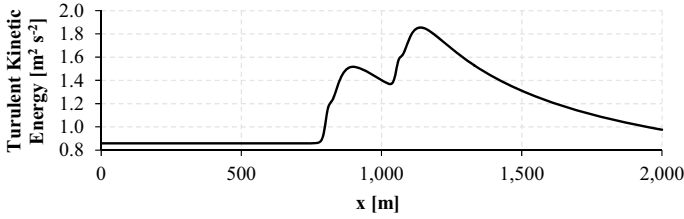
Fig. 5 Turbulent kinetic energy along the centre line of both wind turbines
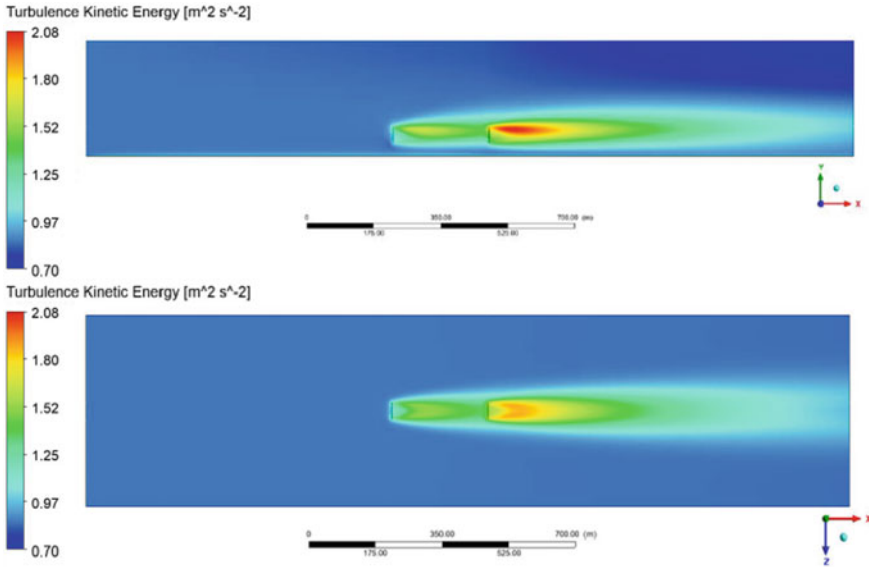
Fig. 6 Turbulent kinetic energy contours of the in-line wind turbines simulation along the side plane and top plane, respectively

4 Conclusions

This CFD analysis allowed to study the impact of upstream rotors on downstream turbine and how their wake would affect their productivity. The simulation showed a decrease in the oncoming wind velocity for the downwind turbine of roughly 10%. The wind velocity (u_0) of the first turbine is 8.5 m/s, whereas the velocity of the wind at the second turbine was reduced to 7.63 m/s. Because of this decrease in velocity, a power loss was also noticed as it went from 220 kW in the first turbine to 160 kW in the second turbine, demonstrating a 28% power loss due to the wake shadowing.

References

1. Letcher TM (2017) Wind energy engineering: a handbook for onshore and offshore wind turbines. Academic Press, London
2. Sawant M, Thakare S, Rao AP, Feijóo-Lorenzo AE, Bokde ND (2021) A review on state-of-the-art reviews in wind-turbine- and wind-farm-related topics. Energies 14(8):2041. https://doi.org/10.3390/en14082041
3. O'Brien JM, Young TM, O'Mahoney DC, Griffin PC (2017) Horizontal axis wind turbine research: a review of commercial CFD, FE codes and experimental practices. Prog Aerosp Sci 92:1–24. https://doi.org/10.1016/j.paerosci.2017.05.001
4. Miller A, Chang B, Issa R, Chen G (2013) Review of computer-aided numerical simulation in wind energy. Renew Sustain Energy Rev 25:122–134. https://doi.org/10.1016/j.rser.2013.03.059
5. Chang C-Y, Schmidt J, Dörenkämper M, Stoevesandt B (2018) A consistent steady state CFD simulation method for stratified atmospheric boundary layer flows. J Wind Eng Ind Aerodyn 172:55–67. https://doi.org/10.1016/j.jweia.2017.10.003
6. Richards P, Hoxey R (1993) Appropriate boundary conditions for computational wind engineering models using the k-ε turbulence model. J Wind Eng Ind Aerodyn 46–47(C):145–153. http://doi.org/10.1016/0167-6105(93)90124-7
7. Richards PJ, Norris SE (2011) Appropriate boundary conditions for computational wind engineering models revisited. J Wind Eng Ind Aerodyn 99(4):257–266. https://doi.org/10.1016/j.jweia.2010.12.008
8. Hargreaves DM, Wright NG (2007) On the use of the k–ε model in commercial CFD software to model the neutral atmospheric boundary layer. J Wind Eng Ind Aerodyn 95(5):355–369. https://doi.org/10.1016/j.jweia.2006.08.002
9. Tang H, Lam K-M, Shum K-M, Li Y (2019) Wake effect of a horizontal axis wind turbine on the performance of a downstream turbine. Energies 12(12):2395. https://doi.org/10.3390/en12122395
10. Uchida T et al (2020) A new wind turbine CFD modeling method based on a porous disk approach for practical wind farm design. Energies 13(12):3197. https://doi.org/10.3390/en13123197
11. Barthelmie RJ et al (2009) Modelling and measuring flow and wind turbine wakes in large wind farms offshore. Wind Energy 12(5):431–444. https://doi.org/10.1002/we.348
12. Porté-Agel F, Bastankhah M, Shamsoddin S (2020) Wind-turbine and wind-farm flows: a review. Boundary-Layer Meteorol 174(1):1–59. https://doi.org/10.1007/s10546-019-00473-0
13. Radünz WC, Mattuella JML, Petry AP (2020) Wind resource mapping and energy estimation in complex terrain: a framework based on field observations and computational fluid dynamics. Renew Energy 152:494–515. https://doi.org/10.1016/j.renene.2020.01.014
14. Avramenko A, Agafonova O, Sorvari J, Haario H, Avramenko Y (2017) Fast numerical modelling method for wind flow investigation based on depth-averaged equations, pp 21–29. http://doi.org/10.2495/CMEM170031
15. Ansys Inc., ANSYS FLUENT theory guide. http://www.afs.enea.it/project/neptunius/docs/fluent/html/th/main_pre.htm
16. Burton T, Jenkins N, Bossanyi E, Sharpe D, Graham M (2021) Wind energy handbook, 3rd edn. Wiley, Hoboken, NJ
17. Vestas Wind Systems A/S. Vestas V42—600,00 kW. https://en.wind-turbine-models.com/turbines/109-vestas-v42#marketplace. Accessed 24 Sept 2021
18. Mollestad KA (1997) Butinge wind power plant—pre-engineering, Steinkjer
19. Silva D, Silva J, Pinto P, Teixeira S, Teixeira J (2022) Analysis and validation of a CFD simulation of the wind through a horizontal axis wind turbine as an actuator disc with a porous jump condition. In: ICIE'2022, International conference innovation in engineering, June 2022, Guimarães, Portugal (accepted for presentation)

Simulation and Yield Analysis of a Solar Tower Plant Combined with Seawater Desalination System in the Mediterranean Area

Hanane Ait Lahoussine Ouali, Mujeeb Iqbal Soomro, Samir Touili, and Ahmed Alami Merrouni

Abstract In this study, we assessed the yield of a solar power tower integrated with a desalination unit, under the climate of Nador city located in the Mediterranean part of Eastern Morocco (35.17° N latitude and −2.93° E longitude). The considered solar tower is equipped with a two-tank direct thermal energy storage system which operates with a mixture of potassium and sodium nitrate molten salt integrated with a power block of 111 MWe turbine gross output performed using System Advisor Model (SAM) software. As for the desalination unit, it consists of a Direct Contact Membrane Distillation System (DCMD) solved in MATLAB environment based on heat and mass transfer equations. The obtained findings show that the annual energy generation and cooling water demand were expected to be 446.7 GWh and 1,141,527 m^3, with a SM = 2.4, a capacity factor of 51.1%, and 12 h of full load thermal storage. This combination leads to a freshwater production ranging from 4046 to 28,833 L/day in December and August, respectively.

Keywords Desalination · Concentrating solar power · Morocco · MATLAB · SAM software

1 Introduction

The world's freshwater resources are decreasing drastically due to population growth, the expansion of industrial activity and climate change. The lack of drinking water concerns mainly some countries with limited reserves or located in arid areas such as the Middle East or North Africa. These countries have resorted to the desalination of seawater to meet their needs. Indeed, 97% of the water available on our planet

H. Ait Lahoussine Ouali (✉) · S. Touili · A. Alami Merrouni
LPTPME Laboratory, Materials Science, New Energies and Application Research Group,
Department of Physics, Faculty of Science, Mohamed First University, Oujda, Morocco
e-mail: hananeaitlahoussine@gmail.com

M. Iqbal Soomro
Department of Mechanical Engineering, Mehran University of Engineering and Technology,
SZAB Campus, Khairpur Mir's, Sindh 66020, Pakistan

is found in the seas and oceans; it must then be treated to be made drinkable. The kingdom of Morocco is one of the countries interested in finding new resources and new ways to provide fresh water. This need for water is mainly due to the rapid growth of the population, the development of different sectors such as agriculture and industry, but also to the drought which has become a factor threatening even the way of life. To fill this deficit, we need to invest in solar desalination technology, as one of the future sources of water, which will contribute to and carry the proposed future solutions for the production of freshwater, given that the country is characterized by a high intensity of solar radiation during most periods of the year [1, 2].

In recent years, several studies in the literature have been interested in the investigation of different desalination systems combined with solar energy. For instance, Manolakos et al. [3] presented a techno-economic comparison between a reverse osmose (RO) desalination unit combined with a photovoltaic system and a RO-Solar Organic Rankine system under the climate of Theresia island in Greece. The findings show that the desalination with the PV-RO system cost reached 7.77 €/m³ whereas that of the system with RO-Solar Rankine was 12.53 €/m³. Similarly, Askari and Ameri [4] carried out a techno-economic analysis of a linear Fresnel reflector solar field combined to a multi-effect desalination using MATLAB and the System Advisor Model software. The results obtained showed that the minimum cost of water was obtained at the lowest solar share (27.54%). Moharram et al. [5] investigated the feasibility of concentrated solar power plant coupling with water desalination systems using a multi-effect desalination plant and a reverse osmose plant under the weather conditions of Ras Ghib in Egypt, using the MATLAB/Simulink to predict power generation and freshwater production. It was found that a solar field of 250,000 m² can produce a total of 22,775 m³/day of desalinated water along with 15.5 MWe supplied to the grid at optimum conditions in July. Ortega-Delgado et al. [6] performed a techno-economic comparative study between MED and RO in South of Spain, in order to find the best coupling strategy for a 5 MWe CSP plant. The results show that the reverse osmose unit connected to the local electric grid is the best coupling option, which produces the lower levelized water cost. Valenzuela et al. [7] presented CSP and photovoltaic plants integrated with multi-effect desalination (MED) system for energy and freshwater production in Chile using TRNSYS software. The findings indicate that a CSP + PV + MED plant presents a capacity factor 7.6% lower than CSP + PV plant. Finally, Mata-Torres et al. [8] carried out a techno-economic investigation of a parabolic trough collector plant integrated with a MED system, in transient conditions, under the weather conditions of Venezuela and Chile. The results obtained showed that the proposed scheme is achievable and can save 15% of installed cost for Chile. The objective of this paper is to assess the performance of a solar tower plant integrated with a DCMD system for both electricity and freshwater production under the weather conditions of Nador city located in the northeast of Morocco. This is the first study of its kind in Morocco combining these two technologies and dealing with the city under investigation.

2 Methodology and Simulation Inputs

2.1 Mathematical Modeling

This section provides a brief mathematical description for the solar power plant and DCMD system obtained from [9–13]:

The amount of thermal energy incident on the heliostat field can be determined from the relationship given as follows:

$$Q_{\text{field}} = \text{DNI} * A_{\text{heliostat}} \tag{1}$$

where DNI is direct normal irradiation, and $A_{\text{heliostat}}$ is the total reflective area of heliostats.

The amount of solar radiation incident on the receiver can be determined from the relationship given as follows:

$$Q_{\text{rec}} = Q_{\text{field}} * \eta_{\text{field}} \tag{2}$$

$$Q_{\text{HTF}} = m'_{\text{HTF}} * (h_{\text{out}} - h_{\text{in}}) \tag{3}$$

where m'_{HTF}: is the mass flow rate of heat transfer fluid (HTF), h_{in}: is the specific enthalpy of HTF at receiver inlet, and h_{out} is specific enthalpy of HTF at receiver outlet.

The solar multiple (SM) is defined as the ratio of the actual size of the solar field to the size of the solar field needed to power the turbine at nominal conditions and expressed by the following formula:

$$\text{SM} = \frac{Q_{\text{field}}}{Q_{\text{PB,ref}}} \tag{4}$$

where $Q_{\text{PB,ref}}$ is the thermal power required by the power block.

The temperature polarization coefficient (TPC), which is the ratio of the actual driving force to the theoretical driving force, can be obtained from the following equation:

$$\text{TPC} = \frac{T_{\text{mf}} - T_{\text{mp}}}{T_{\text{bf}} - T_{\text{bp}}} \tag{5}$$

where T_{mf} is the membrane surface temperature for feed side, T_{mp} membrane surface temperature for permeate side, and T_{bp} and T_{bf} are the bulk permeate and side temperatures, respectively.

T_{mf} and T_{mp} can be estimated the following equations:

$$T_{mf} = \frac{k_m * \left(T_{bp} + \left(\frac{h_f}{h_p}\right) * T_{bf}\right) + \left(\left(\delta_m * (h_f * T_{bf} - J_m * \Delta H_v)\right)\right)}{(k_m) + \left(h_f * \left(\delta_m + \left(\frac{k_m}{h_p}\right)\right)\right)} \tag{6}$$

$$T_{mp} = \frac{k_m * \left(T_{bf} + \left(h_p / h_f\right) * T_{bp}\right) + \left(\left(\delta_m * (h_p * T_{bp} + J_m * \Delta H_v)\right)\right)}{(k_m) + \left(h_p * (\delta_m + (k_m / h_f))\right)} \tag{7}$$

where k_m is the thermal conductivity of membrane, h_p and h_f are convective heat transfer coefficient for permeate and feed side respectively, δ_m is the membrane thickness and H_v is the latent heat of vaporization of water.

Besides, the permeate flux (J_m) through the membrane can be obtained from the following equation:

$$J_m = D_m * \left(P_a^v - P_b^v\right) \tag{8}$$

where D_m is the coefficient for membrane distillation, P_a^v is the vapor pressure of the feed water, and P_b^v is the vapor pressure of permeate.

Finally, the quantity of freshwater produced within the DCMD unit which is the ultimate objective of this study can be expressed by the following formula:

$$Q_w = (J_m * A_{DCMD} / \rho) \tag{9}$$

where A_{DCMD}: is the area of membrane and ρ is the water density.

2.2 Solar Plant and Desalination System Design Parameters

As mentioned above, the CSP part consists of a solar tower with a two-tank direct thermal energy storage system which operates with a mixture of $NaNO_3$ and KNO_3 of 60% and 40%, respectively. The thermal field is integrated with a power block of 111 MWe gross consisted of a conventional Rankine operating using wet cooling with seawater as cooling agent using SAM software [14]. The desalination part consists of Direct Contact Membrane Distillation System achieved in MATLAB environment based on heat and mass transfer equations. The specific parameters chosen for the simulation are presented in Table 1 [10, 15, 16]. Besides, the optimized configuration of the heliostats field found in the Nador city is presented in Fig. 1.

3 Resource Assessment of the Investigated Location

The study of CSP plant integrated with desalination unit performances was examined using the meteorological data of the Nador location (35.17° N latitude and −2.93° E

Table 1 Input data for simulation of the solar power plant and DCMD desalination system

Parameter	Value	Unit
	Solar tower data	
Design turbine gross output	111	MWe
Estimated net output (nameplate)	100	MWe
HTF type	Molten salt (60% $NaNO_3$ + 40% KNO_3)	–
HTF hot temperature	574	°C
HTF cold temperature	290	°C
Solar multiple	2.4	
Tower height	192.8	m
Receiver diameter	17.32	m
Number of panels	20	–
Tube wall thickness	1.25	mm
Tube outer diameter	40	mm
Heliostat width	12.2	m
Heliostat height	12.2	m
Single heliostat area	144.37	m^2
Thermal storage fluid	Molten salt (60% $NaNO_3$ − 40% KNO_3)	–
Full load hours of storage	12	h
Storage tanks	2	–
	DCMD unit data	
Membrane type	Flat-sheet	–
Effective area	0.14	m^2
ε	91	%
δ_m	120	μm
τ	1.098	–
C_f	35	g/L
T_{bf}	20	°C

longitude) extracted from [17]. Nador is a city in the northeast of Morocco, located in the Eastern Rif region, Nador enjoys a Mediterranean climate with an acceptable amount of solar irradiance. Figure 2 illustrates the monthly averages of direct normal irradiation (DNI). As indicated in Fig. 2, the DNI, which is the most important parameter to be assessed for CSP plants, has its peak around noon. The highest DNI records are of 733 W/m^2 and they were measured in June. While the lowest values
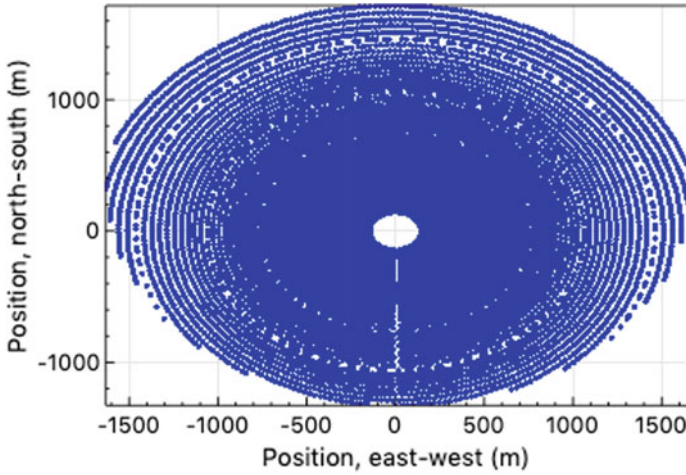
Fig. 1 Layout of the heliostat field for the study area

occur in December. Moreover, as we mentioned in previous section, typically used cooling water was switched by seawater. For that, the monthly average values of the sea water temperature are the ones used to cool the condenser in the solar tower power plant, and their values are presented in Fig. 3. As it can be observed, the sea water temperature values [18] varied between 13 and 26 °C which is very convenient for the tower plant cooling requirements.
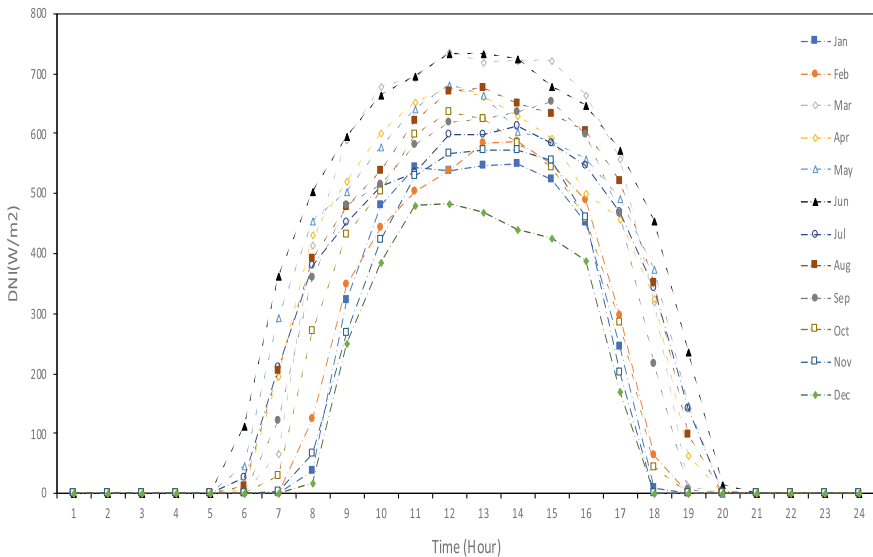
Fig. 2 Direct normal radiation at Nador, Morocco
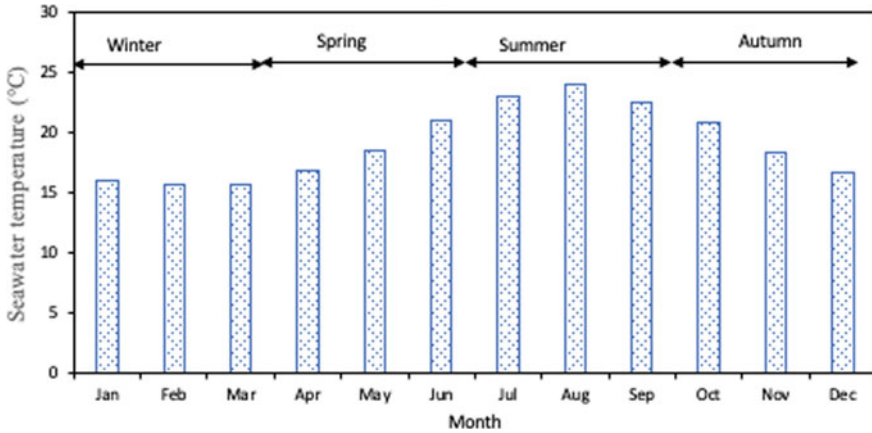
Fig. 3 Monthly change in sea water temperature in Nador city

4 Results and Discussion

4.1 Solar Tower Assessment

Figure 4 presents the daily averages of the hourly values simulated for each day of the month, every month of plant performance considering the field thermal power, the receiver incident thermal power produced and TES charge and discharge thermal energy. As seen from figure, the field and the receiver incident thermal power values varied throughout the year, and an increase/decrease in the field incident thermal power increased/decreased the supply of the receiver incident thermal power. Concerning the charge and discharge TES system curves, it's obvious that the TES system stores a good amount of thermal energy; it exceeded the 200 MW$_{th}$ from March to June. This amount of stored energy gives the plant the capacity to run the power cycle after sunset and in the cloudy time. Figure 5 shows the monthly electricity production from solar plant integrated with thermal storage studied in Nador, Morocco. As indicated in the figure, the month of June produced the most energy (55.69 GWh) which represents 12% of the total energy produced, while December generated the least energy (17.67 GWh) which represent only 4% of the total energy produced. It is worth mentioning that for the 100 MWe solar tower plant with thermal storage system (12 h), the annual capacity factor was found to be 51.1% which is considered good for a case of Nador. Concerning the cooling water demand as shown in Fig. 6 the monthly cooling water requirement has the same trend as electrical energy production. Hence, highest and lowest cooling water requirements for the investigated solar plant were found to be 142,424 m^3 in June, while 45,158 m^3 was found in December.
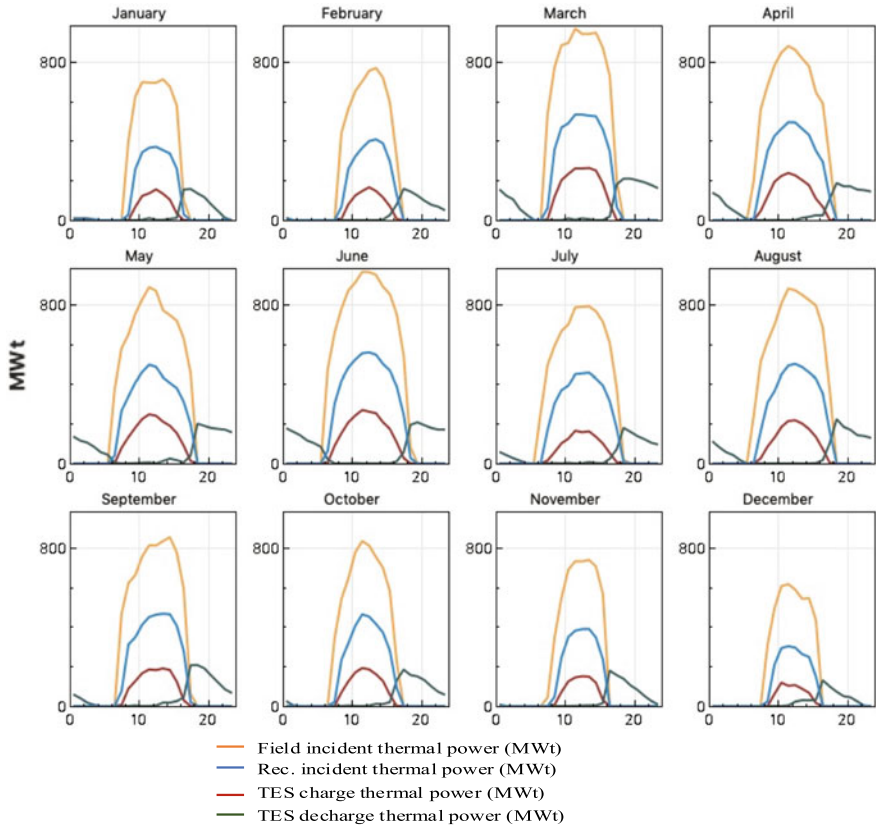
Fig. 4 Field thermal power, the receiver incident thermal power produced, and TES charge and discharge thermal energy

4.2 Desalination Assessment

In this section, permeate flux and freshwater production will be presented. The effect of feed temperature on permeate flux is presented in Fig. 7. It can be seen in Fig. 7 that permeate flux was 4.178 kg/m^2-h at 25 °C; however, it increased exponentially to 17.8 kg/m^2-h when feed temperature was increased to 35 °C. This is in relation with Antione equation. Figure 8 displays the production of freshwater from the DCMD unit for a year (from 0 to 8760 h). As it can be seen, the freshwater production unit reaches its maximum during the summer months. And the highest average value was expected to be 28,833 L/day in August, while the minimum production occurred during the winter season and the lowest freshwater production was 4046 L/day in December.
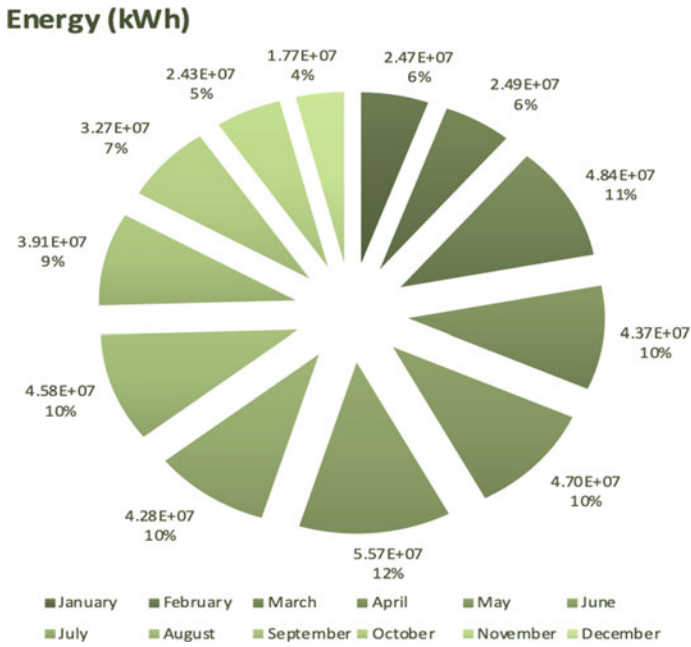
Fig. 5 Monthly energy generation for the study area
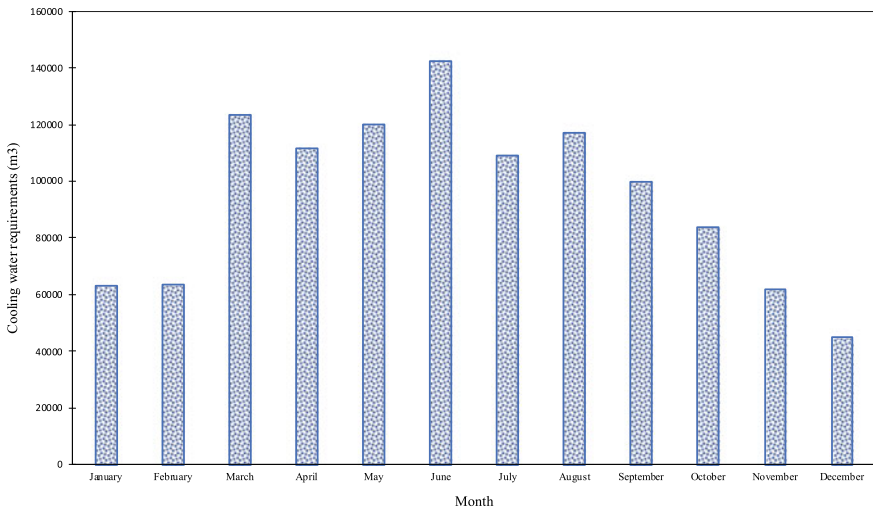
Fig. 6 Cooling water requirements in the studied area
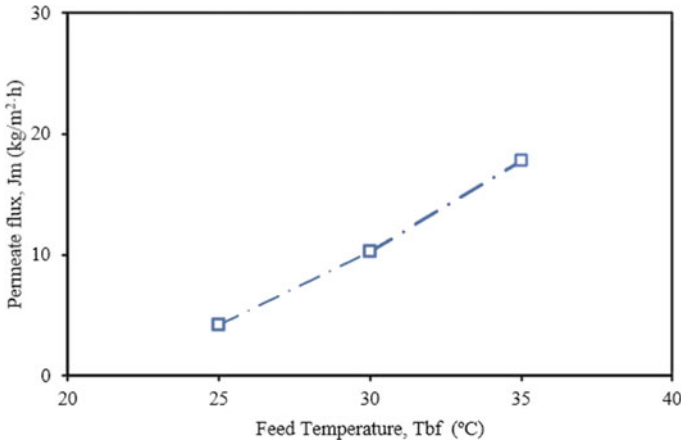
Fig. 7 Effect of inlet temperatures on the permeate flux
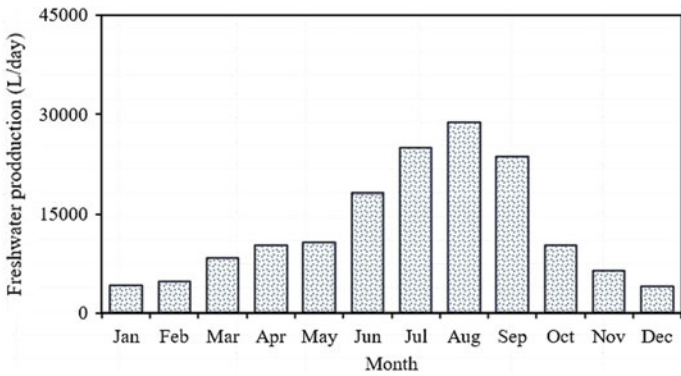
Fig. 8 Monthly freshwater production from the investigated system

5   Conclusion

The paper presented the feasibility study of a solar tower plant coupled with a seawater desalination DCMD system under the weather conditions of Nador in Morocco. The simulations were conducted using the SAM and MATLAB software to predict power generation and fresh water production, respectively. The main concluded remarks could be summarized as follows:

- With a SM = 2.4 and 12 h full load thermal storage, the maximum and minimum electrical energy production was predicted to be to be 55.69 GWh and 17.67 GWh in June and December, respectively.
- The cooling water demand varied from 45,158 to 142,424 m^3.

- The freshwater production fluctuated from 4046 to 28,833 L/day in December and August, respectively.

The results of this study can be used to pave the way for a sustainable realization of solar desalination plants in the coastal region of Morocco.

References

1. Merrouni AA, Ghennioui A, Wolfertstetter F, Mezrhab A (2017) The uncertainty of the HelioClim-3 DNI data under Moroccan climate. AIP Conf Proc 27:1850
2. Azouzoute A, Alami Merrouni A, Touili S (2020) Overview of the integration of CSP as an alternative energy source in the MENA region. Energy Strateg Rev 1(29):100493
3. Manolakos D, Mohamed ES, Karagiannis I, Papadakis G (2008) Technical and economic comparison between PV-RO system and RO-Solar Rankine system. Case study: Thirasia island. Desalination 221(1–3):37–46
4. Askari IB, Ameri M (2016) Techno economic feasibility analysis of linear Fresnel solar field as thermal source of the MED/TVC desalination system. Desalination 15(394):1–17
5. Moharram NA, Bayoumi S, Hanafy AA, El-Maghlany WM (2021) Techno-economic analysis of a combined concentrated solar power and water desalination plant. Energy Convers Manag 15(228):113629
6. Ortega-Delgado B, García-Rodríguez L, Alarcón-Padilla DC (2016) Thermoeconomic comparison of integrating seawater desalination processes in a concentrating solar power plant of 5 MWe. Desalination 15(392):102–117
7. Valenzuela C, Mata-Torres C, Cardemil JM, Escobar RA (2017) CSP + PV hybrid solar plants for power and water cogeneration in northern Chile. Sol Energy 15(157):713–726
8. Mata-Torres C, Escobar RA, Cardemil JM, Simsek Y, Matute JA (2017) Solar polygeneration for electricity production and desalination: case studies in Venezuela and northern Chile. Renew Energy 1(101):387–398
9. Solar field optimization and its impact on overall design and performance of solar tower thermal power plant in Bangladesh | Enhanced Reader
10. Soomro MI, Kim WS (2018) Performance and economic investigations of solar power tower plant integrated with direct contact membrane distillation system. Energy Convers Manag 15(174):626–638
11. Montes MJ, Abánades A, Martínez-Val JM, Valdés M (2009) Solar multiple optimization for a solar-only thermal power plant, using oil as heat transfer fluid in the parabolic trough collectors. Sol Energy 83(12):2165–2176
12. Qtaishat M, Matsuura T, Kruczek B, Khayet M (2008) Heat and mass transfer analysis in direct contact membrane distillation. Desalination 219(1–3):272–292
13. Khayet M (2011) Membranes and theoretical modeling of membrane distillation: a review. Adv Colloid Interface Sci 164(1–2):56–88
14. Home—System Advisor Model (SAM) [Internet] [cited 2022 Jan 5]. Available from: https://sam.nrel.gov/
15. Ouali HAL, Moussaoui MA, Mezrhab A, Naji H (2020) Comparative study between direct steam generation and molten salt solar tower plants in the climatic conditions of the eastern Moroccan region. Int J Renew Energy Dev 9(2):287–294
16. Ait Lahoussine Ouali H, Raillani B, Amraqui S, Moussaoui MA, Mezrhab A, Mezrhab A (2020) Analysis and optimization of SM and TES hours of central receiver concentrated solar thermal with two-tank molten salt thermal storage. Lecture notes in electrical engineering 684, pp 666–673

17. Geographical Information System [Internet] [cited 2021 Dec 14]. Available from: https://re.jrc.ec.europa.eu/pvg_tools/en/#TMY

18. World Water Temperature [Internet] [cited 2022 Jan 10]. Available from: https://www.seatemperature.org/

Evaluation of Photovoltaic, Wind and Hybrid Energy Systems for the Power-to-Hydrogen (PtH) Concept in Eastern Morocco

Samir Touili, Salaheddine Amrani, Hanane Ait Lahoussine Ouali, Ahmed Alami Merouni, and Hassane Dekhissi

Abstract The objective of this study is to conduct a feasibility study on the power-to-hydrogen (PtH) projects in the eastern region of Morocco. For this reason, we simulated the hydrogen production from a photovoltaic, wind and hybrid plants with a capacity of 50 MW_p under the climatic conditions of Oujda city taken as a case of study. We calculated the cost of hydrogen production for each plant, and we analyzed the impact of the system components. The results show that the highest amounts are generated by the wind energy with an annual production of 1806 tons, and the lowest cost of production is achieved by the photovoltaic plant with a cost of production of 4.05 \$/kg. As for the hybrid plant, it can produce hydrogen higher than the PV plant with a cost lower than the wind energy plant, and therefore, it can be an attractive option for PtH projects if an optimal configuration is achieved between PV and wind; especially the results showed that the costs related to the electricity production account for 94% of the hydrogen production cost.

Keywords Power-to-hydrogen · Eastern Morocco · Economic analysis

1 Introduction

Nowadays, the energy consumption worldwide is dominated by fossil fuels with almost 80%. Such fuels are the main cause to the global warming because they release greenhouse gases during combustion. Furthermore, fossil fuels are depleting rapidly due to the excessive use, thus the need of more sustainable sources of energy [1]. For this reason, many governments around the globe are adding more renewable energy sources into their energy systems, and consequently, the renewable energy

S. Touili (✉) · S. Amrani · H. Ait Lahoussine Ouali · A. Alami Merouni
LPTPME Laboratory, Materials Science, New Energies and Application Research Group, Department of Physics, Faculty of Science, Mohammed First University, Oujda, Morocco
e-mail: touili.univ@gmail.com

H. Dekhissi
LPTPME Laboratory, Physique Nucléaire, Physique des Particules et Applications, Research Group, Faculty of Sciences, Mohammed First University, Oujda, Morocco

capacity installed around the world is knowing a significant and continuous increase especially for wind and solar energy [2].

However, the main drawback with renewable energy like wind and solar is their fluctuant nature, which can cause several technical issues with respect to the reliability to satisfy the energy demand. In addition, it's difficult for other sectors like in the industry and transport to become sustainable on only renewable electricity. Therefore, it is necessary to convert the electricity production from the intermittent sources into other form that are reliable and can be store more efficiently [3]. To overcome this challenge, a new concept has recently attracted a lot of attention called power-to-hydrogen (PtH), in which the generated electricity is converted to hydrogen by the mean of the water electrolysis process [2]. Indeed, hydrogen offers several advantages like its high energy density compared to other fossil fuels, possibility to be stored efficiently for long periods without losses; it releases only water in the course of the combustion process, and it can be used as an energy vector in all the application requiring fossil fuels in addition to a wide range of industrial application [4].

Therefore, the aim of this study is to evaluate hydrogen production from photo-voltaic (PV), wind and hybrid plant in order to assess the potentiality of PtH projects in the eastern region of Morocco taken the city of Oujda as a case study. The methods used to achieve this goal as well as the pertinent results are presented in the next sections.

2   Methodology

In order to evaluate hydrogen production from the PV, wind and hybrid plant, we firstly simulate the electricity production from each plant, then we calculate the hydrogen production, and finally we quantify the cost of production.

2.1   Electricity Production

The simulation of the electricity from the different plants is conducted by the Green Energy System Analysis Tool (Greenius) developed by the German Aerospace Center (DLR). Greenius is a widely used tool for the simulation of various renewable energy systems, and it has been used in many studies available in the literature, for example, we cite: [5–10]. The designed plant's capacity has been fixed at 50 MW_p for the PV and wind plants, while for the hybrid the overall 50 MW_p capacity is equally divided between PV and wind with 25 MW_p for each technology. As for the simulation, we selected the Yingli Panda YL250P-29 b PV module which has an efficiency of 15.4% for the PV plant, while for the wind park we selected the Gamesa G114 turbine which has a nominal power of 2.5 MW and 80 m hub height.

2.2 Hydrogen Production

The electricity production from the plants mentioned above is used to drive the water electrolysis process. To this end, various electrolyzers exist with different stages of technological development. In this study, we used the proton exchange membrane (PEM) technology because it's one of the most used due to its maturity, high efficiency, high lifetime and most importantly its compatibility with fluctuant source of electricity like in the case of renewable energy source. The hydrogen production is estimated as follows [11–16]:

$$M_{H_2} = \frac{E \times \eta_{ele}}{HHV_{H_2}} \tag{1}$$

M_{H_2} represents the mass of hydrogen production in kg, E is the electricity production in kWh, η_{ele} is efficiency of the electrolyzer taken as 75%, and HHV_{H_2} is the hydrogen higher heating value: 39.4 kWh/kg.

2.3 Levelized Cost of Production

For the evaluation of the economic viability of renewable energy projects, many indictors exist; one of the most used is the levelized cost of production. Indeed, this indicator takes in consideration all the project costs during its lifetime like the capital, installation, operation and maintenance costs as well as the technical aspect of the project. The levelized cost of hydrogen production (LCO_{H_2}) in $/kg is calculated as follows [13, 16–19]:

$$LCO_{H_2} = \frac{\sum_{i=0}^{N} \left((C_E + C_{ele})(1 + T)^{-i} \right)}{\sum_{i=0}^{N} \left(M_{H_2,i}(1 + T)^{-i} \right)} \tag{2}$$

T is the discount rate, N is the project lifetime (20 years), $M_{H_2,i}$ is the annual hydrogen production in the year i, and C_E and C_{ele} are the investment costs of the electricity production and the electrolyzer plant, respectively.

3 Results and Discussion

The daily hydrogen production from the PV, wind and hybrid plants is presented in Fig. 1. The results show that the daily hydrogen production from the PV plant is in the range of 0.6–6 tons, and between 0.13 and 18.67 tons for the wind energy plant, while for the hybrid plant it ranges between 0.51 and 10.62 tons. It can also be noticed that the high amounts of hydrogen are generated during spring and summer in the
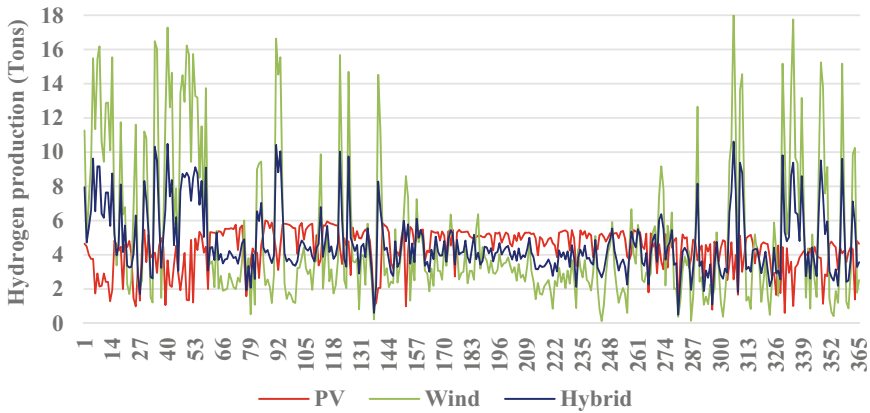
Fig. 1 Daily hydrogen production

case of the PV plant which is due to the fact that in this period high amounts of solar irradiation are received and therefore, electricity and consequently hydrogen can be produced more efficiently. As for the wind energy plant, the profile of hydrogen production is highly intermittent which is due to fluctuant nature of wind resources that varies greatly in time, with the majority of days with large production observed during fall and winter. Also, it can observe that hydrogen production from the PV plant is higher than the one from the wind energy in the summer and spring period while its lower in the fall and winter period. It's worth to mention that the daily average hydrogen production is 4.35, 4.94 and 4.64 tons for the PV, wind and hybrid plant, respectively.

The monthly hydrogen generation from PV and wind is presented in Fig. 2. It can be noticed that the hydrogen production from wind energy exhibits high fluctuation through the months of the year with no apparent profile, which were expected after the previous results. Indeed, the monthly production ranges between 76.76 and 292.17 tons recorded in August and February, respectively. In contrast, hydrogen production from the PV plant has a less important difference between the month. In addition, a production profile can be clearly observed as the production is at its lowest values during winter months and then, its gradually increases until the production reached the maximum in the spring and summer months, then it decreases during the fall months. Also, the difference of hydrogen production is less significant as it ranges between 97 and 157.34 tons with the highest amount recorded in April while the lowest is observed in February. Moreover, the hydrogen production from wind energy is remarkably higher than the PV plants during winter months especially in February where the production from wind energy is almost 3 the production from the PV plant, while the production from PV exceeds the one from wind energy during summer as its almost the double in July and August. As for the monthly average hydrogen production, its 132.46 tons for the PV plant and 150.49 tons for the wind energy plant.
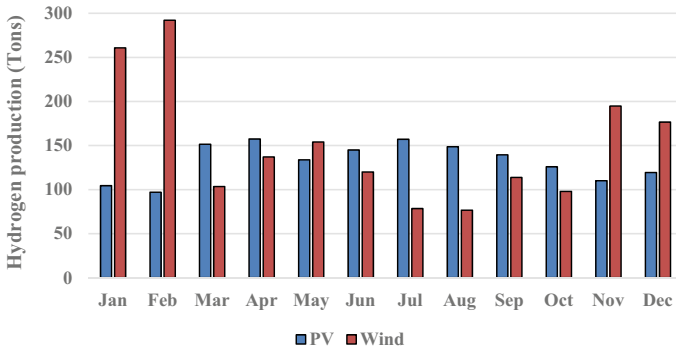
Fig. 2 Monthly hydrogen production from the wind and PV plant

With respect to the monthly production from the hybrid plant, it is displayed in Fig. 3, and as it can be seen, the production pattern is less intermittent in comparison to the wind energy; besides, high amounts are generated through the year, with the PV represents the highest share of hydrogen production during spring/summer months, whereas wind energy has the higher share during winter/fall months. In fact, the highest share of hydrogen production from wind energy occurs in February with 75% of the production, while the highest share from PV is in July with 66.6% of the production. The monthly hydrogen production from the hybrid plant varies between 111.94 and 194.6 tons recorded in October and February respectively with a monthly average of 141.48 tons.

Regarding the annual hydrogen production, the highest amounts are generated by the wind energy plant with 1806 tons followed by the hybrid plant with 1697.79 tons and finally the PV plant with 1589.56 tons. In addition, the results show that the
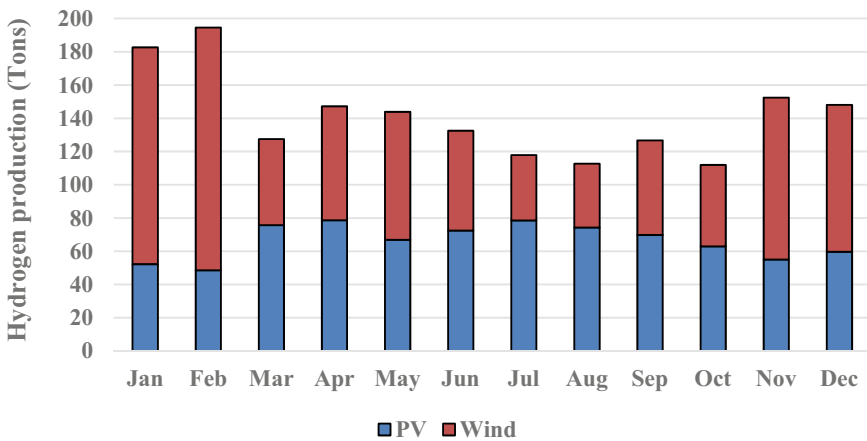
Fig. 3 Monthly hydrogen production from the hybrid plant

wind energy is responsible of 53.18% of the hydrogen production, while the PV plant generates the remaining 46.81%, which were expected since the production from the wind energy plant is higher than the one from the PV plant.

In order to have an idea on the hybrid plat hydrogen production pattern, we calculated the mean hourly hydrogen production in every month, the results are presented in Fig. 4. The first observation is that in each month, the profile of hydrogen production is in general similar to the one from the PV technology since the production start to increase early in the in morning around 6:00, and it continues to rise till noon around 12:00–15:00 depending on the month where it reaches its maximum, then it starts to decrease gradually to reach the lowest values during the night. Nonetheless, it can be noticed that hydrogen is produced during nighttime as well, due to the fact that wind turbines can operate all day long since wind resources are available also after sunset. In fact, the lowest hydrogen production is between 3 and 176 kg while the highest ranges between 400 and 600 kg.

The cost of hydrogen production is presented for each plant in Fig. 5 (left). The lowest cost is achieved by the PV plant with an $LCOH_2$ of 4.05 \$/kg, followed by the hybrid plant with 7.94 \$/kg, while the production cost from the wind energy plant is the highest with 11.37 \$/kg. It can be noticed that the $LCOH_2$ of the wind energy is almost 3 times the PV $LCOH_2$ even though its hydrogen production is higher than PV. This is due to the high costs related to the wind energy technology in one hand and the limited wind resource in the selected site in the other hand. In fact, the cost of production of any renewable technology depends mainly on two parameters, the technology investment cost and its capacity to generate electricity and consequently hydrogen.

Accordingly, in order for wind energy-based project to be economically viable they must be installed in locations with high wind resources to be able to compensate for its high costs with large amount of electricity and consequently hydrogen production, especially that wind turbines have a large operating period, in fact the hydrogen production by the wind park in the current location exceed the production from the
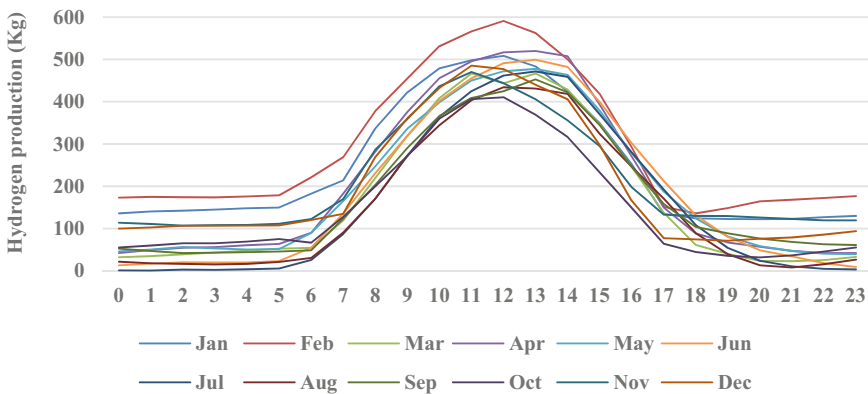
Fig. 4 Mean monthly hourly hydrogen production
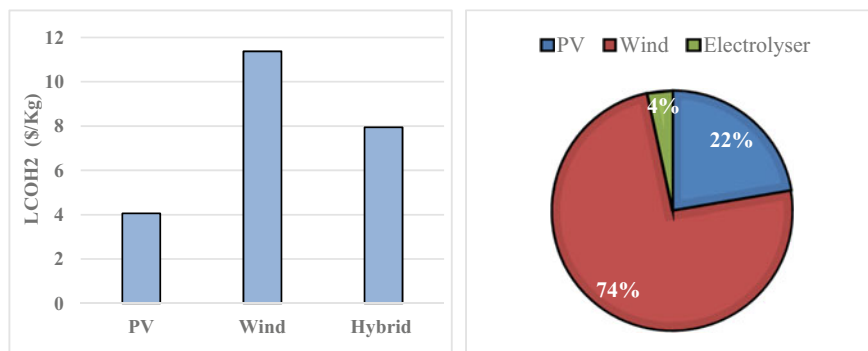
Fig. 5 LCOH$_2$ from the different plants (left); and the hydrogen production cost fraction for the hybrid plant (right)

PV plant even if its wind resources are considered as limited (the mean annual wind speed in reference height is 2.43 m/s). In contrast, the investment cost of the PV technology is much lower which explain its low LCOH$_2$ even if its operation period is limited during the day.

For a better understanding of the impact of the investment cost of each technology on the overall hydrogen production cost, we plotted the share of each one in the case of the hybrid plant in Fig. 5 (right). As expected, the costs related to wind energy represent the majority of the total cost with 74%, then the PV costs with 22% and finally the electrolyzer costs with only 4%. In addition, it can be perceived that the cost of hydrogen production is largely dominated by the costs related to the electricity production to drive the water electrolysis process.

4 Conclusion

In this paper, we present a technical an economical evaluation of hydrogen production from PV, wind and hybrid plant in the city of Oujda located in eastern Morocco. The results show that the highest production is achieved by the wind energy plant with an annual production of 1806 tons followed by the hybrid plant with 1697.79 tons and finally the PV plant with 1589.56 tons. However, the lowest cost of production is achieved by the PV plant with 4.05 $/kg, followed by the hybrid plant with 7.94 $/kg, and finally the wind energy plant with 11.37 $/kg. Furthermore, it was found that the cost of hydrogen production is dominated by the costs related to the electricity production, and therefore, we recommend that an optimal configuration must be selected between PV and wind to take advantage of the technical and economic benefits of both technologies in order to have a larger operation period and a more efficient hydrogen production to the aim of increasing the economic viability of power-to-hydrogen projects.

References

1. Bian Z (2020) A review on perovskite catalysts for reforming of methane to hydrogen production. Renew Sustain Energy Rev 134:10291
2. Allouhi A (2020) Management of photovoltaic excess electricity generation via the power to hydrogen concept: a year-round dynamic assessment using artificial neural networks. Int J Hydrogen Energy 45:21024–21039
3. Touili S (2018) A technical and economical assessment of hydrogen production potential from solar energy in Morocco. Int J Hydrogen Energy 43:22777–22796
4. Touili S (2020) Analysis of the yield and production cost of large-scale electrolytic hydrogen from different solar technologies and under several Moroccan climate zones. Int J Hydrogen Energy 45(51):26785–26799
5. Touili S (2019) Performance analysis of large scale grid connected PV plants in the MENA region. Int J Eng Res Africa 42:139–148
6. Touili S (2020) Design and simulation of PV/electrolyser system for smart, green and efficient residential building in Morocco. In: 2019 international conference on intelligent systems and advanced computing sciences, pp 1–6
7. Touili S (2021) Technical and economic analysis of solar hydrogen production in Morocco. In: Proceedings of the 2nd international conference on electronic engineering and renewable energy systems. ICEERE 2020. Lecture notes in electrical engineering 681, pp 777–783
8. Touili S (2020) Techno-economic investigation of electricity and hydrogen production from wind energy in Casablanca. IOP Conf Ser Mater Sci Eng 948:012012
9. Touili S (2018) A comparative study on hydrogen production from small-scale PV and CSP systems. In: Proceedings of the 1st international conference on electronic engineering and renewable energy. ICEERE 2018. Lecture notes in electrical engineering 519
10. Touili S (2019) A techno-economic comparison of solar hydrogen production between Morocco and Southern Europe. In: 2019 international conference on wireless technologies, embedded and intelligent systems
11. Dagdougui H (2011) A regional decision support system for onsite renewable hydrogen production from solar and wind energy sources. Int J Hydrogen Energy 36:14324–14334
12. Rahmouni S (2017) Prospects of hydrogen production potential from renewable resources in Algeria. Int J Hydrogen Energy 42:1383–1395
13. Gouareh A (2015) GIS-based analysis of hydrogen production from geothermal electricity using CO_2 as working fluid in Algeria. Int J Hydrogen Energy 40:15244–15253
14. Honnery D (2009) Estimating global hydrogen production from wind. Int J Hydrogen Energy 34:727–736
15. Posso F (2016) Preliminary estimation of electrolytic hydrogen production potential from renewable energies in Ecuador. Int J Hydrogen Energy 41:2326–2344
16. Rahmouni S (2016) GIS-based method for future prospect of hydrogen demand in the Algerian road transport sector. Int J Hydrogen Energy 41:2128–2143
17. Gökçek M (2010) Hydrogen generation from small-scale wind-powered electrolysis system in different power matching modes. Int J Hydrogen Energy 35:10050–10059
18. Rahmouni S (2014) A technical, economic and environmental analysis of combining geothermal energy with carbon sequestration for hydrogen production. Energy Procedia 50:263–269
19. Sigurvinsson J (2007) Can high temperature steam electrolysis function with geothermal heat? Int J Hydrogen Energy 32:1174–1182

Improvement of the Mirror Cleanliness Control Methodology in a Concentrating Solar Power (CSP) Power Plant

Fatima-Ezahra El Haddad, Yousra Jbari, and Souad Abderafi

Abstract The maintenance of the solar field of a CSP power plant has a major importance in order to ensure good electricity production. This maintenance is ensured by cleaning the collector mirrors. However, before proceeding with the cleanup, prior planning of the priority areas to be cleaned is necessary. The objective of this work is to analyze and improve the action plan of controlling the solar field of the NOOR I power station located in Ouarzazate City of Morocco. The monitoring method currently used via seven templates only provides information on the state of the solar field in two days. This has a negative impact on the precision of the cleaning planning. However, we proposed the reduction of the number of samples of each template, in order to be able to carry out a daily check of the integrity of the solar field. Using statistical software and industrial data related to the measured reflectivity, the reduction of samples from each template was tested, with two different percentages 30 and 50%. A comparison of the different results obtained, allowed us to retain the most optimal reduction, which means the one with 30%. It allows for keeping good visual distribution in the map of the solar field used, and the cleanliness factor is obtained with relative average error less than 0.10%.

1 Introduction

In solar thermal power plant, operating with cylindrical parabolic collectors, the solar field is considered to be the most important part which contains mirrors intended to intercept the solar radiation and to concentrate it in a focal point [1]. This solar

F.-E. El Haddad · Y. Jbari (✉) · S. Abderafi
Mohammadia Engineering School, Mohammed V University in Rabat, Ibn Sina, B.P. 765, Agdal, 10090 Rabat, Morocco
e-mail: yousrajbari@research.emi.ac.ma

S. Abderafi
e-mail: Abderafi@emi.ac.ma

field requires the most maintenance and control, especially since industrial units are generally located in desert areas [2, 3]. The efficiency of the solar field is influenced by the dirt and dust that are collected on top of the mirrors of the CSP plant [4]. This results in a decrease in their reflectance and energy production. To ensure good efficiency of the solar installation, an appropriate cleaning method must be frequently applied, to remove dirt and dust [5]. It is an expensive procedure, especially since it is based on demineralized water. However, daily control of the various plant equipments is a plied, mainly to the solar field, before cleaning it. Unfortunately, the information can only be available after two days, due to the large capacity of the factory. However, this can result in a minimization of electricity production due to a dirty solar field which must be urgently cleaned. In this study, we will seek to minimize cleaning control time, by focusing on the NOOR I CSP plant, located in Morocco [6].

2 Solar Field and Control Methodology

NOOR I is an industrial unit that produces electricity using CSP technology. Its collectors are mobile with linear focus, made up of parabolic mirrors focusing the solar irradiation on the tubular receiver. It is located in a desert area in the south of Morocco, considered to be an appropriate site benefiting from a high DNI [7]. However, its solar field can be considerably affected by dust, followed by a red rain event, as has been shown in [8, 9].

2.1 Solar Field Description

The solar field of NOOR I is divided into eight areas as can be seen in Fig. 1a. Each area is represented by a color. The solar field contains seven platforms, and each one of it has two subfields: for a total of 14 subfields. Each subfield has a loop assembly which consists of two rows called Filas, and each one of it has two collectors. Between two collectors there is a drive pylon that provides power and allows remote control of rows from the control room. To check the cleanliness of these mirrors, it is difficult to carry out a direct test on the entire solar field, because it is very expensive in time and money. The mirrors are cleaned by trucks equipped with brushes and high-pressure cleaners. Brushes are used for very dirty mirrors, while high-pressure cleaners are used for less dirty mirrors [6]. Cleaning trucks use an average of 1.5 m^3 of demineralized water to clean a single loop [10]. About 1 h is required to clean the mirrors of a loop. So, the samples corresponding to the sections distributed according to the area, most sensitive to soiling are chosen. Knowing that fouling is greater in collectors near the edges of the field, on main roads and in electrical islands, another subdivision has been implemented (Fig. 1b). The parts with dark colors are the area most likely to be dirty, called "High fouling." The parts with lighter colors are the areas least exposed to dust called "Low fouling."
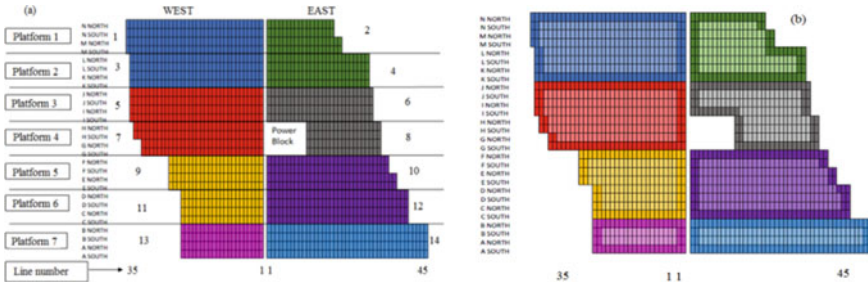
Fig. 1 Diagram representing the solar field (**a**) and its statistical distribution (**b**)

2.2　Control Procedure

The control procedure is based on the calculation of the cleanliness factor (CF), which is a parameter for evaluating the effect of soiling on the mirrors, to deduce their performance [9]. It is calculated based on the measure of reflectivity that is obtained using a portable specular reflectometer to move it between the different areas of the solar field and position it on the surface of the reflector [11]. For reasons of ease of measurement and accessibility, the collectors would be moved to the cleaning position during the measurement. At NOOR I, the reflectivity of the mirrors is measured every morning, before sunrise in each area of the factory, with a 15R-USB type reflectometer. In order to simplify, further, the sampling process for measuring reflectivity, SENER (Spanish Company) proposed to divide the solar field into blocks of samples. Each sample block consists of a half collector on the left and the other one on the right (Fig. 2a). To identify the sample blocks, we attributed to each one of them an address. Each address begins with a letter (N or M), followed by the direction NORTH or SOUTH and the number of the row. For example, the sample highlighted in red has the coordinates (N, NORTH, 35), the one in blue corresponds to (M, NORTH, 33) and the sample highlighted in green has (M, SOUTH, 30). Note that the NORTH–SOUTH designation is more appropriate to respond to the distribution of soiling in the solar field than the EAST–WEST. The EAST–WEST designation also designates the complete collector, but only the left or right of the loop.

The reflectivity measurements must be taken on randomly selected samples. The personnel is supposed to make several measurements, for the mirror in its dirty state and only one in a cleaned state, to take it as reference and calculate, thereafter the CF. By browsing the different collectors, the reflectivity results are entered in a template. Then the cleanliness factor is calculated, for each section, each zone, and for the entire solar field. This information will be used to plan the cleaning; according to the results obtained, the area with the weakest CF is the one that will be cleaned that day. Figure 2b shows a map illustrating the samples of each template; each one of it contains the addresses of the collectors sampled from each area separately. This organization of the templates established by SENER facilitates the measurement of
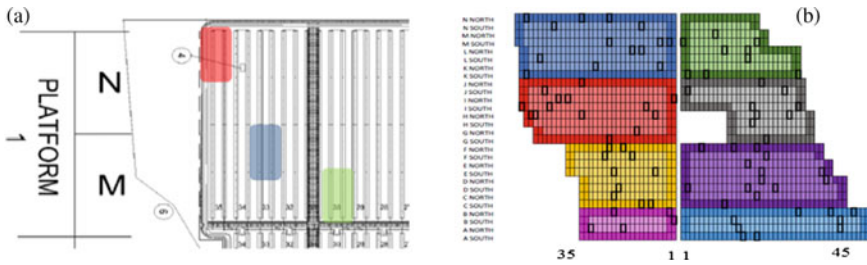
Fig. 2 Diagram representing, an example of assignment of an address to a sample (**a**) and a sample block selected, for template 1 (**b**)

the reflectivity to the personnel and allows a control of the eight main areas of the solar field.

The control procedure is carried out, in NOOR I, by seven templates, in order to ensure the maintenance of the solar field. Each template is filled in two days with the reflectivity measurements taken by the staff. Information on the state of the solar field is therefore only obtained after two days. In this case, the cleaning schedule will not be precise, since the climate can change overnight. It is therefore necessary to reduce the number of samples to be checked and to finish the entire template in one day.

3 Improvement of the Control Methodology

To improve the control methodology, the number of samples has been reduced from each area of the initial template. Then the average cleanliness factor was calculated, for the samples obtained after extraction. To test the reliability of the method, the values obtained were compared with those collected from the NOOR I, relating to the history of reflectivity measurements for each month of 2018. The followed procedure's first step is to remove 30% of the samples and then 50%. To perform data extraction, two criteria were applied. The first criterion is that for each template, we must keep the same average value of the CF of the solar field as the one obtained with the original template. As for the second, we must keep a good visual distribution in the used map of the solar field, to schematize the location of the chosen sample. From the analysis of the seven existing templates, we found out that the samples of two templates were similar. In this case, we have chosen to study only the five templates that are different from each other.

3.1 Results and Discussion

The five templates containing the 84 samples have been reduced by 30% to keep only 58 samples in total. These templates have been chosen so that they contain different collector addresses from each other, in order to ensure varied control of the different areas of the solar field. This extraction carried out using the statistical software was subsequently adjusted manually, in order to have a difference between the calculated values of CF and those obtained in 2018, less than 0.10%. If the difference is significant between the two values, the average of each zone of the solar field is compared and the samples in this zone are modified until the error becomes minimal. Table 1 gives the results obtained of the cleanliness factor, after extraction and their relative error. Analysis of these results shows that the values of the CF are generally similar to those already existing, with the exception of a few values of high relative error. They were noticed for the months of February, April, July, August, and September, distributed differently, in the considered templates. These differences can be due either to a problem in one of the equipment in the solar field, or to a strong disturbance that experienced one of the months of the year. This period is known by its very disturbed climate, which makes it possible to have dirtier areas than others and consequently also disturbed calculations. Based on the average relative error, the form factor values are obtained with a very satisfying average relative error, which does not exceed the tolerated error difference of 0.10% for the five studied templates. The visual distribution of the samples allows us to see that this distribution is satisfying and allows us to control the entire solar field. Figure 3a shows this distribution for template 1, with 30% reduction. Analysis of the cleanliness factor showed us that this factor has an average value of 96%, which slightly affects production.

The same procedure was followed, for a reduction of the number of samples by 50%. The results obtained showed a visual distribution of the samples that do not allow to control the entire solar field (Fig. 3b), on the one hand. On the other hand, the average relative error exceeds the tolerated threshold which was fixed to 10%. Figure 4 shows a comparison of the average absolute error (AAE), obtained for the templates after reduction of 30 and 50%. The AAE is more remarkable for the templates relating to an extraction of 50% exceed the difference in error tolerated, for the various templates. In this case, we can only reduce the samples by 30%, which gives results close to the experimental data with a minimum error difference. The proposed templates allow daily monitoring of the solar field, which increases the performance of cleaning planning with a rate of 50%. The five tested templates allow plant's staff to have an overview of the general state of the solar field in five days instead of fourteen days.

Table 1 Calculated cleaning factor (CF) and its relative error

Template	1		2		3		4		5	
Months	CF (%)	Er (%)	CF (%)	Er (%)	CF (%)	Er (%)	CF (%)	Er (%)	CF (%)	Er (%)
January	97.3	0.03	98.46	0.01	98.47	0.02	97.72	0.03	96.9	0.06
February	98.2	0.01	94.39	0.75	96.57	0	93.9	0	97.76	0.03
March	98.3	0	97.82	0.03	97.65	0	97.73	0.01	97.75	0
April	96.61	0.02	97.11	0.06	96.84	0.04	96.36	0.1	97.14	0.07
May	96.28	0.02	95.75	0.06	97.34	0.01	94.71	0	94.11	0.34
June	97.41	0.03	97.71	0.02	97.61	0.02	97.06	0.01	96.93	0.01
July	97.91	0.04	97.99	0.01	96.51	0.02	92.14	0.24	95.86	0.1
August	94.36	0.2	93.91	0.02	91.37	0.92	94.13	0.03	95.08	0.05
September	93.73	0.16	94.09	0.02	90.17	0.02	95.21	0.03	93.31	0.25
October	98.14	0.03	97.25	0.02	97.62	0.05	97.82	0.01	98.09	0.04
November	97.82	0.02	96.82	0.01	96.44	0.02	97.82	0.01	97.9	0.02
December	98.06	0.02	97.65	0.04	98.35	0.01	98.37	0.01	98.28	0.02
AAE (%)		0.05		0.09		0.09		0.04		0.08
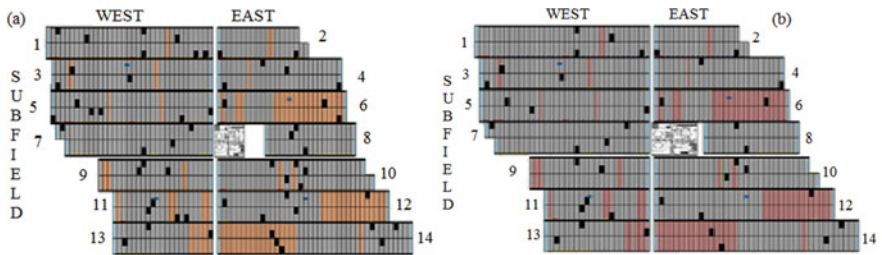
Fig. 3 Template 1 solar field map, with 30% reduction (**a**) and 50% (**b**)
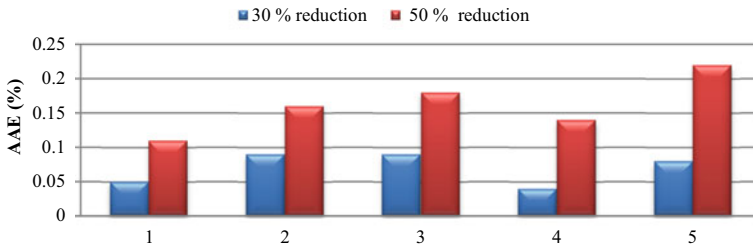
Fig. 4 Comparison of AAE for the 5 templates, after reducing 30 and 50% of the samples

4   Conclusion

The objective of this work was to analyze and minimize the solar field control period, to guarantee a good level of efficiency of the NOOR I thermal power plant. The improvement of the solar field control methodology was carried out by reducing the number of samples. This reduction has been validated, respecting two criteria. Each template must keep the same average value of the CF as the original templates, on the one hand, and on the other hand, a good visual distribution in the map of the solar field must be obtained. Based on this study, it was shown that reducing the test samples by 30% is sufficient to reduce the solar field test time by 50%. In perspective, a techno-economic analysis of the performance of the NOOR I plant will be evaluated, following this procedure.

References

1. Suman S, Khan MK, Pathak M (2015) Performance enhancement of solar collectors—a review. Renew Sustain Energy Rev 49:192–210
2. Ben Fares MS, Abderafi S (2018) Water consumption analysis of Moroccan concentrating solar power station. Sol Energy 172:146–151
3. Giostri A, Binotti M, Astolfi M, Silva P, Macchi E, Manzolini G (2012) Comparison of different solar plants based on parabolic trough technology. Sol Energy 86:1208–1221
4. Alami Merrouni A, Wolferfstetter F, Mezrhab A, Wilbert S, Pitz-Paal R (2015) Investigation of soiling effect on different solar mirror materials under Moroccan climate. Energy Procedia 69:1948–1957
5. Fernández-Garcíaa A, Álvarez-Rodrigob L, Martínez-Arcosc L, Aguiard R, Márquez-Payés JM (2014) Study of different cleaning methods for solar reflectors used in CSP plants. Energy Procedia 49:80–89
6. Aqachmar Z, Allouhi A, Jamil A, Gagouch B, Kousksou T (2019) Parabolic trough solar thermal power plant Noor I in Morocco. Energy 178:572–584
7. Ouammi A, Zejli D, Dagdougui H, Benchrifa R (2012) Artificial neural network analysis of Moroccan solar potential. Renew Sustain Energy Rev 16:4876–4889
8. Conceição R, Merrouni AA, Lopes D, Alae A, Silva HG, Bennouna EG, Collares-Pereira M, Ghennioui A (2019) A comparative study of soiling on solar mirrors in Portugal and Morocco: preliminary results for the dry season. In: AIP conference proceedings, vol 2126, p 220001
9. Bouaddi S, Ihlal A, Fern_andez-García A (2017) Comparative analysis of soiling of CSP mirror materials in arid zones. Renew Energy 101:437–449
10. Raza A, Higgo AR, Alobaidli A, Zhang T (2016) Water recovery in a concentrated solar power plant. AIP Conf Proc 1734:160014
11. Fernández-García A, Sutter F, Martínez-Arcos L, Sansom C, Wolferfstetter F, Delord C (2017) Equipment and methods for measuring reflectance of concentrating solar reflector materials. Sol Energy Mater Sol Cells 167:28–52

Seasonal Variation of Atmospheric Absorptivity in Ouarzazate

Ouassila Salhi, Mohammed Diouri, Abdelmoula Ben-tayeb, Ibtissam Marsli, Sara El Hassani, and Mohammed Ammine Moussaoui

Abstract Solar radiation is one of the most affordable energy intensive of all renewable energies as in Ouarzazate, a city known by the use of solar energy resource. The climate change system manifests many changes in this energy by absorption of solar radiation by the atmosphere and earth surface, and infrared radiation is also affected by this absorption. Using our model ACRA19, that describes the regional radiative balance with corresponding atmosphere components like temperature, incoming solar radiation at the top of the atmosphere, optical depth and surface albedo, we determine the absorptivity of solar radiation a_1 that observes high values during summer 0.27 for 500 nm. The reflectivity of solar radiation by the atmosphere r_1 is close to 0.36. This shows that the desert aerosols may attenuate solar radiation by either reflection or absorption at the level where the solar incident flux reaching the earth surface drops to 170 W.m^{-2}.

Keywords Dust aerosols · Absorption · Climate models · Solar radiation

1 Introduction

In the last years, the study of climate change and incident solar radiation helps to use solar energy resources. The absorption of solar and infrared radiation, as well as the effect of aerosols, is expected to have a significant impact on thermal equilibrium of the atmosphere, causing perturbation and significant changes that influence temperatures and incident solar flux at surface. The rise of particle concentrations in the air depending on season and climatic components has a crucial impact.

O. Salhi (✉) · M. Diouri · A. Ben-tayeb · I. Marsli · S. El Hassani · M. A. Moussaoui
Laboratory of Mechanics and Energetics, Team of Physics of Atmosphere, Faculty of Sciences, Mohammed First University, Oujda, Morocco
e-mail: salhi.ouassila@ump.ac.ma

A. Ben-tayeb
Mechanical Engineering Department, Faculty of Sciences and Technologies, Sultan Moulay Slimane University, Beni-Mellal, Morocco

Morocco's historical record indicates a scenario of increasing temperatures and a decrease in precipitation, which can cause severe droughts. Morocco is located in a region that receives very significant solar radiation, with over 2500 h of sunshine per year [1]. This energy is affected by climatic parameters such as optical thickness, greenhouse gases and albedo, which modify the absorptivity by the atmosphere or the reflectivity in space.

In this research, we chose Ouarzazate as a city that describes the situation, as well as its important site about renewable energies near to The Noor Ouarzazate Solar Complex, the largest solar farm in Africa. Desert dust [1, 2] affects the climate of Ouarzazate [3] and also the surface of solar panels. In terms of solar irradiance, aerosols play the key role of attenuation [4].

2   Materials and Methods

2.1   Measures Used

The study of this work is based on data from AERONET. AERONET (AErosol RObotic NETwork) is a consortium of terrestrial aerosol networks and remote sensing created by NASA and Lille 1 University, also in collaboration with national scientists. The AERONET collaboration provides globally distributed measurements of spectral aerosol optical depth (AOD), inversion products and perceptible water at over 500 sites, for eight wavelengths (340, 380, 440, 500, 675, 870, 1020 nm and 1640 nm). In this study, we focused only on the spectral total optical thickness at 0.5 μm ($TOT_{0.5}$) and ($TOT_{1.20}$) (Fig. 1).
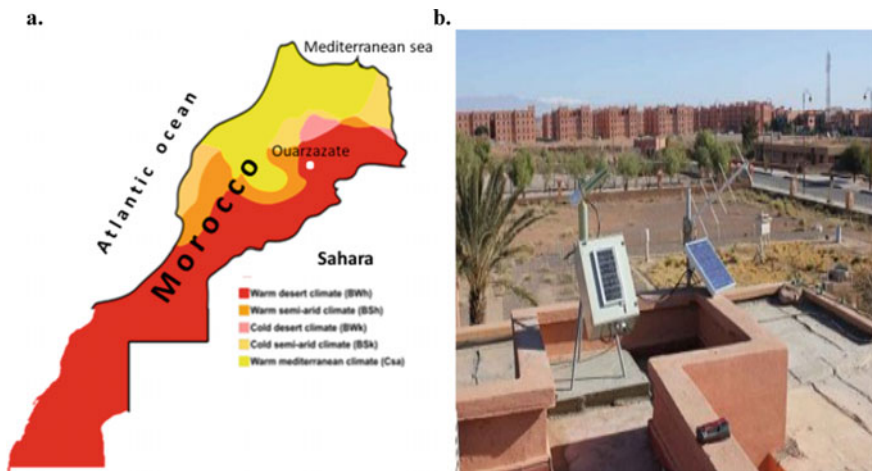
Fig. 1   a Site location. **b** A view of the sun photometer in Ouarzazate

Data from the Prediction of World Energy Resources (POWER) [5] project is also used. This work analyzes results on a large scale that is relevant to the renewable energy industry's and climate change research, particularly surface albedo, surface temperature and solar incident flux at the top of the atmosphere and at earth surface.

Ouarzazate is situated southeast of the Atlas Mountains and northwest of the Sahara. The meteorological station is located (30.92° N, 6.91° W and 1136 m) [6]. Ouarzazate is one of the sites with almost no industry but with a large renewable energy activity, specially the solar energy. The climate is hot and dry. The average maximum temperature in August is 35 °C for 2017 (Fig. 3). According to the Köppen-Geiger climate classification, the region is considered as BWh, and clouds are infrequent in spring.

2.2 Data Analysis Procedure

TOT (λ) is the sum of the spectral contributions of Rayleigh interactions τ_{Ray}, τ_{aer} of aerosols and τ_{gas} obtained from AERONET [4, 6].

$$\tau(\lambda) = \tau_{Ray}(\lambda) + \tau_{aer}(\lambda) + \tau_{gas}(\lambda) \tag{1}$$

τ_{Ray}: optical thickness due to Rayleigh scattering, $\tau_{aer}(\lambda)$: optical thickness due to aerosols and $\tau_{gas}(\lambda)$: optical thickness due to gas absorption (O_3, NO_2, CH_4 and H_2O).

The monthly average fluctuation of $TOT_{0.5}$ and $TOT_{1.02}$ for 2016 and 2017, data taken from AERONET, shows the period of the desert dust season that varies throughout the year (Fig. 2). The total atmospheric thickness for short wavelengths 500 nm is greater than long wavelengths of 1020 nm. $TOT_{0.5}$ reaches the maximum of 0.54 on August 2017 and the minimum of 0.10 on January. $TOT_{1.02}$ observes the same change tendencies with higher values on April and August.

Because of its location compared to the equator, the incident solar radiation for Ouarzazate reaches high levels of 474.5 W/m^2 on June and minimum of 226.21 W/m^2 on December (Fig. 2) (Fig. 3).

3 Results and discussion

Using the simple radiative balance model [7] (Fig. 4). The absorptivity and annual average of solar radiation by the atmosphere a_1 (Table 1) of the wavelength 1020 nm are 0.18 (2016) and 0.19 (2017); however, for 500 nm, the values are more important 0.37 (2016) and 0.42 (2017). The absorptivity a_1 takes the same seasonal fluctuation because the aerosol optical depth has a monthly period during spring and summer. The reflectivity of solar radiation by the atmosphere r_1 reaches its maximum of 0.33

Fig. 2 Monthly average of
TOT (λ) (2016 and 2017)
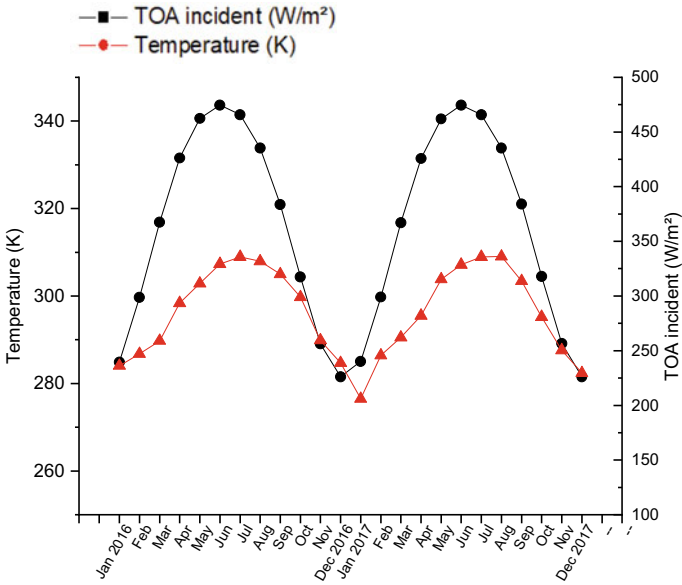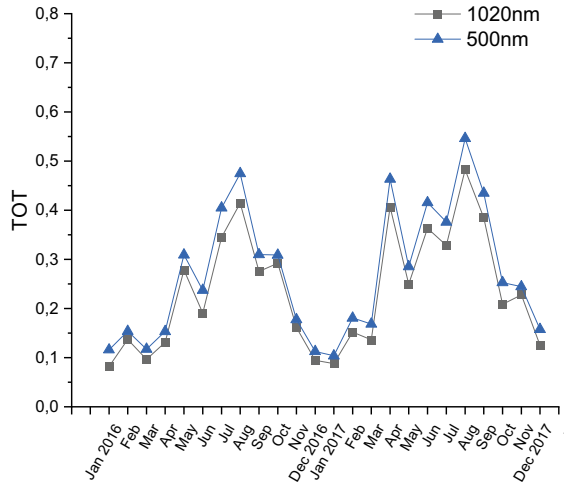
Fig. 3 Monthly average of temperature (K) and TOA solar incident (2016 and 2017)

on June (2016) and 0.36 (2017), and the minimum of 0.14 is observed on December (2016) and 0.11 on November (2017) (Fig. 5).

The variation of infrared radiation absorbed by the atmosphere exhibits high values, the infrared radiation by the surface to the atmosphere is almost totally absorbed, and the annual average $a_2 = 0.9$ (2016) and 0.85 (2017).
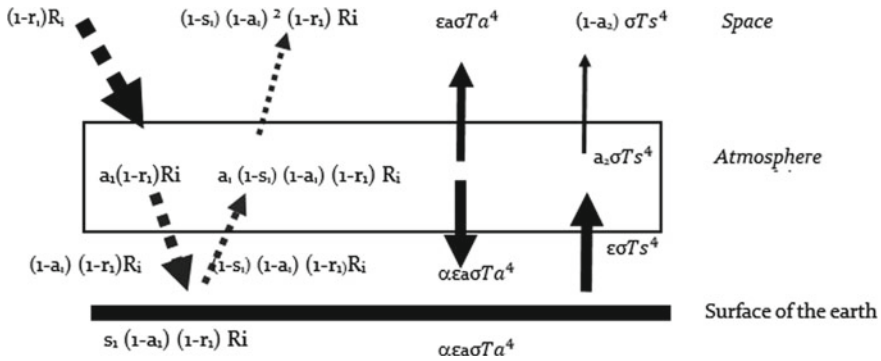
Fig. 4 Model of the greenhouse effect and the different parameters of absorptivity and reflectivity

Table 1 Optical parameter definitions

Symbol	Definition
a_1	Atmospheric visible absorptivity
a_2	Atmospheric thermal absorptivity
s_1	Absorptivity of solar radiation by the surface
r_1	Reflectivity of solar radiation by the atmosphere
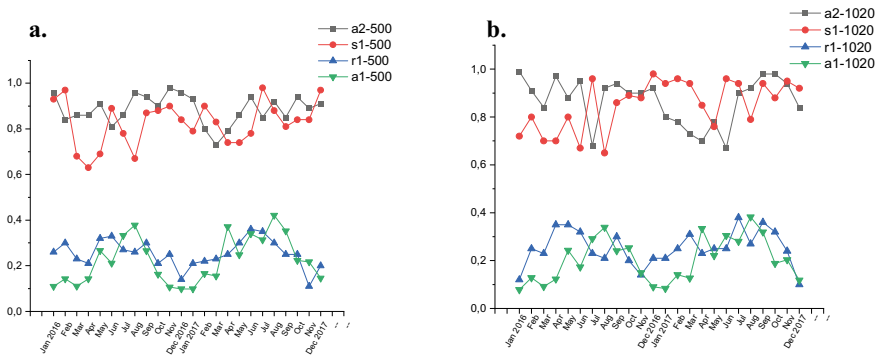
Fig. 5 Monthly average of atmospheric visible absorptivity a_1, absorptivity of solar radiation by the surface s_1, reflectivity of solar radiation by the atmosphere r_1 and atmospheric thermal absorptivity. **a** at 500 nm and **b** at 1020 nm

4 Conclusions

Solar energy is a very important source of energy, and areas with a large amount of insolation are an opportunity to make good use of it in the field of renewable energy. However, the determination of solar radiation absorptivity, for two years, which observes high values in summer and autumn in agreement with the high registered

AOD, and the determination of the reflectivity of solar radiation by the atmosphere show the important aerosol contribution to these two factors with the concerned radiation balance. This indicates that desert dust aerosols have large contribution to the absorptivity and reflectivity of solar radiation. In addition, infrared range absorptivity of thermal radiation by the atmosphere at the origin of the heating of the atmosphere registers near total absorption (0.9) for 2016.

References

1. Middleton NJ (2017) Desert dust hazards: a global review. Aeolian Res 24:53–63. https://doi.org/10.1016/j.aeolia.2016.12.001
2. Moorthy KK, Babu SS, Satheesh SK, Srinivasan J, Dutt CBS (2007) Dust absorption over the "Great Indian Desert" inferred using ground-based and satellite remote sensing. J Geophys Res 112(D9):D09206. https://doi.org/10.1029/2006JD007690
3. Elias T et al (2016) Aerosols attenuating the solar radiation collected by solar tower plants: the horizontal pathway at surface level. In: AIP conference proceedings, Cape Town, South Africa, p 150004. https://doi.org/10.1063/1.4949236
4. Ben-tayeb A, Diouri M, Meziane R, Steli H (2020) Solar radiation attenuation by aerosol: application to solar farms. Air Qual Atmos Health 13(2):259–269. https://doi.org/10.1007/s11869-020-00790-1
5. Zhang T, Chandler WS, Hoell JM, Westberg D, Whitlock CH, Stackhouse PW (2008) A Global perspective on renewable energy resources: Nasa's prediction of worldwide energy resources (power) project. In: Goswami DY, Zhao Y (eds) Proceedings of ISES world congress 2007 (vol I–V). Springer, Berlin, Heidelberg, pp 2636–2640. https://doi.org/10.1007/978-3-540-75997-3_532
6. Tahiri A, Diouri M (2018) Radiative forcing of the desert aerosol at Ouarzazate (Morocco). E3S Web Conf 37:03004. https://doi.org/10.1051/e3sconf/20183703004
7. Salhi O, Diouri M, Amine Moussaoui M, Marsli I, Meziane R (2021) Estimation of absorptivity of Earth-atmosphere system over the MENA areas. E3S Web Conf 314:03006. https://doi.org/10.1051/e3sconf/202131403006

Development of a New Monitoring Method for Rotating Machines Based on Maintenance 4.0—A Case Study of Unbalance and Misalignment

El Mahdi Bouyahrouzi and **Bachir El Kihel**

Abstract Implementing Maintenance 4.0-type solutions for Industry 4.0 has become an exciting topic for industrial and researchers. Integrating new technologies such as Maintenance 4.0 in factories is one of the most critical issues as long as these factories still need digitally their infrastructures since it aims at improving the performance of factories and strengthening the economic model. This paper aims to develop an innovative solution for diagnosing, monitoring, and real-time tracking faults in lathe machines, including unbalanced and misalignment faults. This solution is based on creating a fast Fourier transform (FFT) algorithm in Python, which facilitates global engine analysis according to the ISO 10816 standard.

1 Introduction

In 2011, the concept of Industry 4.0 appeared for the first time at the Hannover Exhibition in Germany [1], also known as the future industry, which is a technological development and organizational change introduced in the industrial world [2, 3]. With the advent of the Internet, and digitalization of production chains, new technologies are becoming part of everyday life in the factory [4], and Industry 4.0 is considered as an industrial revolution in the factory [5, 6].

Maintenance 4.0 is a technology for preventing machine issues by analyzing data in real time [7], identifying machine operating modes, and predicting failures before they occur [8], resulting in reducing unnecessary interventions and extends the machines' life. In this context, several researchers and industrialists are interested

E. M. Bouyahrouzi (✉) · B. El Kihel
Laboratory of Industrial Engineering and Seismic Engineering, National School of Applied Sciences ENSA, Mohammed First University, Oujda, Morocco
e-mail: elmahdi.bouyahrouzi@ump.ac.ma

B. El Kihel
e-mail: belkihel@yahoo.fr

Table 1 Summary of diagnostic and prognostic approaches discussed in the literature

Approach	Reference	Principles
Physical-based model	[11–13]	The use of an analytical model derived from laws of physics (mechanics, chemistry, electricity, hydraulics, …)
Data-driven model	[9, 14–18]	The use of a neural network models The use of statistical, probabilistic, and stochastic methods
Hybrid approach	[19, 20]	Using both models to estimate the system's current state and predict how long it will run before failure

in the development of maintenance at the level of the new generation of industry [9, 10].

Most Maintenance 4.0 articles focus on one of three approaches:

- Physical-based model,
- Data-driven model,
- Hybrid model.

Table 1 summarizes a review of articles that address these three approaches of Maintenance 4.0.

The monitoring and diagnosis of the condition of rotating machines have been based for years on vibration analysis to apply an effective Maintenance 4.0. Vibration analysis is a technique adapted to studying the maintenance of rotating machines, and it allows to foresee and plan the maintenance operations according to the production. Thanks to the fast Fourier transform (FFT) and IoT this approach has become more effective and efficient.

The interest of this study is the development of a new solution for the diagnosis, monitoring, and real-time tracking of faults in lathe machines, including unbalance and misalignment faults [21].

This paper is organized as follows: The first section presents an introduction of the processing systems used and the proposed method for real-time data acquisition. The second explains the overall architecture of the system and the tools needed to develop the monitoring and diagnostic process to detect engine faults and finally, a conclusion and perspectives.

2 Processing System

The processing of the system is based on two steps: The first step is intrusion detection system (IDS), and the second is data acquisition system (DAQ); the description of this steps is presented in the following paragraphs.
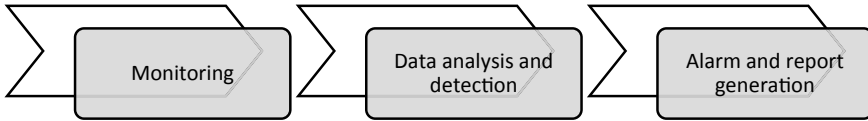
Fig. 1　Overall diagram of the IDS
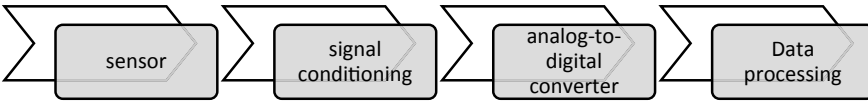
Fig. 2　System architecture of data acquisition systems

2.1　Intrusion Detection System (IDS)

An intrusion detection system (IDS) is a detection and tracking system that monitors machines and issues alerts when such activity is discovered [22].

The IDS is based on three main components, as illustrated in Fig. 1.

2.2　Data Acquisition System (DAQ)

The data acquisition system (DAQ) plays a role in monitoring systems and is used to collect data from various sensors on a machine. Then, this data is digitized for storage, and the DAQ sends the data to the control center for processing and subsequent presentation [23].

The block diagram of the DAQ data acquisition system is shown in Fig. 2. The data acquisition unit consists of four main modules.

3　Proposed Methodology

3.1　Project Architecture

The test bench used makes it possible to highlight a large majority of the vibration phenomena existing on rotating machines as well as the performances of the components, for example: bearings, motor (1.5 KW), and pump under controlled conditions. The experiments are performed on the testbed, as shown in Fig. 3.

In this experiment, the validation has been performed in two parts: The first part is to artificially create, on the test bench, two types of defects:
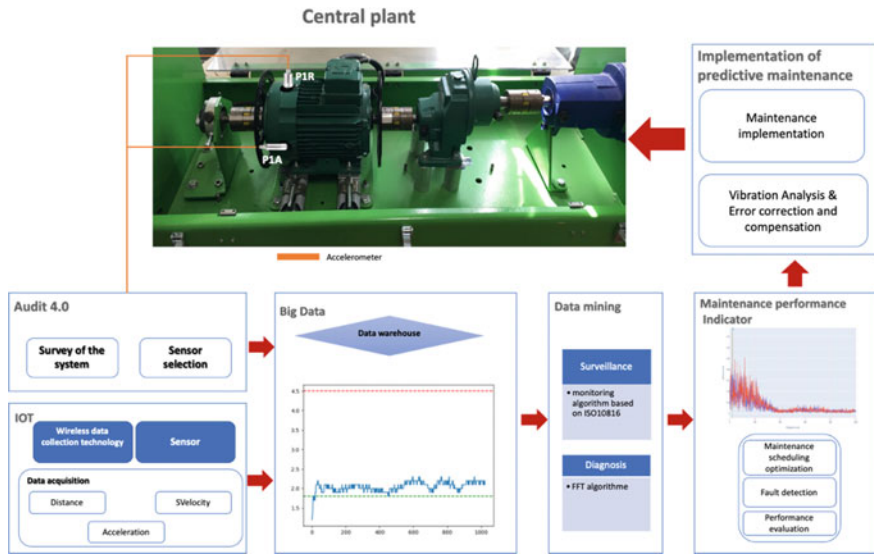
Fig. 3 Global project architecture

- The unbalance is highlighted on two planes by the graduated unbalance of two discs in rotation.
- The operating part allows the misalignment of the motor about the axis of rotation of the adjacent elements.

The second part is fixing the engine by three elastic supports to make it possible to study the filtration of the vibrations according to the frequency of rotation then to put under fatigue the engine which contains the two defects.

For data acquisition, accelerometers are placed on specific points on the motor in the axial and radial directions, and then the signals are transmitted to the computer via an electronic card in real time to be stored and processed with an algorithm based on the FFT.

3.2 Methods

First, the accelerometers are connected to specific points: axial and radial of the engine, and the sensors with an electronic card, to realize a data acquisition system (DAQ) that acquires the data in real time. An algorithm based on Python language has been developed to make the processing part, which makes a global analysis of the engine according to the ISO10816 standard [9, 17]. If the value of the RMS speed exceeds specific limits, the algorithm goes directly to the frequency analysis by applying an FFT algorithm to the acceleration. Finally, the database and curves are generated and saved automatically for further processing.

4 Results Analysis

4.1 Results of Monitoring

To ensure vibration monitoring, the overall level of vibration in velocity is measured. For this purpose, vibration measurements have been made at points P1A and P1R as shown in Fig. 4.

Figure 4 represents the RMS vibration velocity values in both axial and radial directions under healthy engine conditions.

Figures 4 and 5 show a huge difference between the vibrations recorded at the healthy condition and the vibrations recorded at the fault moment at the same point. Moreover, the vibration values of the motor exceed the critical alarm thresholds recommended by the ISO10816 standard [24] (the admissible point for small machines to work is 1.8 mm/s and at the value of 4.5 mm/s the machine breaks down), and the motor continues to work normally without any malfunction.

(a) measurement position axial P1A   (b) measurement position radiale P1R
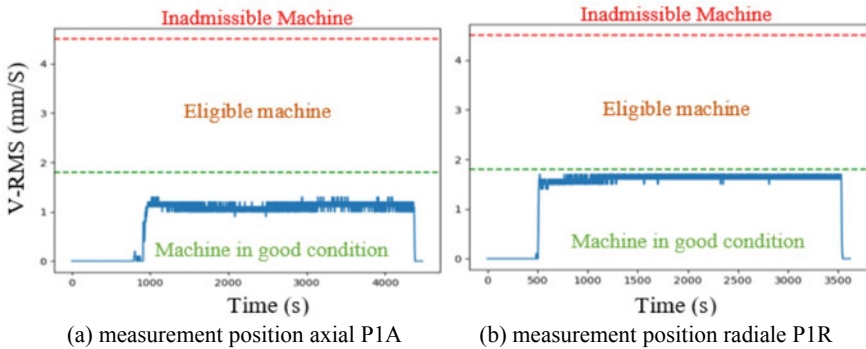
Fig. 4 RMS values in radial and axial directions: case of healthy motor

(a) measurement position axial P1A   (b) measurement position radiale P1R
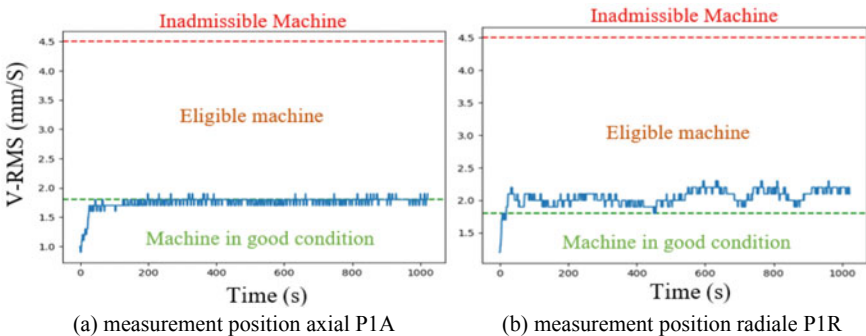
Fig. 5 RMS values in radial and axial directions: case of faulty motor

Once the fault detection is done, the monitoring and diagnosis start, allowing to track the state of the motor fault and its evolution over time to help in the decision-making process of the maintenance task to be performed for the monitored equipment. Determining the diagnosis part of the engine that is defective after the detection of an irregular manifestation of the vibrations is noted during the monitoring.

4.2 Results of Diagnostic

For any motor, unbalance and misalignment faults generate vibration spectra with unique frequency components. These frequency components and their amplitudes are used to diagnose the condition of the motor. The fault frequencies are linear functions of the motor speed. In our case, the speed rotation is set at 1200 rpm, which means that the frequency of rotation is around $f_0 = 20$ Hz and its equal harmonics $2f_0 = 40$ Hz, $3f_0 = 60$ Hz, and so on.

At the diagnostic level, the developed FFT algorithm is applied on the RMS acceleration.

Figure 6 shows the signal with a case of unbalance and misalignment fault on the motor. Both faults are characterized by the presence of several higher frequencies corresponding respectively to the frequency f_0 of the signal and its harmonics concerning the healthy state of the motor (Fig. 7).

According to the frequency analysis of point P1R, the highest peak is 0.07 m/s at the frequency of 20 Hz, and this frequency corresponds to f_0. Therefore, the increase in the amplitude of this peak proves that there is an unbalance problem at the motor.

The three spectra $2f_0$, $3f_0$, and $4f_0$ present harmonics of frequency 20 Hz. This indicates a misalignment of the motor shaft.

After processing, two files were generated:

- The first file is a.csv file which contains the values measured by the sensor in real time.
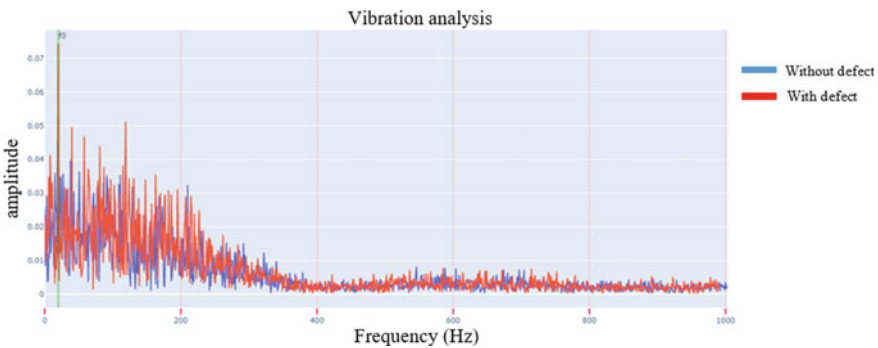
Fig. 6 Frequency spectrum of main control point P1R
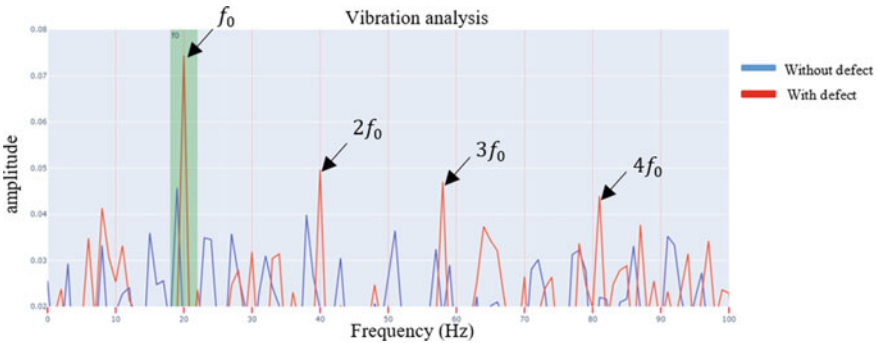
Fig. 7 Zoom on interval 0–100 Hz part
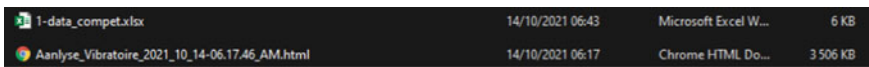
Fig. 8 Files generated by the algorithm

- The second file is a.html file which contains the FFT of the processed signal and the report generated by the algorithm (Fig. 8).

5 Conclusion and Perspectives

The main objective of this research work is to design a dynamic algorithm based on the Python language to diagnose, monitor, and track the defects of rotating machines in real time.

For the proposed algorithm, the selection of indicators is fundamental to giving effective results. Therefore, the following indicators have been extracted: V_RMS, A_RMS, A_PEAK, A_CREAT, followed by their evolution of the default values. For the acquisition and monitoring of the machine condition, ISO10816 is adopted and an IoT-based system to track the degradation of the machine in real time. Also, to distinguish changes due to the evolution of faults from load increases. For the diagnostic part, creating a FFT algorithm by the Python language gave a better result than other methods [25, 26].

The experimental evaluation, the method allowed to obtain better results, monitors the machine's state and diagnose it. The main advantage of this method is the efficient and automatic detection of faults without influencing the machine indicators, which makes it useful for Maintenance 4.0.

One of the perspectives of this paper can also touch on integrating a digital twin in the monitoring process. This digital twin could bring reliability by diagnosis and estimating the residual life. It can also do a simulation to test and predict operational changes in different scenarios. By leveraging the digital twin, companies can realize

substantial benefits such as improved operations, product and service innovation, and faster time to market.

References

1. Drath R, Horch A (2014) Industrie 4.0: hit or hype? [industry forum]. IEEE Ind Electron Mag 8(2):56–58. https://doi.org/10.1109/MIE.2014.2312079
2. Vaidya S, Ambad P, Bhosle S (2018) Industry 4.0—a glimpse. Procedia Manuf 20:233–238. https://doi.org/10.1016/j.promfg.2018.02.034
3. Pacchini APT, Lucato WC, Facchini F, Mummolo G (2019) The degree of readiness for the implementation of industry 4.0. Comput Ind 113:103125. https://doi.org/10.1016/j.compind.2019.103125
4. Sung TK (2018) Industry 4.0: a Korea perspective. Technol Forecast Soc Change 132:40–45. https://doi.org/10.1016/j.techfore.2017.11.005
5. Deniaud I, Marmier F, Michalak J-L (2020) Méthodologie et outil de diagnostic 4.0: définir sa stratégie de transition 4.0 pour le management de la chaîne logistique. Logistique Manag 28(1):4–17. https://doi.org/10.1080/12507970.2019.1693914
6. Aheleroff S et al (2020) IoT-enabled smart appliances under industry 4.0: a case study. Adv Eng Inform 43:101043. https://doi.org/10.1016/j.aei.2020.101043
7. Li Z, Wang K, He Y (2016) Industry 4.0—potentials for predictive maintenance. In: Presented at the 6th international workshop of advanced manufacturing and automation, Manchester, UK. https://doi.org/10.2991/iwama-16.2016.8
8. Chiementin MX, Rasolofondraibe ML, Contribution au processus de surveillance intelligente des machines tournantes: cas des roulements à billes. Thèse dirigée par Xavier Chiementin et Lanto Rasolofondraibe, p 137
9. Okoh C, Roy R, Mehnen J (2017) Predictive maintenance modelling for through-life engineering services. Procedia CIRP 59:196–201. https://doi.org/10.1016/j.procir.2016.09.033
10. Uhlmann E, Laghmouchi A, Geisert C, Hohwieler E (2017) Decentralized data analytics for maintenance in industrie 4.0. Procedia Manuf 11:1120–1126. https://doi.org/10.1016/j.promfg.2017.07.233
11. Chelidze D, Cusumano JP (2004) A dynamical systems approach to failure prognosis. J Vib Acoust 126(1):2–8. https://doi.org/10.1115/1.1640638
12. Kacprzynski G, Sarlashkar A, Roemer M, Hess A, Hardman B (2004) Predicting remaining life by fusing the physics of failure modeling with diagnostics. JOM J Miner Met Mater Soc 56:29–35. https://doi.org/10.1007/s11837-004-0029-2
13. Qiu J, Seth BB, Liang SY, Zhang C (2002) Damage mechanics approach for bearing lifetime prognostics. Mech Syst Signal Process 16(5):817–829. https://doi.org/10.1006/mssp.2002.1483
14. Elkihel A, Derouiche I, Elkihel Y, Bakdid A, Gziri H (2020) Artificial intelligence based on the neurons networks at the service predictive bearing, p 10
15. Daniyan I, Mpofu K, Oyesola M, Ramatsetse B, Adeodu A (2020) Artificial intelligence for predictive maintenance in the railcar learning factories. Procedia Manuf 45:13–18. https://doi.org/10.1016/j.promfg.2020.04.032
16. Xu G, Hou D, Qi H, Bo L (2021) High-speed train wheel set bearing fault diagnosis and prognostics: a new prognostic model based on extendable useful life. Mech Syst Signal Process 146:107050. https://doi.org/10.1016/j.ymssp.2020.107050

17. Einabadi B, Baboli A, Ebrahimi M (2019) Dynamic predictive maintenance in industry 4.0 based on real time information: case study in automotive industries. IFAC-Pap 52(13):1069–1074. https://doi.org/10.1016/j.ifacol.2019.11.337
18. Lin Y, Li X, Hu Y (2018) Deep diagnostics and prognostics: an integrated hierarchical learning framework in PHM applications. Appl Soft Comput 72:555–564. https://doi.org/10.1016/j.asoc.2018.01.036
19. Medjaher K, Zerhouni N (2013) Hybrid prognostic method applied to mechatronic systems. Int J Adv Manuf Technol 69(1):823–834. https://doi.org/10.1007/s00170-013-5064-0
20. Borutzky W (2020) A hybrid bond graph model-based—data driven method for failure prognostic. Procedia Manuf 42:188–196. https://doi.org/10.1016/j.promfg.2020.02.069
21. Li C, Sánchez R-V, Zurita G, Cerrada M, Cabrera D (2016) Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning. Sensors 16(6):895
22. Youcef Khodja A, Guersi N, Saadi MN, Boutasseta N (2020) Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks. Int J Adv Manuf Technol 106(5–6):1737–1751. https://doi.org/10.1007/s00170-019-04726-7
23. Shaoxiang M et al (2021) Design of the data acquisition system of acceleration grid power supply for CFETR N-NBI prototype. Fusion Eng Des 169:112643. https://doi.org/10.1016/j.fusengdes.2021.112643
24. Carounagarane C, Chelliah TR, Appa Sarma S (2020) Simulation and experimental illustration of vibration at load rejection in a continuously overloaded large hydrogenerator. IET Renew Power Gener 14(9):1550–1558. https://doi.org/10.1049/iet-rpg.2019.0847
25. Abouelanouar B, Elkihel A, Gziri H, Jeffali F (2020) Heat generation in materials by mechanical vibrations. Mater Today Proc 31:S168–S174. https://doi.org/10.1016/j.matpr.2020.07.177
26. Patil SS, Gaikwad JA (2013) Vibration analysis of electrical rotating machines using FFT: a method of predictive maintenance. In: 2013 Fourth international conference on computing, communications and networking technologies (ICCCNT), Tiruchengode, July 2013, pp 1–6. https://doi.org/10.1109/ICCCNT.2013.6726711

CSP Mirror Soiling Modeling from Measured Weather Factors and Forecasting Using OpenWeatherMap Server

Ayoub Oufadel, Alae Azouzoute, Massaab El Ydrissi, Hicham Ghennioui, El Ghali Bennouna, and Alami Hassani Aicha

Abstract The problem related to dust accumulation and its impact on the performance of concentrating solar power (CSP) installations, as well as the problem of optimal management of cleaning frequency to optimize the consumption of water and cleaning products makes the quantification of the soiling rate indispensable. In this study, artificial neural network (ANN) model is developed to predict the soiling rate in function of meteorological parameters such as temperature, humidity, wind speed, and wind direction. The data of soiling rate and the meteorological data are extracted using tracking cleanliness sensor (TraCS) and weather sensors, respectively, installed at the weather station in Missour. The results show that the ANN model is a powerful model to predict the soiling rate regarding the complexity of this phenomenon. This will be with high interest for the operation and maintenance sector to develop a cleaning strategy with high efficiency and consequently decrease the operation and maintenance costs.

Keywords CSP · Dust · Soiling · ANN · Modeling · OpenWeatherMap

1 Introduction

The solar energy market is growing in the MENA (Middle East and North Africa) region with high solar radiation as the direct normal irradiation (DNI) could reach 2800 kWh/m^2 [1]. For this reason, several solar power plants are already operational while others are under construction. The total installed capacity from CSP technology

A. Oufadel (✉) · M. El Ydrissi · H. Ghennioui · A. H. Aicha
Laboratory of Signals, Systems, and Components, University Sidi Mohamed Ben Abdellah, Fez, Morocco
e-mail: ayoub.oufadel@usmba.ac.ma

A. Azouzoute · M. El Ydrissi · E. G. Bennouna
Institut de Recherche en Energie Solaire et Energies Nouvelles—IRESEN, Green Energy Park, Km 2 Route Régionale R206, Benguerir, Morocco

A. Azouzoute
MEAT, University Mohamed V in Rabat, EST Sale, 227 Avenue Prince Héritier, Sale, Morocco

will reach 1000 GW by 2050, which will avoid 2.1 Gt of CO_2 emissions each year and contribute to 11% of the world's electricity production [2].

However, the most important parameters influencing site selection or project feasibility are solar radiation and climatic conditions [3]. Dust is a key parameter generated by a complicated process that depends on many parameters like localization, environmental factors, and the mirror's surface characteristics [4]. This natural phenomenon known as "soiling" considerably affects the optical and electrical performance of solar power plants [5–8]. Dust accumulation on the surface of concentrators and PV modules could decrease the electrical production by, respectively, 26% and 27%, especially in the summer season in a region with a semi-arid climate [9, 10].

With the deployment and development of the global navigation satellite system (GNSS), satellite stations can obtain predictions of temperature (T), wind speed (Ws), wind direction (Wd), and relative humidity (RH) with high resolution. Using those meteorological parameters for soiling prediction is very promising. In this work, a common model for soiling prediction including analysis and processing of Missour weather data is developed. The OpenWeatherMap server from the satellite database is used as a source of environmental parameters. This model is important to the operation and maintenance sector in order to define the best cleaning strategy for solar power plants.

2 Material and Methods

2.1 Soiling Data Collection

In this study, the meteorological data are extracted from the soiling station installed in Missour, Morocco (32.86031° N, − 4.10725° E), this station contains different sensors for the measurement of meteorological data including wind speed and direction, ambient temperature, relative humidity, and solar resource measurements. Tracking cleanliness system "TraCS" is used to quantify effect of soiling (see Fig. 1). This system is developed by the German Aerospace Center (DLR) and commercialized by CSP Services in order to help the stakeholders and the operators of concentrated solar plant (CSP) to mitigate against soiling phenomenon. This system is equipped with two pyrheliometers from Kipp and Zonen and a secondary mirror mounted on a sun tracker. The first pyrheliometer is directed toward the sun in order to measure the direct normal irradiance (DNI), and the second is directed toward a secondary mirror to measure the reflected DNI. The reflected DNI measured from the secondary mirror is affected with soiling accumulated on the surface of the mirror. The comparison between the direct DNI and the reflected DNI defines the soiling rate at the specific site. The details of the instrument of measurements at the soiling station are described in Table 1. All the sensors offer analog outputs for easy connection to virtually any data logger. With a Modbus interface and an amplified analog

Fig. 1 Soiling station equipped with TraCS and meteorological instrument measurement

Table 1 List of equipment of the meteorological center

Database notation	Measurement	Instrument
Temp	Air temperature in [°C]	Campbell scientific CS215 sensor
RH	Relative humidity in [%]	Campbell scientific CS215 sensor
WS	Wind speed in [m/s]	R. M. Young 05103-5 sensor or wind speed in m/s, measured with NRG 40H anemometer
WD	Wind direction in [°]	R. M. Young 05103-5 sensor or wind direction measured with NRG 200 wind direction sensor
DNI	Direct normal irradiance in W/m^2	K&Z CHP1 pyrheliometer

output, these measurements are recorded in Excel files, each file for each day with a sampling step of 10 min from 01/01/2015 to 01/08/2017.

2.2　Soiling Index Measurement

The soiling index λ is defined as the normalized ratio between the direct normal irradiance reflected by the secondary mirror DNIref and the direct normal irradiance measured directly from the sun DNI, as shown in (Fig. 2a). The difference between the two signals of DNI is explained by the effect of soiling on the surface of the secondary mirror, which decreases the intensity of the measured irradiance. In order to quantify the soiling rate, only data between 12 and 2 pm will be averaged to avoid disturbance data [11, 12].
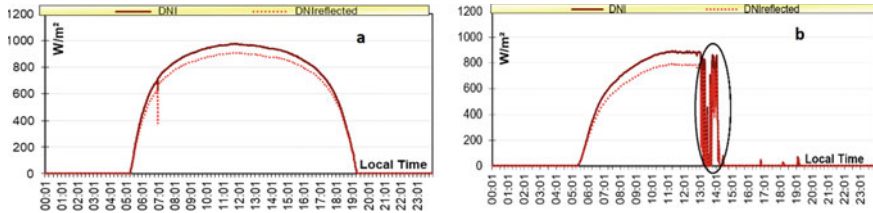
Fig. 2 Variation of the DNI on a clear sky (**a**) and cloudy sky (**b**) during a day

To solve the problem of fluctuation in data, especially for cloudy days, all data with DNI is less than 20% of the maximum value is removed as it can be in (Fig. 2b). On the other hand, data corresponding to cleaning days (human intervention) are also not considered in the database. The soiling index is calculated using the following equation:

$$\lambda = 1 - \text{Cleanliness with} : \text{Cleanliness} = \frac{\rho}{\rho_0} \text{and} \ \rho = \frac{\text{DNIref}}{\text{DNI}} \qquad (1)$$

ρ_0 corresponds to the maximum weighted reflectance measured at the clean state of the mirror.

There is no direct relationship between meteorological parameters and λ, but rather with the change of these parameters. To this end, the modelization of soiling is done considering the difference from a day to the following day $\Delta\lambda = \lambda i - \lambda i - 1$. The positive value of this parameter represents an increase of soiling, however, a negative value defines a decrease of soiling.

2.3 Meteorological Parameters Versus Soiling

In the literature, the average daily value of the meteorological parameter is used to describe the change during the day, but the average value cannot describe the daily variation of this parameter. For this reason, for each parameter, the average value and extremes (minimum and maximum) are considered, except for the wind direction, where the average value is considered as the same extremes are noticed every day.

These factors include WS, T, and RH.

Figure 3 shows the variation in cleanliness over 64-days with relative humidity, wind speed, and temperature. It can be inferred that an increase in relative humidity leads to a decrease in cleanliness. High humidity led to the formation of dew, which increase the cementation of dust particles on the surface of the secondary mirror [12], and eventually leads to severe optical losses of the reflector. The temperature is in the opposite variation regarding relative humidity in which the temperature with high values especially in summer increase dry deposition of dust particles. This impact is relatively not significant in comparison with the presence of high relative humidity
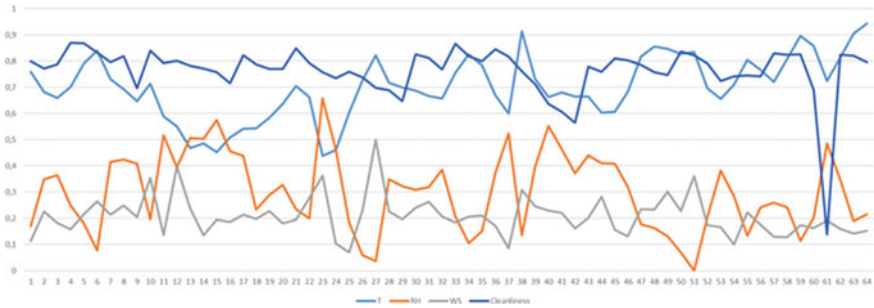
Fig. 3 Soiling index of the current day minus the day before

values. On the other hand, wind speed is considered one of the important factors increasing the effect of soiling. As it can be noticed from Fig. 3, an increase in wind speed is followed by a decrease in the cleanliness value.

2.4 Soiling Modeling

2.4.1 Linear Regression

Single MLR: a linear model is a statistical model in which the soiling function is defined with a linear form in function of meteorological parameters, as follows:

$$\Delta\lambda = \beta_1 + \beta_2 * X_1 + \cdots + \beta_n * X_n \tag{2}$$

With X_i is the meteorological factor and β_i is the coefficients corresponding to each factor.

MLR with interaction: interaction model represents the effect of meteorological variables to each other. This model is expressed as follows:

$$\Delta\lambda = \beta_0 + \beta_1 * X_1 + \cdots + \beta_n * X_n$$
$$+ \mu_1 * X_1 * X_2 + \cdots + \mu_i * X_{n-1} * X_n \tag{3}$$

To calculate the β_i coefficients for both models, XLSTAT extension methods is used.

2.4.2 ANN

The ANN model used is composed of an input layer, a hidden layer, and an output layer. In addition, the sigmoid transfer function is used in the output layer and the

Table 2 Performance indexes of MLR models

Model	Observations	R^2	RMSE	MSE
MLR simple	920	0.452	0.045	0.002
MLR with interaction	920	0.483	0.031	0.0007

predicted soiling rate is between 0 and 1, while the hyperbolic tangent transfer function is selected for the input and hidden layers. The input layer was composed of 12 neurons representing the input parameters each parameter relates to an input layer neuron. In addition, the Keras library and the sequential mode is used as the architecture building of the model. For the compilation, model the mean-squared-error and loss function Adam-optimizer are defined. Finally, the ANN model is trained and evaluated 10 times to obtain the best results and performances.

3 Results and Discussion

3.1 MLR Models

By applying the two linear regression models to the database, XLSAT generates a file containing the regression results and the results are illustrated in Table 2. The results show that the best fit of the model is with a correlation factor of $R^2 = 0.48$, which means that the MLR with interaction could explain 48% of the soiling data. It can be concluded that the linear model is limited to explaining the relationship between soiling and meteorological parameters.

3.2 ANN

After the execution of our program, the model starts in the learning phase to update the weights of each interconnection. After that, it is able to automatically stop the learning using the performance of our model on the validation data. The three phases (training, validation, and test) converge toward the minimum root mean square error, and the learning is stopped after 12 iterations or the network enters a phase of overlearning (the rote learning of the training base examples), and Table 3 shows the details of each step.

Figure 4 shows the regression graph between the real data $\Delta\lambda$ and the predicted data $\Delta\lambda'$, and the solid line represents the best-fitting linear regression. The learning data indicate a good fit, and the validation and test results also show significant *R*-values. The scatter plot is useful to show that some data points are poorly fitted. For example, there is a data point defined as an outlier, which could be explained by measurement errors or the existence of other influencing factors.

Table 3 Performance indexes of ANN model

Data	Observations	R^2	RMSE	MSE
Training	736	0.830	0.0189	0.151×10^{-2}
Validation	92	0.857	0.0106	0.105×10^{-2}
Test	92	0.801	0.0345	0.396×10^{-2}

Fig. 4 Regression graph
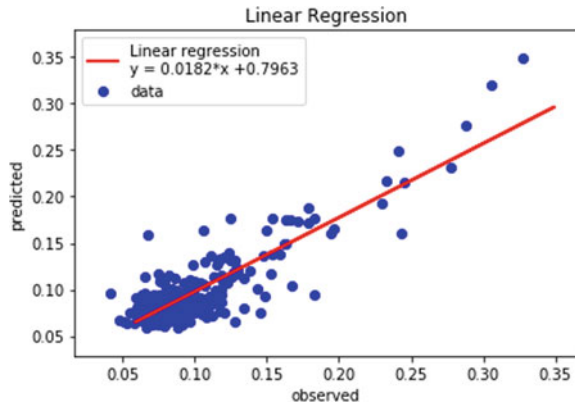
3.3 Soiling Index Prediction Using OpenWeatherMap Server

In this section, OpenWeatherMap is used as a source of meteorological parameters to predict $\Delta\lambda$ based on satellite images. OpenWeatherMap server is derived from the object notation of the JavaScript language (JSON) which is a formal configuration object describing a list of entries, Json-GUI is able to build a form frame interface [13]. It allows the representation of structured information as XML, for example, by specifying the geographical coordinates of the site and the server license, the result is in the form of hourly observations, for 48 h for the trial version and can be increased by purchasing a premium license.

For the prediction of $\Delta\lambda$, Python models and data from the OpenWeatherMap are developed. After the execution of the model, first, saving the model in two files, one with the extension ". json" containing the neural network architecture and another one with the extension ".h5" containing the weights "Wij". On the other hand, satellite images data are analyzed in order to retrieve only the hourly data and calculate the mean values and extremes of each factor, then the standardization of these data to predict Δ_λ by the function, "model.predict(data)". Finally, in order to simulate the cleaning effect, soiling classification approach is determined. The cleanliness value (measured in the field) is divided into four classes: (i) clean, (ii) semi-clean, (iii) dirty, and (iv) very-dirty (see Fig. 5).
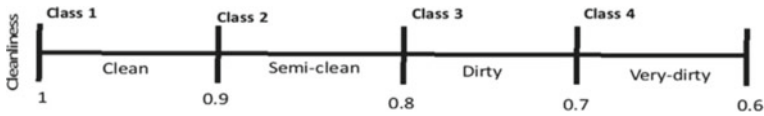
Fig. 5 Cleanliness classes [14]

4 Conclusion

The data collected from Missour soiling station are used to understand the relationship between the soiling rate and the meteorological parameters such as relative humidity, temperature, wind speed, and wind direction. The data show that the soiling phenomenon is significantly related to the meteorological parameters and is relatively high in the presence of high temperature and high wind speed as in the summer season. The relative humidity is also an important parameter that increases the soiling phenomenon, especially in the presence of high dust concentration.

Estimating soiling losses using the neural network shows great potential to address the complexity of soiling by establishing a combined effect of the factors influencing it. A direct-fed multilayer neural network model with 12 neurons in the cache layer is created by combining different meteorological parameters. The results show that the ANN model is with a good correlation factor up to 0.85, while the linear models failed to exactly describe the variation of soiling. On the other hand, the OpenWeatherMap is introduced as a powerful tool to predict soiling from satellite images.

References

1. Azouzoute A, Merrouni AA, Touili S (2020) Overview of the integration of CSP as an alternative energy source in the MENA region. Energy Strategy Rev 29:100493
2. Merrouni AA, Elalaoui FE, Ghennioui A, Mezrhab A, Mezrhab A (2018) A GIS-AHP combination for the sites assessment of large-scale CSP plants with dry and wet cooling systems. Case study: Eastern Morocco. Solar Energy 166:2–12
3. Azouzoute A, Chouitar M, Garoum M, Bennouna EG, Ghennioui A (2019) A new PV soiling monitoring device for optimized cleaning strategy. 2190(1):020068
4. Picotti G, Borghesani P, Cholette M, Manzolini G (2018) Soiling of solar collectors–Modelling approaches for airborne dust and its interactions with surfaces. Renew Sustain Energy Rev 81:2343–2357
5. Azouzoute A, Merrouni AA, Garoum M (2020) Soiling loss of solar glass and mirror samples in the region with arid climate. Energy Rep 6:693–698
6. Azouzoute A, Merrouni AA, Garoum M, Bennouna EG (2019) Comparison of soiling effect of two different solar mirrors in mid-south of Morocco. In: AIP conference proceedings, vol 2126, no 1, Art. no 1. https://doi.org/10.1063/1.5117699
7. Azouzoute A, Garoum M, Jeffali F, Ghennioui A (2020) Experimental study of dust effect on the transmission of a glass PV panel for a fixed and tracking system. Mater Today Proc 27:3091–3094
8. Zitouni H et al (2021) Experimental investigation and modeling of photovoltaic soiling loss as a function of environmental variables: a case study of semi-arid climate. Sol Energy Mater Sol Cells 221:110874

9. AAzouzoute A et al (2021) Modeling and experimental investigation of dust effect on glass cover PV module with fixed and tracking system under semi-arid climate. Solar Energy Mater Solar Cells 230:111219. https://doi.org/10.1016/j.solmat.2021.111219

10. Azouzoute A, Merrouni AA, Garoum M, Bennouna EG, Ghennioui A, Ydrissi ME (2019) The impact of optical soiling losses on the electrical production of CSP power plant. In: AIP conference proceedings, vol 2123, n 1, Art. no 1. https://doi.org/10.1063/1.5117017

11. Azouzoute A et al (2021) Developing a cleaning strategy for hybrid solar plants PV/CSP: case study for semi-arid climate. Energy 228:120565. https://doi.org/10.1016/j.energy.2021.120565

12. Maftah A, Azouzoute A, El Ydrissi M, Oufadel A, Maaroufi M (2021) Soiling investigation for PV and CSP system: experimental and ANN modelling analysis in two sites with different climate. Int J Sustain Energy 1–17

13. Galizia A, Zereik G, Roverelli L, Danovaro E, Clematis A, D'Agostino D (2019) Json-GUI—a module for the dynamic generation of form-based web interfaces. SoftwareX 9:28–34

14. Gallassi HE et al (2021) Proceedings of the 2nd international conference on electronic engineering and renewable energy systems ICEERE, 13–15 Apr 2020, Saidia, Morocco

Modeling a Real-Time Prediction System for Solar Collector Reflectivity Using Fuzzy Petri Net

A. Serji, E. B. Mermri, and M. Blej

Abstract This paper presents a simple and very efficient prediction system where we utilized fuzzy Petri nets (FPN) to construct a real-time prediction system as a modeling tool. Our model aimed at determining future reflectivity based on rules. An expert system (ES) extracts several rules from our database to predict reflectivity, with a necessary certainty factor (CF) for every rule. The essential step of constructing this model is transitioning from fuzzy production rules (FPR) to FPN. The based model describes a real-time system that reacts to the environment and speed of the environment. After getting the results, we validate the model by comparing the real data with predicted data. Then, we can start predicting the reflectivity in real time, represented by a dynamic graph and dynamic database for selecting an appropriate time for cleaning the collectors by a human expert or an ES.

Keywords FPN · FPR · Real-time system · ES

1 Introduction

Scientific research has worked for years to solve renewable energy difficulties to make power plants more effective and productive, but climate change has interrupted the proper functioning of solar power plants in recent years. Furthermore, these plants are situated in dusty environments with traits and situations that influence dust deposition. Extreme weather conditions collect dust on solar collectors (mirrors), necessitating cleaning regularly to keep them in good working order. Because this problem is site specific and the dust qualities of each site are different, awareness of

A. Serji (✉) · E. B. Mermri · M. Blej
Mohamed I University, Oujda, Morocco
e-mail: abdelilah.serji@ump.ac.ma

E. B. Mermri
e-mail: e.mermri@ump.ac.ma

M. Blej
e-mail: m.blej@ump.ac.ma

numerous aspects affecting solar collector cleaning activities is required. A prediction of reflectivity is one of the most powerful solutions in the world of renewable energy. A cleaning decision is made after an expert system, an FPN model, or a human expert observes the results for a period determined by the human expert. FPN prediction models are widely utilized in various research, such as predicting confidence values for bases called in DNA sequencing [1] and predicting the distance with recognition of a soccer ball to emulate human vision [2]. Our FPN model reflectivity prediction system was created to use predicted data as a fundamental rule to conclude the cleaning decision of solar collectors. The use of a reinforced learning algorithm neural network is one of the recommended solutions [3], and it is an effective method because this system uses an extensive database, the weather conditions as inputs, and the reflectivity as an output. In terms of data, our model is similar, but the issue is that we have little data to produce accurate results if we use the neural network. Therefore, we decided to use FPN because it is a rule-based tool and is also highly handy for simulating a real-time system. This paper presents a new prediction system that combines Petri nets and fuzzy logic. The truth-degree tokens of FPN are set according to the data provided by the fuzzification system in this model. However, the fuzzy reasoning mechanism is based on certainty factors, and the concurrent calculation mechanism of Petri nets dominates the prediction in real time.

2 Modeling and Results of the Prediction System

2.1 Preparation of the Rules Base Using FPR and ES

Most solar power plants employ cleaning procedures devised by human experts but with a high error rate. We cannot wholly replace him in this domain, but we can develop these decisions by using a model based on ground rules extracted from the data to simulate the performance of a human expert. FPR is widely used in knowledge-based ES to represent fuzzy, uncertain, imprecise, and ambiguous concepts. They are usually presented in the form of fuzzy IF-THEN (1) rules in which fuzzy sets designate both the antecedent and the consequent. These rules are extracted with an ES that identifies association rules with their CF and threshold λ. The main elements for modeling are rules describing the fuzzy relationship between several propositions. The rules base is conditional phrases with AND-OR (1) made up of words considered linguistic variables, variables whose values are not numbers but words or sentences in a natural or artificial language [4]. We extract a set R of FPR using an ES, where $R = R_1, R_2, \ldots, R_n$ and the general formulation of the ith FPRs is as follows [5]:

$$R_i : \text{IF } d_j \text{ THEN } d_k \text{ (CF} = \mu_i, \lambda_i) \tag{1}$$

where d_j and d_k are fuzzy propositions that describe meteorological conditions, with d_j = DNI (Direct Normal Irradiation), T (Temperature), H (Humidity), WS (Wind

Fig. 1 Membership function of the input 'Temperature'
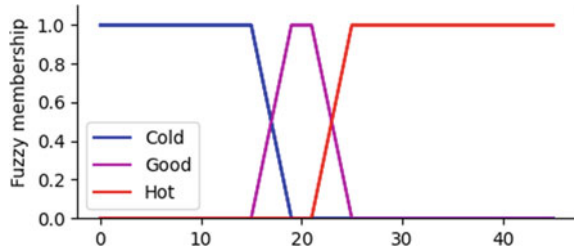
Speed) presenting four conditions, each with three fuzzy intervals, quantities created
by sensors present these conditions, so it is required to translate these quantities
into fuzzy variables DNI = "cloudy, intermittent, clear", T = "cold, good, hot", H =
"low, moderate, high", WS = "low, moderate, high". Fuzzification is an operation
performed by the membership function to establish a relationship between the degree
of truth α of the fuzzy variable and the corresponding input quantity. Each interval
refers to a trapezoidal membership function (see Fig. 1) that allows you to define
the degree of truth of the fuzzy variable. The network's output is R (Reflectivity),
determined by three fuzzy intervals R = "weak, medium, strong".

The first step is to prepare the basic rules, and after converting data from quantita-
tive to qualitative, we can extract the rules with their certainty factor μ and threshold
λ; the two parameters are derived using conditional probability. Our expert system
generated 43 rules with a factor of certainty greater than or equal to 0.5 and a thresh-
old greater than or equal to 0.3. The 43 rules contain four propositions in the form
of fuzzy variables called "linguistic terms" and a consequent proposition.

$$R_n: \text{IF } a_{1n} \text{ AND } a_{2n} \text{ AND } a_{3n} \text{ AND } a_{4n} \text{ THEN } C \text{ (CF} = \mu_n, \lambda_n, n = 1 \dots 43) \quad (2)$$

where $a_{1n}, a_{2n}, a_{3n}, a_{4n}$ are the input fuzzy variables d_j in (1), C is the output fuzzy
variable d_k in (1), and n is the number of rules, these rules can be graphically rep-
resented using an FPN (see Fig. 2), where each rule corresponds to a rule of the
linguistic description.

2.2 FPN Model for Reflectivity Prediction System

Most FPN models are employed for fuzzy reasoning and decision-making and fuzzy
knowledge representation. Looney [6] is the first researcher who proposed merging
Petri nets (PN) with fuzzy logic (FL) and presented an FPN model with a fuzzy
reasoning algorithm for rule-based decision-making. After two years, Chen et al. [5]
suggested an FPN model representing the fuzzy production rules of a rule-based sys-
tem with an efficient method for automatically performing fuzzy reasoning. FPN, on
the other hand, can be used to discover potential inconsistencies in fuzzy knowledge
bases [7]. A collection of eight tuples can define a generalized FPN [5] structure:

$$FPN = (P, T, D, I, O, f, \alpha, \beta) \tag{3}$$

where $P = \{p_1, p_2, \ldots, p_n\}$ denotes a set of places. $T = \{t_1, t_2, \ldots, t_m\}$ denotes a set of transitions. $D = \{d_1, d_2, \ldots, d_n\}$ denotes a set of propositions, $P \cap T \cap D = \emptyset$ and $|P| = |D|$. $I : T \longrightarrow P^\propto$ is the input function, a mapping from transitions to bags of places. $O : T \longrightarrow P^\propto$ is the output function, a mapping from transitions to bags of places. $f : T \longrightarrow [0, 1]$ is an association function, a mapping from transitions to real values between zero and one. $\alpha : P \longrightarrow [0, 1]$ is an association function, a mapping from places to real values between zero and one. $\beta : P \longrightarrow D$ is an association function, a bijective mapping from places to propositions.

In an FPN model, places represent meteorological conditions; transitions represent conditions for rules represented by FPR, and tokens represent truth degrees $\alpha(p_i)$ extracted from the membership function (see Figs. 2 and 3). The dynamic of our system starts in an FPN by checking the firing conditions for each transition that has active places if the degree of truth is more than or equal to the threshold value of the transition. The evolution of the marking provides a passage from one step to the next, as described by Chen [5], as well as a clear understanding of the steps and actions of our system. We will represent a rule defined by FPR (4) to make our model more understandable. The rule is presented in the form of four input proposals, "DNI, T, H, WS", each of which is defined by a fuzzy interval "DNI = clear, T = cold, H =

Fig. 2 Graphical presentation of the rules base FPR in a FPN
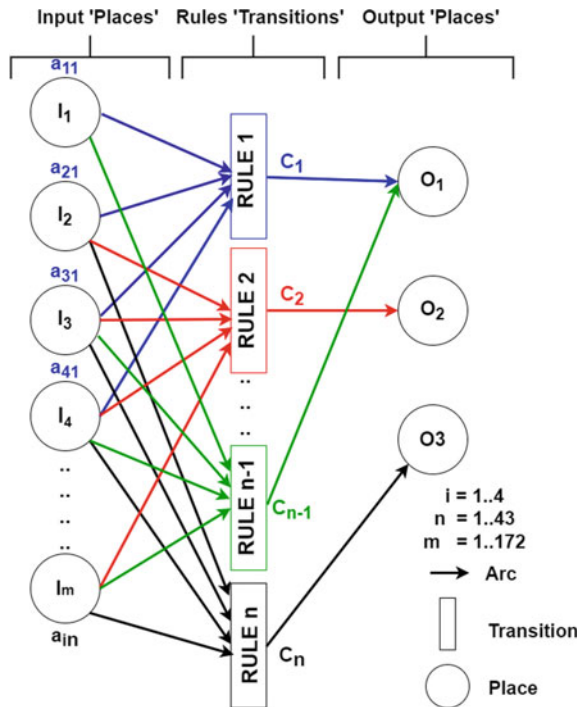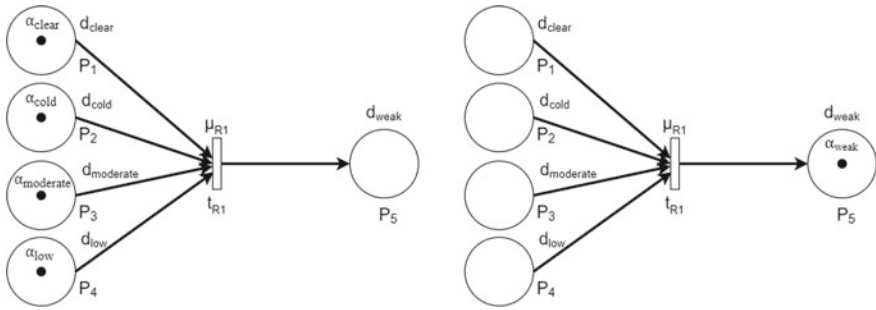
Fig. 3 An example of firing a marked fuzzy Petri net. (left) Before firing transition t_{R1}. (right) After firing transition t_{R1}

moderate, WS = low". The rules describing our system have four inputs and a single output representing reflectivity. Therefore, rule (4) has "R = weak" as an output with a certainty factor of $\mu_{R_1} = 0.78$ and a threshold of $\lambda_{R_1} = 0.31$.

$$R_1: \text{IF clear AND cold AND moderate AND low THEN weak}$$
$$(\mu_1 = 0.78, \lambda_1 = 0.31) \tag{4}$$

After representing rule (4), we can model it using FPN (see Fig. 3) and fuzzify inputs (see Fig. 1), which represent a degree of a value of a fuzzy interval in an FPR and a degree of truth of a token or mark in an FPN specified by value. For example, we have the degrees of the truth of each fuzzy interval quoted in rule (4) for the following entries: "DNI = 320 kWh/m^2, $T = 16\,°C$, $H = 48\%$, WS = 12 km/h", for each entry we have the following degrees of truth: "$\alpha_{\text{clear}} = 0.2$, $\alpha_{\text{cold}} = 0.75$, $\alpha_{\text{moderate}} = 0.9$, $\alpha_{\text{low}} = 0.86$" after the preparation of markings by tokens specified by the degree of truth of the place, we can pass to represent FPN (see Fig. 3). The four degrees of the truth of input places is greater than or equal to the threshold "α_{clear} and α_{cold} and α_{moderate} and α_{low} and $\geq \lambda_{R1}$".

An evolution of marking makes the verification of the conditions very rapid and requires respecting the temporal restrictions of our system. Because our model is a real-time system, failing to follow deadlines results in a calculation failure and an incomprehensible display, which can crash the display dynamically. The output of rule (4) can be determined by $\text{Min}(\alpha_{\text{clear}}, \alpha_{\text{cold}}, \alpha_{\text{moderate}}, \alpha_{\text{low}}) * \mu_{R1}$ after the transition validation requirements have been verified $\alpha_{\text{weak}} = 0.156$, and the marking can be altered with the degree of truth of the output place (see Fig. 3). After determining each degree of truth for each output, the center of gravity is used to get the true value of reflectivity. According to rule (4) and two other rules not specified in our example. We have three degrees of truth "$\alpha_{\text{weak}} = 0.156, \alpha_{\text{medium}} = 0.80, \alpha_{\text{strong}} = 0$" for calculating the real output of our model by the center of gravity, which equals $R_{R1} = 88.61\%$.
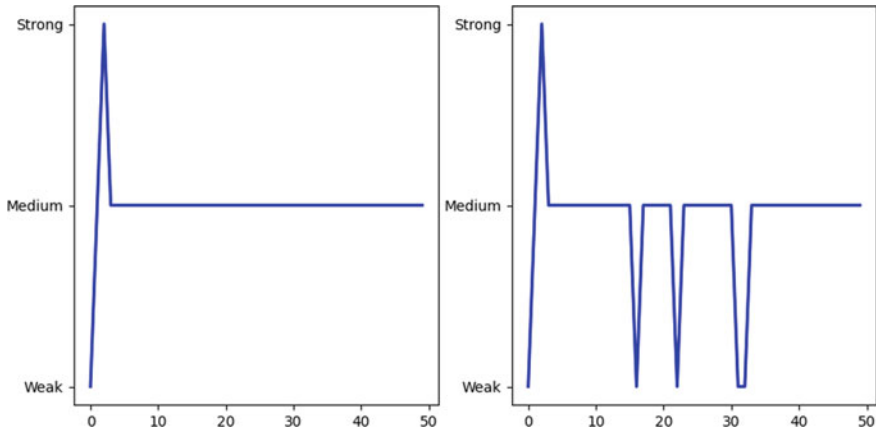
Fig. 4 Fuzzified dynamic display of reflectivity. (left) Real data. (right) Predicted data

2.3 Model Results

After defining our system and modeling it, we validate our model by using real data from a solar power plant located in a dusty area. According to our data, the prediction system error rate is lower than FL and closer to NN, implying that the predicted data approaches the true data. If we have many rules with higher CF and threshold values with membership functions set precisely with a minimum error rate, in that case, the system can give exact results with a minimal error rate, similar to NN but in real time. The outcome of our system is a real-time fuzzified dynamic display (see Fig. 4) of a prediction duration determined by the human expert (daily, per hour, or every 10 min), with values stored in a real-time database and employed in another FPN.

3   Conclusion

Through this work, we described the various modeling elements used to build a reflectivity prediction system using FPN. We created a simple and efficient method for predicting reflectivity, which is less complex than NN yet produces similar results. After preparing the rules by FPR and using an expert system, FPN depicted these rules in the form of a graphic network, each defined by a structure that can describe our system.

The model validity was proved by using real data from a solar power plant, from which rules were collected, and modelization of the prediction system concepts was used. Our system uses general FPN to respect the execution delay in the model output temporal state. Thus, we do not need to employ a temporal controller or another type

of FPN. We can conclude that the system's performance for data prediction is efficient and can be enhanced if we have an extensive rule base.

In future work, we will construct a system that uses the reflectivity predicted by this model to make real-time decisions about whether or not to clean solar collectors.

References

1. Raed IH, Syed IA (2011) Confidence value prediction of DNA sequencing with petri net model. J King Saud Univ Comput Inf Sci 23:79–89
2. Hansen P, Franco P, Kim S, Hirsch T (2018) Soccer ball recognition and distance prediction using fuzzy petri nets. In: IEEE international conference on information reuse and integration for data science, pp 315–322
3. Terhag F, Wolfertstetter F, Wilbert S, Hirsch T, Schaudt O (2019) Optimization of cleaning strategies based on ANN algorithms assessing the benefit of soiling rate forecasts. In: International conference on concentrating solar power and chemical energy systems, Morocco, pp 22-1–22-10
4. Zadeh L (1965) Fuzzy sets. Inf Control 8:338–353
5. Chen S, Ke JS, Chang J (1990) Knowledge representation using fuzzy Petri nets. IEEE Trans Knowl Data Eng 2:311–319
6. Looney CG (1988) Fuzzy petri nets for rule-based decision making. IEEE Trans Syst Man Cybern 18:178–183
7. Scarpelli H, Gomide F (1994) A high level net approach for discovering potential inconsistencies in fuzzy knowledge bases. Fuzzy Sets Syst 64:175–193

Experimental Validation of a Photovoltaic/Electrolysis System Dedicated to Supplying an Alternating Load and Producing Hydrogen

S. Yahyaoui, A. Aziz, A. Messaoudi, J. Blaacha, S. Dahbi, and I. Messaoudi

Abstract In this paper, we present an experimental validation of a photovoltaic/electrolysis system dedicated to supplying an alternating load and producing hydrogen. The system uses new way to produce hydrogen by adapting the electrolysis to a renewable source of energy such as photovoltaic (PV) by electrolysis of water and powering an AC load. Complete modeling and the control of this system are based on the use of several control techniques, and the main components of the proposed system are two converters, one DC/DC buck controlled by an MPPT control, the other DC/DC boost ensures the voltage control. The conversion from AC to DC is done by a modified single-phase full bridge**.** All these controls are digital and implemented in a microcontroller that allows us to minimize wiring and get accurate results. The experimental results show that the proposed system PV-electrolysis equipped with the different control algorithms can supply the energy required by the AC load and of storing the excess energy in the form of hydrogen production.

Keywords Photovoltaic · Hydrogen · Electrolysis · MPPT · Alternative load

1 Introduction

The evolution of technologies has enforced results for clean and optimal product of energy from renewable sources [1]. The main objective of our study is to solve the major drawback of renewable energies which is the problem of intermittency of the energy produced (in excess or in deficit). In the case of deficit, to meet the energy needs, it is necessary to use storage means during sunny periods and to reuse them in the absence of the sun. However, energy storage must be environmentally

S. Yahyaoui (✉) · A. Aziz · J. Blaacha · S. Dahbi
Electrical Engineering and Maintenance Laboratory, Higher School of Technology, Mohamed First University, BP 473, 60000 Oujda, Morocco
e-mail: saloua.yahyaoui@ump.ac.ma

A. Messaoudi · I. Messaoudi
Energy, Embedded System, and Data Processing Laboratory, National School of Applied Sciences, Mohamed First University, BP 473, 60000 Oujda, Morocco

friendly, which means including clean storage such as hydrogen production. In the case of solar energy excess, the most obvious means of storage is the injection of photovoltaic (PV) energy into the electrical grid. However, this means requires a radical rationalization of energy production. We have opted for clean energy storage of photovoltaic through the electrolysis of water to obtain hydrogen gas [2], which is also considered as an energy carrier.

In this context, we propose a reel photovoltaic/electrolysis system that responds:

- To provide, in terms of power, the energy needs to supply an alternative load.
- To switch any excess energy to an electrolysis that produces gaseous hydrogen.

2 Photovoltaic/Electrolysis System

2.1 System Description

The proposed system is shown in the block diagram in Fig. 1. It consists of:

- A photovoltaic panel of 180 W peak power.
- Two DC/DC converters in cascade (buck and boost), the first one controlled by a MPPT control, the second one is controlled by a PI control stabilizing the voltage at the value of a given reference voltage.
- An inverter performing the DC/AC conversion controlled by a bipolar PWM digital control.
- An electrolysis mounted in derivation on the photovoltaic conversion chain for water electrolysis and hydrogen production [3].
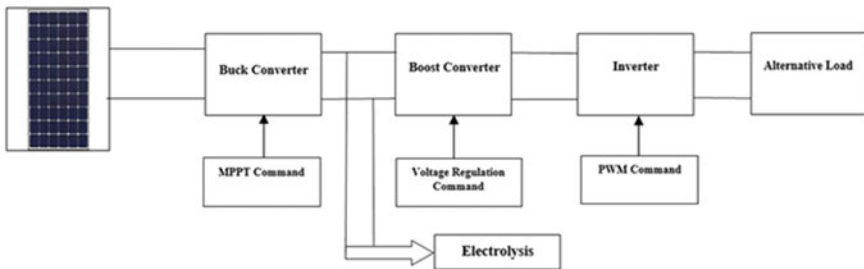
Fig. 1 Block diagram of the proposed system

2.2 Optimization of the PV System

DC/DC Converter Buck

For optimal operation of the photovoltaic panel [3], we have interposed a buck DC/DC converter stage (Fig. 2) whose components are dimensioned in previous work [4]. This stage is designed to be driven by a switching frequency signal of 100 kHz to minimize the inductance footprint.

This fixed frequency and variable duty cycle (PWM) control signal is delivered by a microcontroller that guarantees the tracking of the maximum power point (MPPT) of the PV panel [5]. The algorithm implemented, to realize this tracking, is perturb and observe (P&O) type [6, 7].

DC/DC Converter Boost

A second DC/DC converter of the boost type Fig. 3 is inserted, in the chain, in cascade with the buck to regulate the voltage of the DC bus that feeds the downstream part (inverter) at a constant value. The control of the output voltage of this converter is carried out by means of a digital proportional integral (PI)-type control delivered by the microcontroller [8, 9].

Electrolyzer

To really optimize the energy produced by the PV panel, we have introduced electrolysis in derivation.
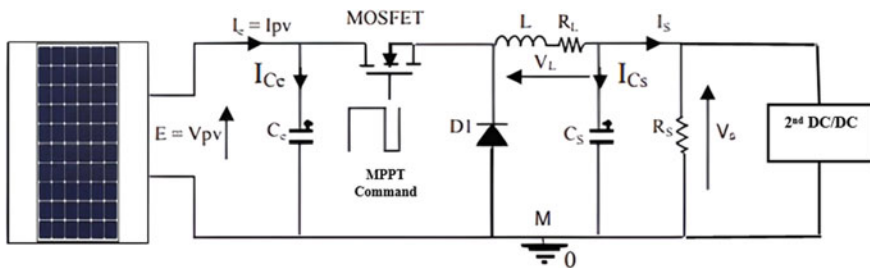
Fig. 2 First buck converter controlled with the MPPT control

Fig. 3 Second boost converter controlled with the PI command
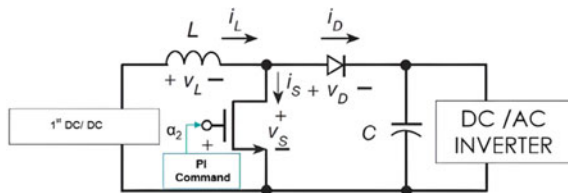
Fig. 4 Fundamental principle of electrolysis
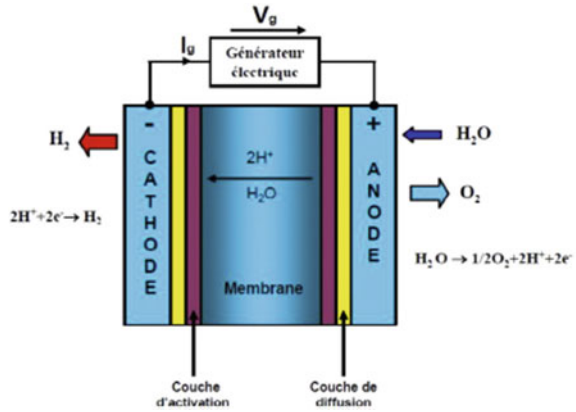
Electrolysis Process: Chemical Phenomena

The diagram of an electrolyzer is shown in Fig. 4 [10]. Its basic elements are two anode–cathode electrodes (chosen in our case cylindrical and concentric), a proton exchange membrane (PEM) and the electrolyte which is water in our case [11–13].

Electrical Model of the Electrolyzer

When the current flows through the solution of the PEM cell, the voltage across the electrolytic cell can be represented by the sum of: the Nernst voltage E_{nernst} [14], the activation overvoltage at the cathode η_c and the anode η_a, the overvoltage due to the membrane η_m and the interfacial overvoltage η_I.

$$E = E_{\text{nernst}} + \eta_c + \eta_a + \eta_m + \eta_I \tag{1}$$

Quantity of Hydrogen Produced by the Electrolysis of Water

According to the second Faraday law, number of moles of hydrogen produced at the cathode is proportional to the amount of electric current passing through the electrolyte:

$$n_{H_2} = \frac{MI_{\text{elec}}tn_c}{m_{H_2}.F} \tag{2}$$

m_{H2} is the mass of hydrogen formed to the electrode (in kg), n_c: number of cells, M: molar mass of hydrogen (in kg. mole^{-1}), I_{elec}: current through the electrolysis (in A), t: time of electrolysis (sec) and F: Faraday's number ($F = 96{,}485$ C/mol).

Synthesis of the Sinusoidal Waveform: Inverter

The system used for the DC/AC conversion is an inverter. Among the topologies described in the literature, we have implemented the full-bridge topology (*H* structure) [8]**.** We opted to work with electronic switches of the MOSFET type.

To obtain a pulse width modulated signal at the output of the inverter, we have controlled the four switches by a sequence of commands that avoid the simultaneous conduction of two switches of the same arm. We have developed and implemented in the microcontroller a unipolar PWM control that generates this command sequence. The generation of this command sequence is based on the comparison of a sinusoidal reference signal of frequency 50 Hz (the desired signal at the output) and a high-frequency triangular carrier.

To obtain a sinusoidal signal to feed the AC load, we have filtered the PWM signal by a filter of type LC dimensioned and intended for this purpose.

3   Results and Discussions

3.1   *Operating Mode*

To verify the performance of the whole PV/electrolysis/load system discussed in the paragraph 2.1, the tests were carried out on a typical day in Oujda characterized by unstable weather to test, under real weather conditions, the performance of the system. The PV module used for the tests is the PVUD180MF5 that produce 180 W peak power. Critical parameters are read and processed by the microcontroller ATmega3228P-PU, then transported to the environment processing via the serial port at a rate of 115200 bps and finally to plot them in Excel.

Some results are read directly on a digital oscilloscope. The results obtained are grouped on the following figures.

3.2   *Discussion*

As it can be seen in Fig. 5a, the maximized power delivered by the PV panel during this recording varies between 40 and 90 W depending on the variable power of the incident sunlight. This power is tracked by the MPPT control algorithm which controls the first DC/DC converter (buck).

The output of the buck converter is connected to an electrolyzer producing hydrogen from water while consuming an average power value is almost 60 W as shown in Fig. 6a.

In parallel to the electrolyzer, a second DC/DC boost converter is connected providing voltage stabilization on a given reference voltage as seen in Fig. 5b (17 V in this case).
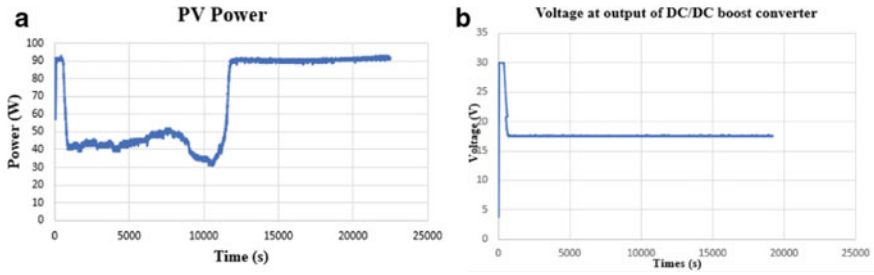
Fig. 5 **a** Power produced by PV panels, **b** controlled voltage at the output of the boost converter
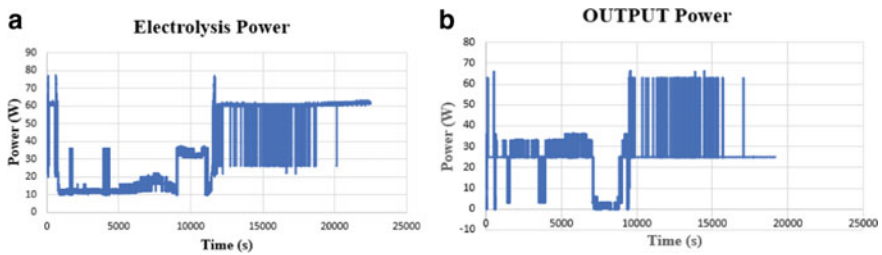
Fig. 6 **a** Power consumed by the electrolyzer, **b** power consumed by the inverter

The controlled DC voltage is then transformed into AC voltage by a single-phase inverter; the power consumed is almost 25 W as shown in Fig. 6b. The total power consumed by the electrolyzer and the AC load at the output of the inverter reaches on average the 40 W resp 90 W produced by the PV panel under the variable irradiation. Some Watts disappeared due to the conversion and switching losses of the DC/DC converters as well as the DC/AC inverter (such as unwanted harmonic losses and MOSFETs switching losses). The power loss will be detailed in future work.

After filtering, we obtained the clean sinusoidal signal as shown in Fig. 7.

Fig.7 Output signal after filtering the whole system
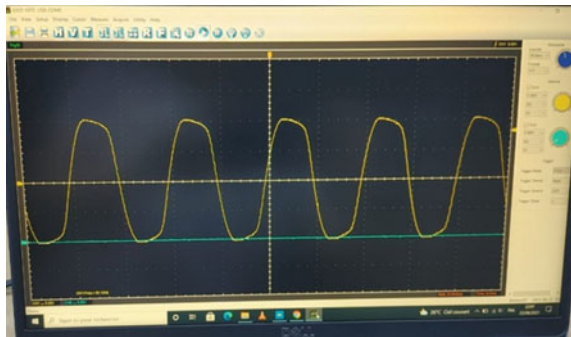
4 Conclusion

The presented system maximizes the power produced by the PV panels, produces clean hydrogen, regulates the voltage and converts DC to AC. The MPPT control follows a P&O algorithm and is implemented by controlling a buck DC/DC converter. The voltage regulation control uses a simple comparison algorithm with a given reference voltage. After power maximization and voltage control, the DC current is transformed into AC current using a single-phase inverter controlled by a bipolar PWM control. The power delivered by the panel exposed to variable radiation is, under all conditions, divided between the electrolyzer and the AC load.

All the above-mentioned algorithms are implemented digitally using hardware containing microprocessors, MOSFET drivers and galvanic isolations. The presented results meet the main goal of the system, which is to maximize the use of the generated energy without using batteries.

In the future work, we are interested in the synchronization of the system parameters (phase, frequency) for the coupling with the electrical grid and the use of the produced hydrogen in new applications.

References

1. Sun Y, Gao J, Wang J, Huang Z, Li G, Zhou M (2021) Evaluating the reliability of distributed photovoltaic energy system and storage against household blackout. Global Energy Interconnection 4(1):18–27. ISSN: 2096-5117
2. Hatti M, Meharrar A, Tioursi M (2011) Power management strategy in the alternative energy photovoltaic/PEM fuel cell hybrid system. Renew Sustain Energy Rev 5104–5110
3. Dahbi S, Mazozi I, Benabdellah N, Benazzi N, Abdelhak A (2017) Efficient power management strategy of Photovoltaic/Load/Electrolysis/Grid hybrid system. In: 2017 8th International renewable energy congress (IREC). Amman, Jordan
4. Nabil M, Allam SM, Rashad EM (2013) Performance improvement of a photovoltaic pumping system using a synchronous reluctance motor. Electr Power Compon Syst 41(4):447–464
5. Aziz A (2006) Propriétés Electriques des Composants Electroniques Minéraux et Organiques, Conception et Modélisation d'une Chaîne Photovoltaïque pour une Meilleure Exploitation de l'Energie Solaire. Thesis, Toulouse III-Paul Sabatier University
6. Dahbi S, Aziz A, Benazzi N, Elhafyani M (2016) New method to improving hydrogen production by an adaptive photovoltaic system and P&O algorithm. J Theor Appl Inf Technol 84:42e7
7. Abdelsalam AK, Massoud AM, Ahmed S, Enjeti PN (2011) High performance adaptive perurb and observe MPPT technique for photovoltaic-based microgrids. IEEE Trans Power Electron 26:1010–1021
8. Dahbi S, Aboutni R, Aziz A, Benazzi N, Elhafyani M, Kassmi K (2016) Optimised hydrogen production by a photovoltaic electrolysis system DC/DC converter and water flow Controller. Int J Hydrogen Energy 41:20858e66
9. Fereidooni M, Mostafaeipour A, Kalantar V, Goudarzi H (2018) A comprehensive evaluation of hydrogen production from photovoltaic power station. Renew Sustain Energy Rev Part 1 82:415–423. ISSN: 1364-0321
10. Swette LL, LaConti AB, McCatty SA (1994) Proton-exchange membrane regenerative fuel cells. J Power Sources 47:343e51

11. Loroi T, Yasuda K, Siroma Z, Fujiwara N, Miyazaki Y (2002) Thin film electrocatalyst layer for unitized regenerative polymer electrolyte fuel cells. J Power Sources 112:583e7

12. Selamet OF, Becerikli F, Mat MD, Kaplan Y (2011) Development and testing of a highly efficient proton exchange membrane (PEM) electrolyzer stack. Int J Hydrogen Energy 36:11480e7

13. Dahbi S, Aziz A, Messaoudi A, Mazozi I, Kassmi K, Benazzi N (2018) Management of excess energy in a photovoltaic/grid system by production of clean hydrogen. Int J Hydrogen Energy 43:5283–5299

14. Dahbi S , Mazozi I, El Ouariachi M, Messaoudi A, Aziz A (2017) Implementation of a multi-control architecture in a Photovoltaic/Grid/Electrolysis system for usual use and clean storage by hydrogen production. Int J Renew Energy Res 7(4)

Numerical Modeling of a Two-Dimensional Multiphase Flow Through a Porous Dam-Break

Abderrahmane Kaouachi, Salah Daoudi, and Imad Elmahi

Abstract In this work, we conducted a study that allowed us to examine the multiphase flow problem through a porous medium. We used the interFoam solver already implemented on the OpenFoam software. This solver is based on the modified Volume Of Fluid (VOF) model. It consists in solving numerically the two-dimensional incompressible Navier-Stokes equations coupled with an equation describing the fraction volume, the system is solved by using the finite volume method. Afterward, we realized test cases treating a dam break in a porous medium. All of the numerical results were compared with those obtained from some experimental works.

Keywords Finite volume method · InterFoam · Multiphase flow · Porous medium · Volume of fluid model

1 Introduction

Hirt and Nichols [1] introduced the Volume of Fluid (VOF) Model, which initiated a new era in multiphase flow modeling. It is based on the specification of an *indicator* function. This function informs us if a cell is filled with one kind of fluid or a mixture of the two.

The idea is to introduce a function α that has a value of 1 at any position occupied by the fluid and 0 elsewhere. The average value of α in a cell would thus indicate the proportion of the cell's volume occupied by fluid. In this case, a unit value of α would mean that the cell was full of fluid, while a zero value would mean that the cell had no fluid. A free surface must be present in cells with α values between 0 and 1.

A. Kaouachi (✉) · S. Daoudi
Laboratoire de Mécanique et Énergétique, FSO, Université Mohammed I, Oujda, Morocco
e-mail: a.kaouachi@ump.ac.ma

I. Elmahi
Laboratoire de Modélisation et calcul scientifique, ENSAO, Université Mohammed I, Oujda, Morocco

There are different methods that may be used to apply this notion, but it is worthwhile to examine the specific approach taken by OpenFOAM in this regard. Ubbink [2] and Rusche [3] provide a detailed description of the methodology, whereas Berberovic et al. [4] provide a brief and up-to-date explanation of the methodology.

2 Governing Equations

In the VOF model [1], the interFoam [5] solver solves the transport equation of the function α which represents the volume fraction of a single phase is coupled with the Navier-Stokes equation for two incompressible, isothermal immiscible fluids which is composed of the continuity equation and the momentum equation.

$$\nabla \cdot \mathbf{U} = 0 \tag{1}$$

$$\frac{\partial \rho \mathbf{U}}{\partial t} + \nabla \cdot (\rho \mathbf{U} \mathbf{U}) = -\nabla p_d - \mathbf{g} \cdot \mathbf{x} \nabla \rho + \nabla \cdot (2 \mu \mathbf{S}) + \sigma \kappa \nabla \alpha + \mathbf{f}_d \tag{2}$$

$$\frac{\partial \alpha}{\partial t} + \nabla \cdot (\alpha \mathbf{U}) = 0 \tag{3}$$

where $\mathbf{U}$ denotes the combined velocity field of the two fluids across the flow domain. ρ is the fluid density and p_d is the pressure minus the hydrostatic potential,

$$p_d = p - \rho \mathbf{g} \cdot \mathbf{x} \tag{4}$$

$\mathbf{g}$ is the gravitational acceleration, $\mu = \rho \nu$, is the dynamic viscosity and ν denotes the kinematic viscosity, and $\mathbf{S}$ is the mean strain rate tensor given by,

$$\mathbf{S} = 0.5[\nabla \mathbf{U} + (\nabla \mathbf{U})^{\mathrm{T}}] \tag{5}$$

The second to last term in Eq. (2) accounts the effect of surface tension, σ, where κ is the curvature of the free surface determined from $\kappa = -\nabla \left(\frac{\nabla \alpha}{|\nabla \alpha|} \right)$. α denotes the volume fraction of one of the phases (water or air) is used to characterize the medium, which takes value 0 in air and 1 in water. It can be defined in terms of the density as:

$$\alpha = \frac{\rho - \rho_{\mathrm{a}}}{\rho_{\mathrm{w}} - \rho_{\mathrm{a}}} \tag{6}$$

We suppose that any intrinsic fluid property, Φ, can be written in terms of α as follows:

$$\Phi = \Phi_{\mathrm{w}} \alpha + \Phi_{\mathrm{a}} (1 - \alpha) \tag{7}$$

Two immiscible fluids are considered as one effective fluid throughout the domain, the physical properties of which are calculated as weighted averages based on the distribution of the water volume fraction, thus being equal to the properties of each fluid in their corresponding occupied regions and varying only across the interface,

$$\rho = \rho_w \alpha + \rho_a (1 - \alpha) \tag{8}$$

$$\mu = \mu_w \alpha + \mu_a (1 - \alpha) \tag{9}$$

where ρ_w and ρ_a are densities of water and air, respectively.

$\mathbf{f}_d$ is the drag force which characterize the porous media, it can be expressed in the general Darcy-Forchheimer [6] case by:

$$\mathbf{f}_d = - \left[\mu D + \frac{1}{2} \rho |\mathbf{U}| F \right] \mathbf{U} \tag{10}$$

where D and F are the Darcy and Forchheimer coefficients, respectively, calculated by the Ergun model [7] as:

$$D = \frac{150(1 - \varepsilon)^2}{d_p^2 \, \varepsilon^3} \quad \text{and} \quad F = \frac{1,75(1 - \varepsilon)}{d_p \varepsilon^3} \tag{11}$$

ε is the porosity and d_p mean the average grain size. InterFoam solves the numerical difficulty of preserving a crisp interface by compressing it numerically and limiting phase fluxes using the "Multidimensional Universal Limiter with Explicit Solution" (MULES) limiter [8–10]. Numerical interface compression is achieved by adding a completely heuristic element to Eq. (3),

$$\frac{\partial \alpha}{\partial t} + \nabla \cdot (\mathbf{U} \alpha) + \nabla \cdot [\mathbf{U}_r \alpha (1 - \alpha)] = 0 \tag{12}$$

where $\mathbf{U}_r = \mathbf{U}_w - \mathbf{U}_a$ is the vector of relative velocity, designated as the "compression velocity" that represents an artificial force that compress the area under consideration.

Equation (12) contains an additional convective term called the "compression" term, which allows the compression of the surface to a cleaner surface. This term contributes significantly to a higher interface resolution. It is present only in the interface region and disappears at both boundaries of the phase fracture. It does not affect the solution outside this region. In addition to correctly interpreting the multiphase flow phenomenon. The main objective of this formulation lies in the advantage of capturing the interface region much more accurately than the classical VOF approach. It is unavoidable to escape the numerical diffusion which occurs by the discretization of the convection terms, but they can be controlled and minimized by the discretization of the compression term. This allows us to visualize the results more accurately at the interface.

3 Numerical Method

Discretization of the compression term in Eq. (12) the relative velocity at cell faces is calculated from the gradient of phase fraction as follows:

$$U_{r,f} = n_f \min \left[C_\gamma \frac{|\phi|}{|\mathbf{S}_f|}, \max \left(\frac{|\phi|}{|\mathbf{S}_f|} \right) \right] \tag{13}$$

where C_γ is a value defined by the user which determines the strength of the compression, ϕ is the face volume flux, and n_f is the face unit normal flux and $\mathbf{S}_f$ is the face area vector normal to the face pointing out of the cell. In [11], the reader can find out more about the numerical implementation works.

To keep the solution process stable for the adaptative time step control, we define the Courant number as,

$$\text{Co} = \frac{\mathbf{U}_f \cdot \mathbf{S}_f}{\mathbf{d} \cdot \mathbf{S}_f} \Delta t \tag{14}$$

where $\mathbf{d}$ is a vector between calculations points of control volumes sharing the face and Δt is the time step.

Discretization of the momentum equation The PISO algorithm [12] is used, it allows the coupling of velocity and pressure correctly. It is therefore convenient to adjust the PISO algorithm to the momentum equation for the interphase.

A discretization of the momentum equation gives:

$$A_p^{\mathbf{U}} \mathbf{U}_p = \mathbf{H}(\mathbf{U}) - \nabla p_d - \mathbf{g} \cdot \mathbf{x} \nabla \rho + \sigma \kappa \nabla \gamma \tag{15}$$

separating velocity at cell centers:

$$\mathbf{U}_p = [A_p^{\mathbf{U}}]^{-1} \left\{ [\mathbf{H}(\mathbf{U}) - \mathbf{g} \cdot \mathbf{x} \nabla \rho + \sigma \kappa \nabla \gamma] - \nabla p_d] \right\} \tag{16}$$

we replace this expression in the continuity equation, we obtain a poisson equation for pressure p_d

$$\nabla \cdot \left\{ [A_p^{\mathbf{U}}]^{-1} \nabla p_d \right\} = \nabla \cdot \left\{ [A_p^{\mathbf{U}}]^{-1} [\mathbf{H}(\mathbf{U}) - \mathbf{g} \cdot \mathbf{x} \nabla \rho + \sigma \kappa \nabla \gamma] \right\} \tag{17}$$

At last we find the flows corresponding to the continuity equation, we get there by:

$$\phi = -(A_p^{\mathbf{U}})^{-1} \mathbf{S}_f \cdot \nabla p_d + (A_p^{\mathbf{U}})^{-1} \mathbf{S}_f \cdot [\mathbf{H}(\mathbf{U}) - \mathbf{g} \cdot \mathbf{x} \nabla \rho + \sigma \kappa \nabla \gamma] \tag{18}$$

4   Results and Discussion

In order to validate the model, a comparative study was conducted based on the experiments performed by Lin [13]. He tested a porous dam break in a glass aquarium of 89 cm in length, 44 cm in width and 58 cm in height (in 2D the study is assimilated to 89 cm × 58 cm), in the middle is a porous block of 29 cm in length and 58 cm in height, consisting of 2 porous materials such as crushed stone and glass beads of properties see Table 1. A base level of water of 2.5 cm was fixed on the entire bottom of the tank, outside the reservoir. Water was isolated in a tank between one side of the reservoir and a gate at 2 cm from the porous dam break in the initial state. The gate was opened and the water flowed freely colliding with the porous materials and percolating into the dry zone. The flow of the porous dam break was recorded by a camera at different moments, more details are founded in [13] (Fig. 1).

The numerical simulations presented show very similar behavior. Normally, water movement inside the porous block is not very well predicted. These divergences are probably due to the water being released differently. In the numerical results, the water is initially at rest and an instantaneous opening of the gate is simulated. In the experimental results, however, a manually opened gate is moved upwards with the estimated time of 0.1 s [13]. The water is accelerated to the bottom, which results

Table 1 Porous materials properties

Material	ε	d_p [cm]
Crushed stone	0.49	1.59
Glass beads	0.39	0.3
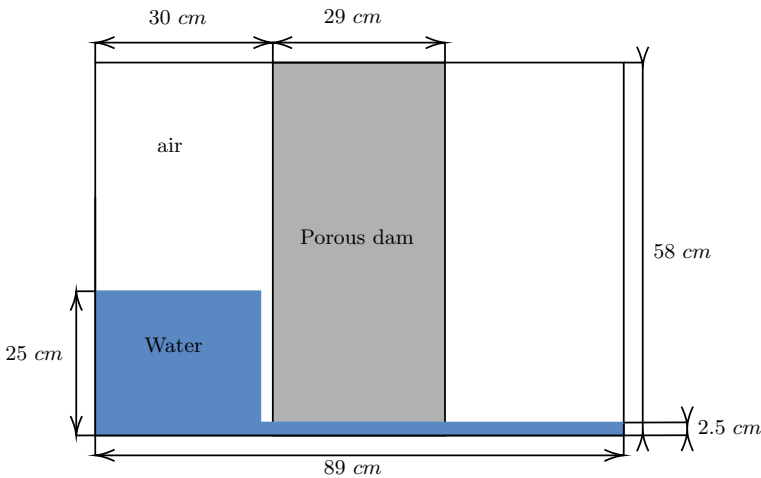
Fig. 1 Initial stage for the porous dam break flow
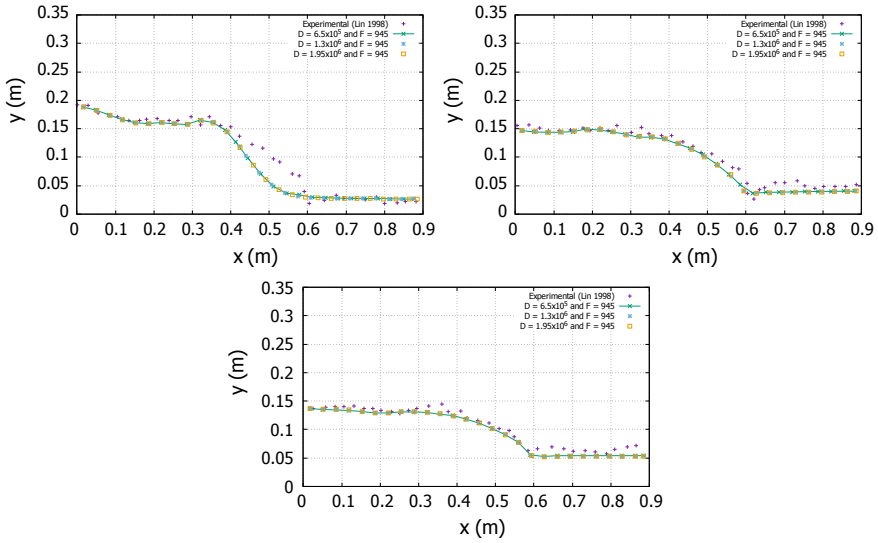
Fig. 2 Results obtained for the phase fraction α at: 0.8 s, 1.6 s and 2.2 s, respectively, using values: $D = (6.5 \times 10^5; 1.3 \times 10^6; 1.95 \times 10^6)$ for the Darcy coefficient and $F = (945)$ for the Forchheimer coefficient
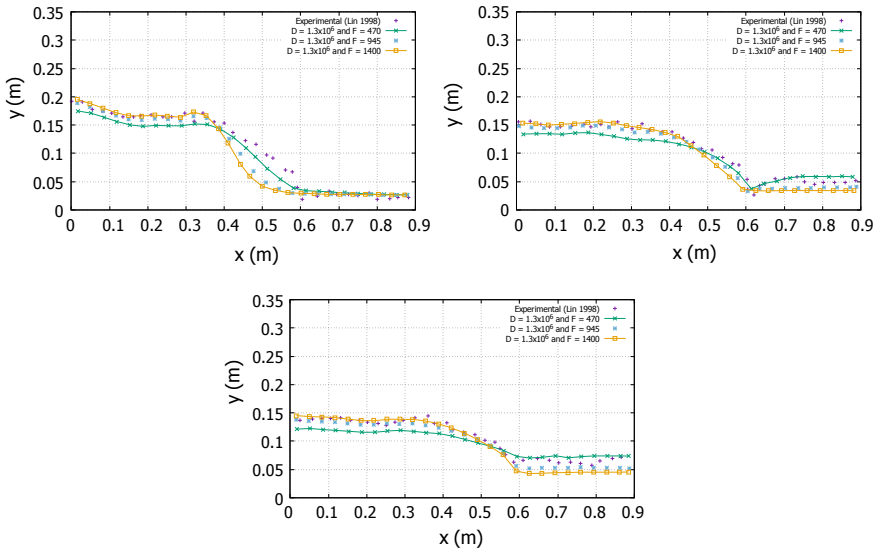
Fig. 3 Results obtained for the phase fraction α at: 0.8 s, 1.6 s and 2.2 s, respectively, using values: $D = (1.3 \times 10^6)$ for the Darcy coefficient and $F = (470; 945; 1400)$ for the Forchheimer coefficient

in a different behavior at the first stage of the simulation. The water starts to flow from the bottom during the opening of the gate, in the other instants (1.8 s and 2.2 s) there is no such deviation between the numerical results and the experiments. We can think that the first stages, influenced by the opening of the gate, affect more the configuration. It is also observed that the infiltration process is slower due to the characteristics of the glass beads, see [13]. This influence is also noticeable at later times when the water takes longer to percolate through the dam. For the higher instants, the model handles the free surface flow well. The water surface levels do not show significant differences either on the sides of the porous dam or inside the porous block. Concerning the Darcy and Forchheimer coefficients that we varied, we observe that when we vary the Darcy coefficient in Fig. 2, the flow is almost the same for all variations. However, when we vary the Forchheimer coefficient in Fig. 3, the flow is more influenced especially when the water passes through the porous block, the green curve (fixed values: $D = 1.3 \times 10^6$ and $F = 470$) shows a behavior very close to the experiment in the instants 0.8 s and 1.6 s but at the instants 2.2 s, we notice that the yellow curve (fixed values: $D = 1.3 \times 10^6$ and $F = 1400$) is better at the porous dam.

5   Conclusion

The VOF problem is solved using the InterFoam solver based on the finite volume method. We treated the flows through the porous media using the solver and compared the results with those of the laboratory. The numerical results show a good agreement with those of the experiments, so we can conclude that this solver performs well and can solve this type of problem. The Ergun model [7] may be used to estimate D and F coefficients for a given porosity and nominal diameter. The numerical results obtained demonstrate that motion behaves very similarly to the experiments data realized by Lin [13] A novel approach, based on guessing D and F coefficients, is required.

References

1. Hirt CW, Nichols BD (1981) Volume of fluid (VOF) method for the dynamics of free boundaries. J Comput Phys 39(1):201–225. https://doi.org/10.1016/0021-9991(81)90145-5
2. Ubbink O (1997) Numerical prediction of two fluid systems with sharp interfaces. Ph.D. thesis, Imperial College of Science, Technology and Medicine, London
3. Rusche H (2002) Computational fluid dynamics of dispersed two-phase flows at high phase fractions. Ph.D. thesis, Imperial College of Science, Technology and Medicine, London
4. Berberovic E, Van Hinsberg NP, Jakirlic S, Roisman IV, Tropea C (2009) Drop impact onto a liquid layer of finite thickness: dynamics of the cavity evolution. Phys Rev E 79. https://doi.org/10.1103/PhysRevE.79.036306

5. Larsen BE, Fuhrman DR, Roenby JR (2019) Performance of interFoam on the simulation of progressive waves. Coast Eng J 61(3):380–400. https://doi.org/10.1080/21664250.2019.1609713

6. Forchheimer P (1901) Wasserbewegung durch Bodem. Z Ver Dtsch Ing 45:1782–1788

7. Ergün S (1952) Fluid flow through packed columns, chemical engineering progress, vol 48

8. OpenCFD Ltd. (2009) OpenFOAM, The open source CFD toolbox, user guide

9. OpenCFD (2005) Technical report no. TR/HGW/02

10. Van Leer B (1974) Towards the ultimate conservative difference scheme II. Monotonicity and conservation combined in a second order scheme. J Comput Phys 14:361–370

11. Deshpande SS, Anumolu L, Trujillo MF (2012) Evaluating the performance of the two-phase flow solver interFoam. Comput Sci Discov 5:014016 (36p). https://doi.org/10.1088/1749-4699/5/1/014016

12. Issa RI (1986) Solution of the implicitly discretised fluid flow equations by operator splitting. J Comput Phys 62:40–65

13. Lin P (1998) Numerical modeling of breaking waves. Cornell University

High Order Scheme for Numerical Simulation of an Oblique Shock Overbreak a Ramp

Youssef Es-Sabry, Hind Talbi, Elmiloud Chaabelasri, and Najim Salhi

Abstract This work deals with the numerical simulation of steady flow involving an oblique shock over a ramp. The numerical approach is based on solving the Euler equations on an unstructured triangular grid. The conservation equations are applied to each volume by considering the flows of variables through each face of the control volume. The calculation of convective fluxes is based on Roe's flux difference splitting scheme. To improve the higher-order accuracy we adopted the Monotonic Upstream-centered Scheme for Conservation Laws (MUSCL) approach with the particular edge-based formulation. For time integration, a second-order Runge-Kutta with local time-stepping and the central implicit residual smoothing to accelerate the convergence to the steady-state are used. Numerical simulations are presented to demonstrate the efficiency and robustness of the method to deal with shock problem.

Keywords Euler equations · Finite volume method · Oblique shock · Roe solver · Runge-Kutta · Supersonic

1 Introduction

Currently, aerodynamic applications and many industrial problems are concerned with the compressible flows of a perfect fluid, which constitute a continuous fluid model and are modeled by the Euler equation system of a perfect fluid.

Y. Es-Sabry (✉) · H. Talbi · N. Salhi
LME, Faculty of Sciences, University Mohammed I, 60000 Oujda, Morocco
e-mail: essabryoussef@gmail.com

H. Talbi
e-mail: h.talbi@ump.ac.ma

E. Chaabelasri
LPTPME, Faculty of Sciences, University Mohammed I, 60000 Oujda, Morocco

The Euler Equations are a system of non-linear hyperbolic conservation laws that govern the dynamics of inviscid compressible flow model. Indeed, the analytical resolution of these equations seems impossible, but the numerical resolution is now quite common thanks to the development of new numerical methods, In effect, this requires special numerical method known as robust that is to say, having the capacity to calculate the flow without producing a non-state physical.

The cell-vertex finite volume method was applied during the present study. A class of methods, based essentially on the resolution of the Riemann problem, has been retained and concerns shock capture schemes. These schemes have been developed following an original method proposed by Godunov [1]. Where he proposes a scheme allowing the exact solution, at each face of a Riemann problem in one dimension. This scheme is, however, very expensive. To reduce the computational time Osher and Solomon [2] suggested to solve the Riemann problem in an approximate way. With the same purpose, Roe [3], by linearising the Riemann problem, developed the first FDS where the flows are evaluated using the exact solution of a linearised problem. The inconvenience of this scheme is that it may admit, as non-physical solutions, the expansion waves. To avoid this problem an entropy correction has been proposed by Harten's [4]. This correction has made Roe's scheme the most used scheme because it offers the advantage of better capturing the shocks and contact discontinuities and to be less expensive. In addition, in [5, 6] augmentation basic upwind dissipation, have been reviewed to enhance Roe scheme. Recently, Musa [7] have proposed a modified Roe solver that can employ high-order schemes with high stability based on splitting of upwind term.

Schemes using upstream differentiation are specified to order one in space. In order to improve the accuracy, several extensions have been proposed. Among them, we find the MUSCL technique (Monotonic Upstream-centered Scheme for Conservation Laws) was proposed by Van Leer [8], and the Total Variation Diminishing (TVD) schemes formulated by Harten's [9], and Chakravarthy and Osher [10]. When computing the supersonic flow, discontinuities will usually exist somewhere in the computational domain. In the vicinity of these discontinuities, the MUSCL reconstruction used to achieve the second-order accuracy will produce oscillations in the flow solution and eventually cause the computation to fail. In order to overcome this problem and extend higher-order schemes to unstructured meshes, Barth and Jespersen [11] have developed a new reconstruction method with a multidimensional limiter according to the maximum principle. The limiter can effectively suppress numerical oscillations with shocks. However, they adopted a non-differentiable limiting function, which adversely affected the convergence property of the flow solver. In analyzing this problem, Venkatakrishnan's [12] enhanced Barth's limiter by replacing the limiting function with a smooth, differentiable one. Nevertheless, Venkatakrishnan's limiter introduced a free parameter, the value of which impacts both accuracy and convergence.

In this study, our main objective is to investigate the efficacy and accuracy of Roe's approximate Riemann solver for high order accurate MUSCL reconstruction scheme as a state-of-the-art tool to study compressible inviscid flow with the presence of the shock wave.

2　Governing Equations

The considered governing equations in this study are the two-dimensional time-dependent Euler equations written in a conservative form as follows [13]

$$\frac{\partial W}{\partial t} + \frac{\partial E}{\partial x} + \frac{\partial F}{\partial y} = 0 \tag{1}$$

where t is the time, x and y are the Cartesian coordinates, W is the conservative variables, E and F represent the fluxes in the x and the y-directions, respectively given by:

$$W = \begin{bmatrix} \rho \\ \rho u \\ \rho v \\ \rho E \end{bmatrix}, E = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ \rho u v \\ \rho H u \end{bmatrix}, F = \begin{bmatrix} \rho v \\ \rho v u \\ \rho v^2 + p \\ \rho H v \end{bmatrix} \tag{2}$$

where ρ is the fluid density, u and v are the Cartesian velocity components, p is the static pressure and E is the total energy per unit volume.

$$E = \left(\frac{P}{(\gamma - 1)} + \frac{1}{2} (u^2 + v^2) \right) \tag{3}$$

and the total enthalpy is defined as:

$$H = E + \frac{P}{\rho} \tag{4}$$

The system is closed by the equation of state for ideal gases:

$$p = \rho R T \tag{5}$$

where T is the static temperature and R is the specific gas constant.
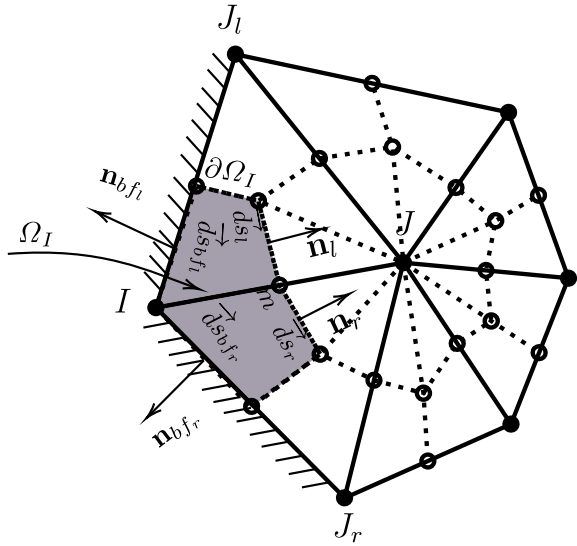
3   Numerical Methods

3.1   *Finite Volume Approach*

In the approach of cell-vertex Finite volume [14], the computational domain is discretized into a finite number of elements. In this work, the control volumes are formed that surround each vertex in the mesh. The flow variables are stored at each mesh vertex. Then Eq. 1 is applied over each control volume of the mesh. These control volumes are formed by connecting the center of each cell to the midpoint of the cell edges, as shown in Fig. 1.

The integrations of Eq. 1 over the control volume gives:

$$\Omega_i \frac{\partial W_i}{\partial t} = - \oint_{\partial \Omega_i} F_c . \mathrm{d}S \tag{6}$$

where Ω_i represents the volume of the control volume, W_i represents the cell averaged vector of dependent variables, $\partial \Omega_i$ represents the boundary of the control volume, $\mathrm{d}S$ represents the infinitesimal area of the boundary, and F_c is the flux projected along with the unit outward normal vector of boundary, n_{ij}.

$$F_c = [E, F].n_{ij} \tag{7}$$

The integral boundary presented in Eq. 6 is computed over the boundary of the control volume using an edge-based representation of the mesh as:

$$\oint_{\partial\Omega_i} F_c dS = \sum_{j=1}^{n_{\text{edges}}} (F_c)_{ij} . n_{ij} \Delta S_{ij} \tag{8}$$

where n_{edges} represent the number of edges connected to the vertex i, $(F_c)_{ij}$ is a numerical flux at the cell edge, n_{ij} is the outward unit normal vector is given by:

$$n_{ij} = n_l + n_r \tag{9}$$

and ΔS_{ij} is total face area is given by:

$$\Delta S_{ij} = \Delta S_l + \Delta S_r \tag{10}$$

Finally, for node i the semi-discrete form of Eq. 1 can be written as:

$$\frac{\partial W_i}{\partial t} = -\frac{1}{\Omega_i} \sum_{j=1}^{n_{\text{edges}}} (F_c)_{ij} . n_{ij} \Delta S_{ij} \tag{11}$$

3.2 Convective Flux Approximations

In this study, Roe's approximate Riemann solver [3] is employed to compute the convective flux. With this solver, the numerical flux at the cell edge is given by

$$(F_c)_{ij} = \frac{1}{2} [F_c(W_R) + F_c(W_L)] - \frac{1}{2} \sum_{k=1}^{4} \tilde{R}_k |\tilde{\lambda}_k| \tilde{\alpha}_k \tag{12}$$

where W_L and W_R are the right and the left state variables, respectively. $\tilde{R}_k$ is the kth column vector of the right eigenvector of the Roe-averaged matrix

$$\tilde{R}_k = \begin{pmatrix} 1 & 1 & 0 & 1 \\ \tilde{u} - \tilde{a} & \tilde{u} & 0 & \tilde{u} + \tilde{a} \\ \tilde{v} & \tilde{v} & 0 & \tilde{v} \\ \tilde{H} - \tilde{u}\tilde{a} & \frac{1}{2}(\tilde{u} + \tilde{v}) & \tilde{v} & \tilde{H} + \tilde{u}\tilde{a} \end{pmatrix} \tag{13}$$

with the Roe-averaged quantities are computed as follows:

$$\tilde{\rho} = \sqrt{\rho^L \rho^R}, \quad \tilde{u} = \frac{\sqrt{\rho^L} u^L + \sqrt{\rho^R} u^R}{\sqrt{\rho^L} + \sqrt{\rho^R}}, \quad \tilde{v} = \frac{\sqrt{\rho^L} v^L + \sqrt{\rho^R} v^R}{\sqrt{\rho^L} + \sqrt{\rho^R}}$$

$$\tilde{H} = \frac{\sqrt{\rho^L} H^L + \sqrt{\rho^R} H^R}{\sqrt{\rho^L} + \sqrt{\rho^R}}, \quad \tilde{a} = \sqrt{\left[(\gamma - 1) \left(\tilde{H} - \frac{1}{2} (\tilde{u}^2 + \tilde{v}^2) \right) \right]} \tag{14}$$

where $\tilde{a}$ represent the speed of sound.

The eigenvalue $\tilde{\lambda}_k$ of the Roe-averaged matrix are given as:

$$\tilde{\lambda}_1 = V_n - \tilde{a}, \ \tilde{\lambda}_2 = \tilde{\lambda}_3 = V_n \ \tilde{\lambda}_4 = V_n + \tilde{a} \tag{15}$$

with $V_n = \tilde{u}.n_x + \tilde{v}.ny$.

And the wavelength $\tilde{\alpha}_k$ are given as:

$$
\begin{aligned}
\tilde{\alpha}_1 &= ((p_R - p_L) - \tilde{\rho}\tilde{a}(V_n^R - V_n^L))/2\tilde{a}^2 \\
\tilde{\alpha}_2 &= ((\rho_R - \rho_L) - (p_R - p_L)/\tilde{a}^2 \\
\tilde{\alpha}_3 &= ((p_R - p_L) + \tilde{\rho}\tilde{a}(V_n^R - V_n^L))/2\tilde{a}^2 \\
\tilde{\alpha}_4 &= \tilde{\rho}(V_t^R - V_t^L)
\end{aligned}
\tag{16}
$$

with $V_n^k = \tilde{u}_x^k.n_x + \tilde{u}_y^k.n_y$ and $V_t^k = -\tilde{u}_x^k.n_y + \tilde{u}_y^k.n_x$, $(k = L, R)$.

The original Roe's scheme produces non-physical expansion shocks in the locations where the eigenvalue of the Roe-averaged matrix vanish (e.g., the sonic region ends the stagnation points). In order to overcome this problem, Harten's entropy correction is implemented [4].

$$
|\tilde{\lambda}| = \begin{cases} |\tilde{\lambda}| & \text{if } |\tilde{\lambda}| > \delta \\ \dfrac{\tilde{\lambda}^2 + \delta^2}{2\delta + \epsilon} & \text{if } |\tilde{\lambda}| \leq \delta \end{cases}
\tag{17}
$$

where $\epsilon = 10^{-12}$ is a small positive parameter chosen in order to avoid divisions by zero of a null quantity, and $\delta = 0.2$.

3.3 Second-Order Accuracy

The Monotone Upstream-centered scheme for conservation laws (MUSCL) technique is used in this study to obtain second-order spatial accuracy. For the conservative variables W, the left W_{ij} and right W_{ji} states are extrapolated to the interface from values of variables at nodes i and j as.

$$
\begin{cases} W_{ij} = W_i + \dfrac{1}{2}(\nabla W_i.\mathbf{r}_{ij}) \\ W_{ji} = W_j - \dfrac{1}{2}(\nabla W_j.\mathbf{r}_{ij}) \end{cases}
\tag{18}
$$

where ∇W_i and ∇W_j are the gradients of W at nodes i and j, respectively. $r_{ij} = x_j - x_i$, represent the vector from node i to node j. The gradients are computed in this study using the green-gauss approximation [15] as

$$\nabla W_i = \frac{1}{\Omega} \sum_{j=1}^{n_{\text{edges}}} \frac{1}{2} (W_i + W_j) n_{ij} \Delta S_{ij} \tag{19}$$

As high-order schemes tend to produce spurious oscillations in the vicinity of large gradients, a slop limiter needs to be used to preserve the local Extrema diminishing property. In order to overcome this problem, the Venkatakrishnan's slop limiter is implemented [12].

4   Time Integration

An explicit five-stage second-order Runge-Kutta (RK) [16], scheme with local time-stepping and central implicit residual smoothing is used to advance the solution in time. The five stages are given by.

$$
\begin{aligned}
W^{(1)} &= W^{(n)} - \alpha_1 \Delta t R(W^{(n)}) \\
W^{(2)} &= W^{(1)} - \alpha_2 \Delta t R(W^{(1)}) \\
W^{(3)} &= W^{(2)} - \alpha_3 \Delta t R(W^{(2)}) \\
W^{(4)} &= W^{(3)} - \alpha_4 \Delta t R(W^{(3)}) \\
W^{(5)} &= W^{(4)} - \alpha_5 \Delta t R(W^{(4)})
\end{aligned}
\tag{20}
$$

where the constants value are $\alpha_1 = 0.0695, \alpha_2 = 0.1602, \alpha_3 = 0.2898, \alpha_4 = 0.5060$, $\alpha_5 = 1.0$, Δt is the time-steeping, and $R(W)$ the residual.

In this study, the local time-steeping is determined by:

$$\Delta t_i = \text{CFL} \frac{\Omega_i}{(|u| + a)\Delta S_x + (|v| + a)\Delta S_y} \tag{21}$$

where CFL represent the courant-Friedrichs-Lewy number, Ω_i the volume of the control volume, ΔS_x, and ΔS_y represent projections of the control volume in the x and y-plane.

5   Result and Discussion

This section present the results achieved to check the validity of employed method. The problem to be solved is the inviscid supersonic flow over a ramp in a 2D channel. The ramp, as well as the boundary conditions, are presented in Fig. 2, in which the slip boundary conditions are imposed for the upper and lower parts of the domain, the free-stream condition is imposed for the left-hand of the domain, and for the right end of the domain, the variables are simply extrapolated from the interior nodes. The Mach number at the inlet is set at 2.0. The static temperature and pressure are 300 K and 101,325 Pa, respectively. The fluid flow is simulated up to the convergence of
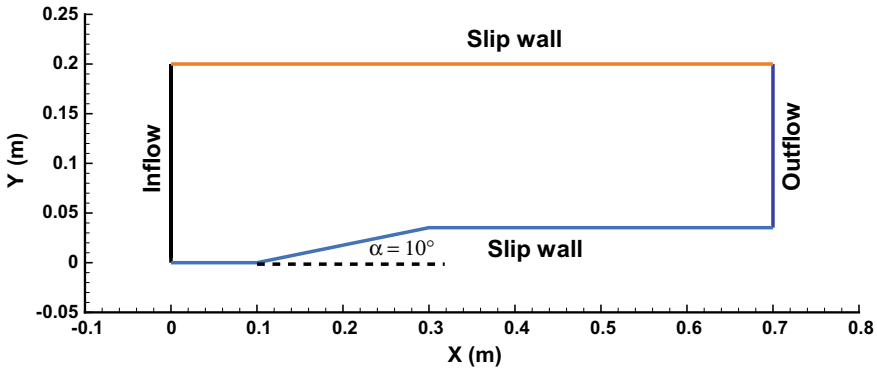
Fig. 2 Computational domain for the supersonic flow over a channel with ramp
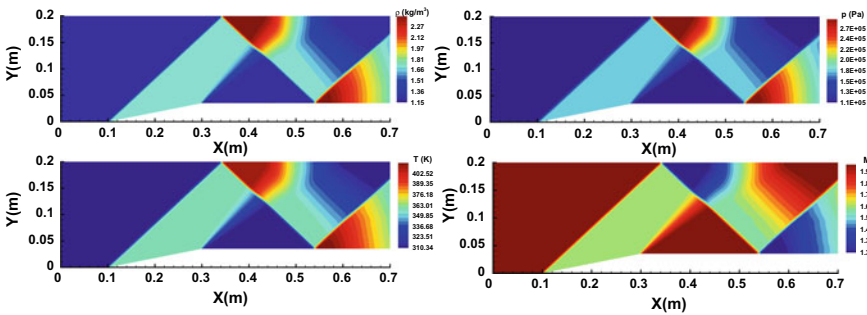
Fig. 3 The contours fields for the inviscid supersonic flow over a channel with ramp

the result to the steady-state with CFL number 5.5. The flow is characterized by an oblique shock generated due to the change in the direction of the supersonic flow caused by the ramp (Fig. 2).

Figure 3 shows the contours of the simulated density, pressure, temperature, and Mach number. The results show physical wave capture, robustness and accuracy in the expansion region, and nice shock capabilities. In Fig. 4a, the computed pressure distribution along the lower wall using three grid points (coarse, medium, and fine) is presented. On the coarsest grid, the solution oscillates, but it is rather smooth on all the other grids. It's worth noting that grid refinement has no effect on the location of the shocks; only the steepness improves.

Figure 4b shows the convergence profile of the residual L2 norm. It shows that the solver is stable and converge with the same error for the three grid point used.

6 Conclusion

A supersonic flow with an oblique shock over a ramp in a channel has been simulated using the Euler equations. The governing equations have been solved numerically

(a): Profile of pressure   (b): Convergence historic
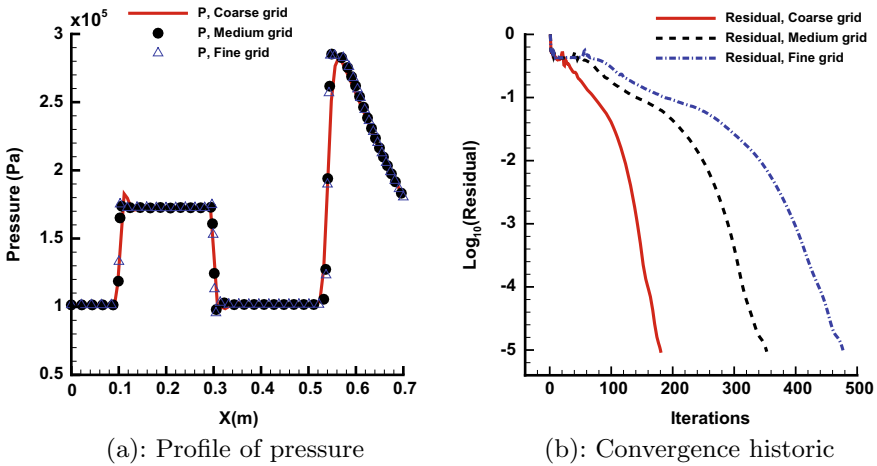
Fig. 4 Pressure distribution along the lower wall and convergence history for inviscid supersonic flow over a channel with ramp

by a cell-vertex finite volume method. The convective flux is assessed by Roe's approximate Riemann solver. The method has shown its reliability and robustness to calculate this type of numerical problem.

References

1. Godunov SK (1959) Finite-difference method for numerical computation of discontinuous solution of the equations of fluid dynamics. Matema Sbornik 47:271–306
2. Osher S, Solomon F (1982) Upwind difference schemes for hyperbolic systems of conservation laws. Math Comp 30(158):339–374
3. Roe PL (1981) Approximate Riemann solvers, parameter vectors, and difference schemes. J Comput Phys 43(2):357–372. https://doi.org/10.1016/0021-9991(81)90128-5
4. Harten A (1983) High resolution schemes for hyperbolic conservation laws. J Comput Phys 49:357–393
5. Kim S-S, Kim C, Rho O-H, Hong SK (2003) Cures for the shock instability: development of a shock-stable roe scheme. J Comput Phys 185(2):342–374
6. Qu F, Yan C, Sun D, Jiang Z (2015) A new roe-type scheme for all speeds. Comput Fluids 121:11–25
7. Musa O, Huang G, Yu Z, Li Q (2020) An improved Roe solver for high order reconstruction schemes. Comput Fluids 207:104591. ISSN 0045-7930
8. Van Leer B (1974) Towards the ultimate conservation difference scheme: II. Monotonic and conservation combined in a second order scheme. J Comp Phys 14:361–370
9. Harten A (1983) High resolution schemes for hyperbolic conservation laws. J Comput Phys 49:357–393
10. Chakravarthy S, Osher S (1985) A new class of high accuracy TVD schemes for hyperbolic conservation laws. AIAA Paper 85-0363
11. Barth TJ, Jespersen DC (1989) The design and application of upwind schemes on unstructured meshes. AIAA Paper 89-0366

12. Venkatakrishnan V (1993) On the accuracy of limiters and convergence to steady state solutions. AIAA Paper 93-0880

13. Toro EF (1999) Riemann solvers and numerical methods for fluid dynamics: a practical introduction. Springer, Berlin

14. Anderson WK, Bonhaus DL (1994) An implicit upwind algorithm for computing turbulent flows on unstructured grids. Comput Fluids 23(1):1–21. https://doi.org/10.1016/0045-7930(94)90023-X

15. Barth TJ, Jespersen DC (1989) The design and application of upwind schemes on unstructured meshes. AIAA Paper 89-0366

16. Jameson A, Schmidt W, Turkel E (1981) Numerical solution of the Euler equations by finite volume methods using Runge-Kutta time stepping schemes. AIAA Paper 81-1259. https://doi.org/10.2514/6.1981-1259

Modeling of the Magnetohydrodynamic Flow of the Nanofluid Confined Between Two Parallel Plates

Check for updates

Manar Ennaouri and El-Kaber Hachem

Abstract Squeezing flow between parallel plates is very applicable in many industrial and biomedical applications. In light of this fact, this research presents an analytical study of unsteady squeezing nanofluid flow and heat transfer between two parallel plates under the effect of a variable magnetic field using the two-phase Buongiorno model. The nonlinear differential equations governing the problem are solved by the Homotopy perturbation method (HPM). The key parameters addressed in this analysis are the Hartmann number, squeeze number, Eckert number, Brownian motion parameter, thermophoresis parameter, and Schmidt number. To demonstrate the high accuracy of the results, the obtained results were compared to the direct numerical solutions (NM) which show good agreement. Then, for each parameter, a graphical analysis is proposed to examine in detail the flow and heat transfer characteristics. The results show that the improvement of the thermophoresis parameter increased the temperature profile while decreasing the concentration profile. Furthermore, the results show that raising the Brownian motion parameter increases the concentration profile while having no influence on the temperature profile.

Keywords Two-phase model · Homotopy perturbation method (HPM) · Magnetic field · Brownian motion · Thermophoresis phenomenon

Nomenclature

c_p specific heat at constant pressure.
D_T thermophoresis diffusion coefficient.
Ha   Hartmann number.

M. Ennaouri · E.-K. Hachem (✉)
Innovative Research and Applied Physics Team (RIPA), Faculty of Sciences, Moulay Ismail University, B.P. 11201 Zitoune, Meknes, Morocco
e-mail: e.hachem@umi.ac.ma

Nt   thermophoresis parameter.
p pressure.
S squeeze number.
T nanofluid temperature.
C nanofluid concentration.
D_B Brownian diffusion coefficient.
Ec   Eckert number.
k_{nf} thermal conductivity of nanofluid.
α rate of squeezing.
Nb   Brownian motion parameter.
Pr   Prandtl number.
Sc   Schmidt number.
Θ Dimensionless temperature.
ρ_{nf} density of nanofluid.
ϕ Dimensionless concentration.
μ_{nf} dynamic viscosity of nanofluid.
σ_{nf} electrical conductivity of nanofluid.

1   Introduction

During the last few decades, the study of heat and mass transfer in unsteady squeezing nanofluid flow between two parallel plates known a great attention for its various industrial and biomedical applications such as processing of polymer, nuclear plants, compression, food processing, and cooling water.

The first work in this sector was reported by Stefan [1] using lubrication approximation. After Stefan's work, many researchers have studied squeeze flow in different aspects. Magnetohydrodynamic squeezing flow of a viscous fluid between infinite parallel plates was reported by Siddiqui et al. [2]. Domairry and Aziz [3] were study the impacts of suction or injection on squeezing flow of magnetohydrodynamic viscous fluid between parallel. Hayat et al. [4] were develop the work of Domairry and Aziz [3] for non-Newtonian fluids and studied the squeezing flow of second-grade fluid. Squeezing flow over a porous surface has been addressed by Mahmood et al. [5], further, a study regarding thermo-diffusion and diffuso-thermo effects on magnetohydrodynamic flow between parallel disks was reported by Khan et al. [6]. Furthermore, El Harfouf et al. [7] studied and discussed squeezing flow between plates using Cattaneo–Christov heat flux model.

Most important engineering problems, particularly heat transfer equations are nonlinear, so some of them are solved using numerical methods and some are solved using different analytical methods. Several analytical methods are used by researchers in this sector, among these methods we find: differential transformation method (DTM) [8–10], variation of parameter method (VPM) [11], homotopy analysis method (HAM) [12], Adomian decomposition method (ADM) [13], Duan–Rach approach (DRA) [14], homotopy perturbation method (HPM) [15, 16].

Based on the findings of the previous investigations, this paper deals with the squeezing flow of nanofluid under the effect of a variable magnetic field, the viscous dissipation effect and the generation of heat due to friction caused by shear in the flow are retained using the two-phase model. This model proposed by Buongiorno [17] takes into consideration the Brownian motion and the thermophoresis effects, which is different from the other models used.

The governing equations were transformed into ordinary differential equations. One of the analytical simulation methods that do not need small parameters is the HPM, this method was proposed and improved by He [18]. In this method, the solution is considered as the summation of an infinite series which converges rapidly to accurate solutions. In the present study, the reduced ordinary differential equations have been solved by HPM compared with the direct numerical solution NM so that the high accuracy of the results is clear.

2 Problem Description and Governing Equations

In this study, the analysis of heat and mass transfer in the unsteady two-dimensional compression flow of an incompressible viscous fluid between two infinite parallel plates is shown in Fig. 1. The two plates separated by a distance equal to 2 h(t), one plate is placed at $y = +h(t)$ and the other at $y = -h(t)$ with

$$h(t) = l(1 - \alpha t)^{\frac{1}{2}}$$

and l is the position of the plates at time $t = 0$ with respect to the origin of the reference. For $\alpha > 0$, the two plates are squeezed until they touch $t = 1/\alpha$ and for $\alpha < 0$ the two plates are separated. The viscous dissipation effect and the generation of heat due to friction caused by shear in the flow are retained, and all body forces are neglected. A variable magnetic field of strength

Fig. 1 Schematic of the problem (nanofluid between parallel plates in the presence of variable magnetic field)
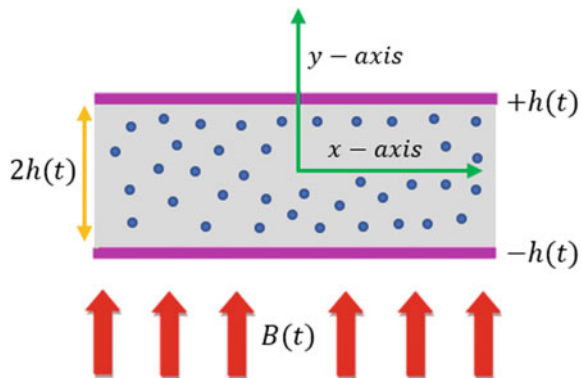
$$B(t) = B_0(1 - \alpha t)^{-\frac{1}{2}}$$

is assumed to be applied toward direction y. The electric current and the electromagnetic forces are defined by:

$$\vec{J} = \sigma\left(\vec{V} \wedge \vec{B}\right) \quad \text{and} \quad \vec{F} = \vec{J} \wedge \vec{B}. \text{ Then } \overrightarrow{F} = \sigma\left(\vec{V} \wedge \vec{B}\right) \wedge \vec{B}.$$

The governing equations for mass, momentum, energy in unsteady two-dimensional flow of a nanofluid are:

$$\overrightarrow{\nabla} \cdot \vec{V} = 0 \tag{1}$$

$$\rho_{\mathrm{nf}} \frac{\partial \vec{V}}{\partial t} + \rho_{\mathrm{nf}} \vec{V} \cdot \overrightarrow{\nabla} \vec{V} = -\overrightarrow{\nabla} \cdot P + \mu_{\mathrm{nf}} \nabla^2 \vec{V} + \sigma_{\mathrm{nf}}\left(\vec{J} \wedge \vec{B}\right) \tag{2}$$

$$(\rho c_{\mathrm{p}})_{\mathrm{nf}}\left[\frac{\partial T}{\partial t} + \vec{V} \cdot \overrightarrow{\nabla} T\right] = \overrightarrow{\nabla} \cdot K_{\mathrm{nf}} \overrightarrow{\nabla} T$$

$$+ (\rho c_{\mathrm{p}})_{\mathrm{p}}\left[D_{\mathrm{B}} \overrightarrow{\nabla} C \cdot \overrightarrow{\nabla} T + D_{\mathrm{T}} \frac{\overrightarrow{\nabla} T \cdot \overrightarrow{\nabla} T}{T_{\infty}}\right] + \mu_{\mathrm{nf}} \Omega \tag{3}$$

$$\frac{\partial C}{\partial t} + \vec{V} \overrightarrow{\nabla} \cdot C = D_{\mathrm{B}} \nabla^2 C + \frac{D_{\mathrm{T}} \nabla^2 T}{T_{\infty}} \tag{4}$$

where u and v are the velocities in the x- and y-directions, respectively, T is the temperature, P is the pressure, σ_{nf} is the electric conductivity of nanofluid, K_{nf} is the thermal conductivity of nanofluid, C is the concentration of nanoparticles, μ_{nf} is the dynamic viscosity of nanofluid, D_{B} and D_{T} the Brownian motion coefficient and the thermophoretic diffusion coefficient, and T_{∞} is the mean fluid temperature.

We can also define the operation of $\overrightarrow{\nabla}$ as $\overrightarrow{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$ and $\mu_{\mathrm{nf}} \Omega$ the expression of the viscous dissipation effect as:

$$\mu_{\mathrm{nf}} \Omega = \mu_{\mathrm{nf}}\left[4\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2\right].$$

The boundary conditions are:

$$u = 0, \quad v = \frac{\mathrm{d}h}{\mathrm{d}t}, \quad T = T_{\mathrm{H}}, \quad C = C_{\mathrm{H}} \quad \text{at} \quad y = h(t) \tag{5}$$

$$\frac{\mathrm{d}u}{\mathrm{d}y} = 0, \quad v = 0, \quad \frac{\mathrm{d}T}{\mathrm{d}y} = 0, \quad \frac{\mathrm{d}C}{\mathrm{d}y} = 0 \quad \text{at} \quad y = 0 \tag{6}$$

The following dimensionless groups are introduced:

$$\eta = \frac{y}{\left[l(1-\alpha t)^{\frac{1}{2}}\right]} = \frac{y}{h(t)}, u = \frac{\alpha x}{2(1-\alpha t)} f'(\eta), v = -\frac{\alpha l}{2(1-\alpha t)^{\frac{1}{2}}} f(\eta) \quad (7)$$

$$\Theta = \frac{T}{T_H}, \quad \phi = \frac{C}{C_H} \quad (8)$$

Substituting the above parameters into Eqs. (2–4) with elimination of the pressure gradient the resulting equations will give:

$$f'''' + S(f f''' - 3f'' - \eta f''' - f' f'') - \mathrm{Ha}^2 f'' = 0 \quad (9)$$

$$\Theta'' + \mathrm{Nb}(\Theta'\phi') + \mathrm{Nt}\,\Theta'^2 + \mathrm{Pr}\,\mathrm{Ec}(f''^2 + 4\delta^2 f'^2) + \mathrm{Pr}\,S(f\Theta' - \eta\Theta') = 0 \quad (10)$$

$$\phi'' + \mathrm{Sc}\,S(f\phi' - \eta\phi') + \left(\frac{\mathrm{Nt}}{\mathrm{Nb}}\right)\Theta'' = 0 \quad (11)$$

The boundary condition in the new similar variables becomes:

$$f''(0) = 0, \, f(0) = 0, \, \Theta'(0) = 0, \, \phi'(0) = 0 \quad \text{at} \quad \eta = 0 \quad (12)$$

$$f'(1) = 0, \, f(1) = 1, \, \Theta(1) = \phi(1) = 1 \quad \text{at} \quad \eta = 1 \quad (13)$$

where S is the squeeze number, Ha is the Hartmann number, Nb is the Brownian motion parameter, Nt is the thermophoretic parameter, Pr is the Prandtl number, and Sc is the Schmidt number which are defined as:

$$S = \frac{\alpha l^2 \rho_{\mathrm{nf}}}{2\mu_{\mathrm{nf}}}, \quad \mathrm{Ha} = l B_0 (\sigma_{\mathrm{nf}}(1-\alpha t)/\mu_{\mathrm{nf}})^{1/2}, \quad \mathrm{Nb} = \frac{D_B C_H}{K_{\mathrm{nf}}}(c_p)_p,$$

$$\mathrm{Nt} = \frac{D_T T_H}{K_{\mathrm{nf}} T_\infty}(\rho c_p)_p, \quad \mathrm{Pr} = \frac{\mu_{\mathrm{nf}}}{K_{\mathrm{nf}}} c_p, \quad \mathrm{Sc} = \frac{\mu_{\mathrm{nf}}}{\rho_{\mathrm{nf}} D_B},$$

$$\mathrm{Ec} = \frac{1}{4c_p}\left(\frac{\alpha x}{(1-\alpha t)}\right)^2, \quad \delta = \frac{l}{x}(1-\alpha t)^{\frac{1}{2}}$$

3 Implementation of Homotopy Perturbation Method

In this section, we employ HPM to solve Eqs. (9–11) subject to boundary conditions Eqs. (12, 13). We can construct Homotopy functions for the Eqs. (9–11) as shown in reference [16].

We obtain:

$$H(f, p) = (1 - p)\left[f'''' - f_0''''(0)\right]$$
$$+ p\left[f'''' + S\left(ff''' - 3f'' - \eta f''' - f'f''\right) - \mathrm{Ha}^2 f''\right] = 0 \qquad (14)$$

$$H(\Theta, p) = (1 - p)\left[\Theta'' - \Theta_0''(0)\right]$$
$$+ p\left[\Theta'' + (\mathrm{Nb})\left(\Theta'\phi'\right) + (\mathrm{Nt})\Theta'^2\right.$$
$$\left. + (\mathrm{Pr\,Ec})\left(f''^2 + 4\delta^2 f'^2\right) + (\mathrm{Pr\,}S)\left(f\Theta' - \eta\Theta'\right)\right] = 0 \qquad (15)$$

$$H(\phi, p) = (1 - p)\left[\phi'' - \phi_0''(0)\right] + p\left[\phi'' + (\mathrm{Sc\,}S)\left(f\phi' - \eta\phi'\right) + \left(\frac{\mathrm{Nt}}{\mathrm{Nb}}\right)\Theta''\right] \qquad (16)$$

We consider as f, Θ, and ϕ follows:

$$f(\eta) = f_0(\eta) + p^1 f_1(\eta) + p^2 f_2(\eta) + p^3 f_3(\eta) + \cdots = \sum_{i=0}^{N} p^i f_i(\eta) \qquad (17)$$

$$\Theta(\eta) = \Theta_0(\eta) + p^1 \Theta_1(\eta) + p^2 \Theta_2(\eta) + p^3 \Theta_3(\eta) + \cdots = \sum_{i=0}^{N} p^i \Theta_i(\eta) \qquad (18)$$

$$\phi(\eta) = \phi_0(\eta) + p^1 \phi_1(\eta) + p^2 \phi_2(\eta) + p^3 \phi_3(\eta) + \cdots = \sum_{i=0}^{N} p^i \phi_i(\eta) \qquad (19)$$

With substituting f and Θ and ϕ from Eqs. (17–19) into Eqs. (14–16) and some simplification and rearranging based on powers of p–terms, we have

$$p^0 : \quad f_0'''' = 0, \quad \Theta_0'' = 0, \quad \phi_0'' = 0 \qquad (20)$$

with the boundary conditions are:

$$f_0''(0) = 0,\ f_0(0) = 0,\ \Theta_0'(0) = 0,\ \phi_0'(0) = 0 \quad \text{at} \quad \eta = 0 \qquad (21)$$

$$f_0'(1) = 0,\ f_0(1) = 1,\ \Theta_0(1) = \phi_0(1) = 1 \quad \text{at} \quad \eta = 1 \qquad (22)$$

$$p^1 : f_1'''' + S(f_0 f_0''' - 3f_0'' - \eta f_0''' - f_0' f_0'') - \mathrm{Ha}^2 f_0'' = 0$$

$$\Theta_1'' + (\mathrm{Pr\ Ec})\left(f_0''^2 + 4\delta^2 f_0'^2\right) + (\mathrm{Pr\ } S)\left(f_0\Theta_0' - \eta\Theta_0'\right) + (\mathrm{Nb})\left(\Theta_0'\phi_0'\right) + (\mathrm{Nt})\Theta_0'^2 = 0 \qquad (23)$$

$$\phi_0'' + (\mathrm{Sc\ } S)\left(f_0\phi_0' - \eta\phi_0'\right) + \left(\frac{\mathrm{Nt}}{\mathrm{Nb}}\right)\Theta_0'' = 0$$

with the boundary conditions are:

$$f_1''(0) = 0, \quad f_1(0) = 0, \quad \Theta_1'(0) = 0, \quad \phi_1'(0) = 0 \quad \text{at} \quad \eta = 0 \qquad (24)$$

$$f_1'(1) = 0, \quad f_1(1) = 0, \quad \Theta_1(1) = 0, \quad \phi_1(1) = 0 \quad \text{at} \quad \eta = 1 \qquad (25)$$

Solving equations of Eq. (20) and of Eq. (23) with these boundary conditions, we have:

$$f_0(\eta) = -\frac{1}{2}\eta^3 + \frac{3}{2}\eta, \quad \Theta_0(\eta) = 1, \quad \phi_0(\eta) = 1 \qquad (26)$$

$$f_1(\eta) = \frac{1}{280} S\eta^7 - \frac{1}{40}\left(\mathrm{Ha}^2 + 4S\right)\eta^5 + \frac{1}{6}\left(\frac{3}{10}\mathrm{Ha}^2 + \frac{159}{140}S\right)\eta^3$$
$$+ \left(-\frac{1}{40}\mathrm{Ha}^2 - \frac{13}{140}S\right)\eta \qquad (27)$$

$$\Theta_1(\eta) = -9\,\mathrm{Pr\ Ec}\left(\frac{1}{30}\delta^2\eta^6 - \frac{1}{6}\delta^2\eta^4 + \frac{1}{12}\eta^4 + \frac{1}{2}\delta^2\eta^2\right) + \frac{33}{10}\mathrm{Ec\ Pr} + \frac{3}{4}\mathrm{Pr\ Ec} \qquad (28)$$

$$\phi_1(\eta) = 0 \qquad (29)$$

The terms $f_i(\eta)$, $\Theta_i(\eta)$ and $\phi_i(\eta)$ when $i > 1$ are too large that is mentioned graphically. The solution of equations is obtained when $p \to 1$ will be as follows:

$$f(\eta) = f_0(\eta) + f_1(\eta) + f_i(\eta) + f_i(\eta) + \cdots = \sum_{i=0}^{N} f_i(\eta) \qquad (30)$$

$$\Theta(\eta) = \Theta_0(\eta) + \Theta_1(\eta) + \Theta_2(\eta) + \Theta_3(\eta) + \cdots = \sum_{i=0}^{N} \Theta_i(\eta) \qquad (31)$$

$$\phi(\eta) = \phi_0(\eta) + \phi_1(\eta) + \phi_2(\eta) + \phi_3(\eta) + \cdots = \sum_{i=0}^{N} \phi_i(\eta) \qquad (32)$$

4 Results and Discussion

In this paper, HPM is applied to solve the problem of the magnetohydrodynamic squeezing flow of nanofluid between parallel plates. The effect of active parameters on heat and mass specifications is investigated. The present code is validated by comparing the obtained result with the direct numerical solution NM for $-f''(1)$ and $-\Theta'(1)$ when Ec = Pr = 1, Ha = Sc = 0, and $\delta = 0.1$ for different values of S as it is seen in Table 1. From Table 1, it can be seen that the results obtained by the homotopy perturbation method are in full agreement with the results of the numerical method.

Throughout the computation, values of existing parameters are taken as:

$$S = 0.5, \text{Ha} = 1, \text{ Pr} = 6.2, \text{ Ec} = 0.01, \delta = 0.1, \text{ Nb} = 0.1, \text{ Nt} = 0.1, \text{ Sc} = 0.2.$$

The impact of Hartman number Ha and squeeze number S on the velocity profile is shown in Fig. 2. It is inferred from the figure that the velocity profile decreases as the Hartmann number increases. From a physical point of view, the magnetic field exerts a force which is called the force of Lorentz, it is a force that acts against the flow if the magnetic field is applied in the normal direction. So, when the Hartmann number increases velocity profile decreases. On the other hand, an increase in the squeeze number is associated with the decrease in the kinematic viscosity, when the plates approach each other they compact the flow, which makes a decrement in velocity. As can be seen in Fig. 3, the effect of Hartman number and squeeze number on the temperature profile. According to this figure, an increase in the value of the Hartman number causes decreases in the temperature profile, as well as the temperature profile decreases with an increment in squeeze number. Furthermore, Fig. 4 demonstrates the impact of Eckert number on the temperature profile and the effect of Schmidt number on concentration profile. It is important to mention that the Eckert number Ec characterizes the influence of self-heating of a fluid as a consequence of dissipation effect, and this plot shows that the augmentation of the Eckert number causes increases in temperature profile because the temperature profile in the nanofluid is not just dominated by the temperature gradients that are present in the nanofluid, but also affected by the dissipation due to internal friction of the nanofluid. On the other hand, Schmidt number is a dimensionless number

Table 1 Comparison between HPM and NM for $-f''(1)$ and $-\Theta'(1)$

S	HPM		NM	
	$-f''(1)$	$-\Theta'(1)$	$-f''(1)$	$-\Theta'(1)$
-1.0	2.170092	3.319860	2.170090	3.319899
-0.5	2.617403	3.129491	2.617403	3.129491
0.5	3.336449	3.026323	3.336449	3.026323
2.0	4.168065	3.127819	4.167389	3.118550

defined as the ratio of momentum diffusivity (*viscosity*) and mass diffusivity. Thus, the concentration profile increases with increasing Schmidt number.

Figures 5 and 6 are prepared in order to see the effects of the thermophoresis parameter Nt and the Brownian motion parameter Nb on the temperature and concentration profiles. It can be seen that these parameters have no significant effect on the temperature profile. While they have a similar effect on the concentration profile, increasing these parameters resulted in an increase in the concentration profile.
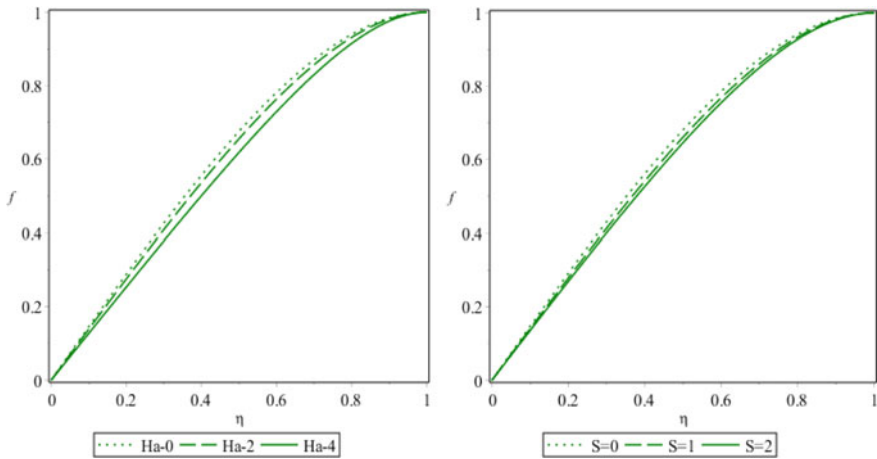
Fig. 2 Effect of Hartman number and squeeze number on the velocity profile
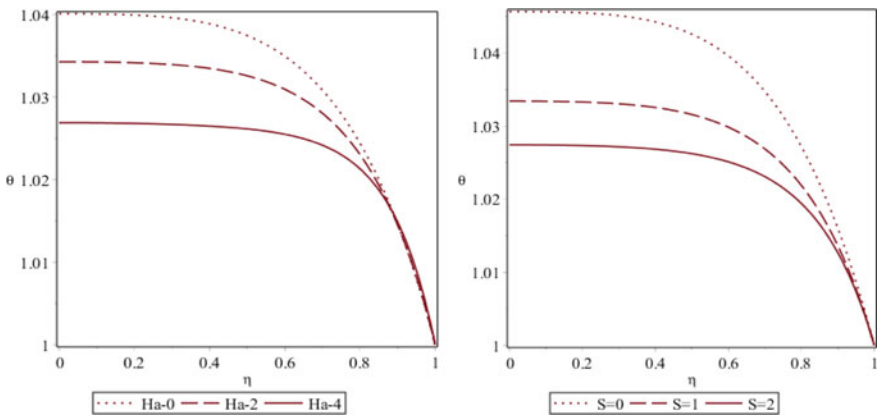
Fig. 3 Effect of Hartman number and squeeze number on the temperature profile
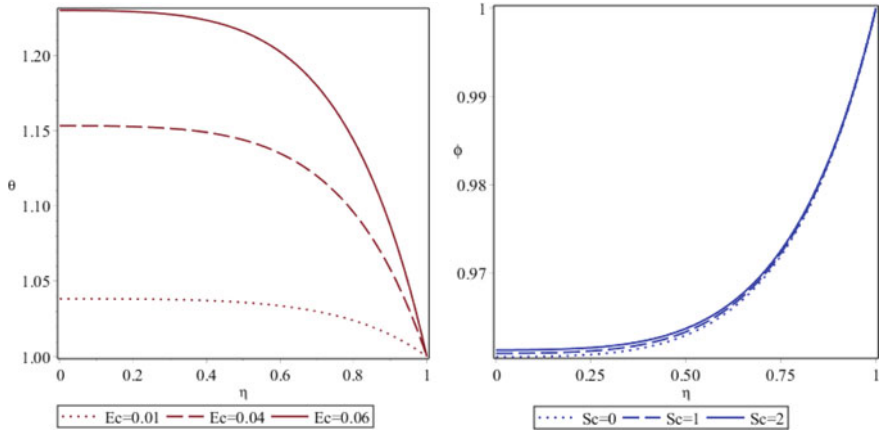
Fig. 4 Effect of Eckert number on the temperature profile and the effect of the Schmidt number on concentration profile
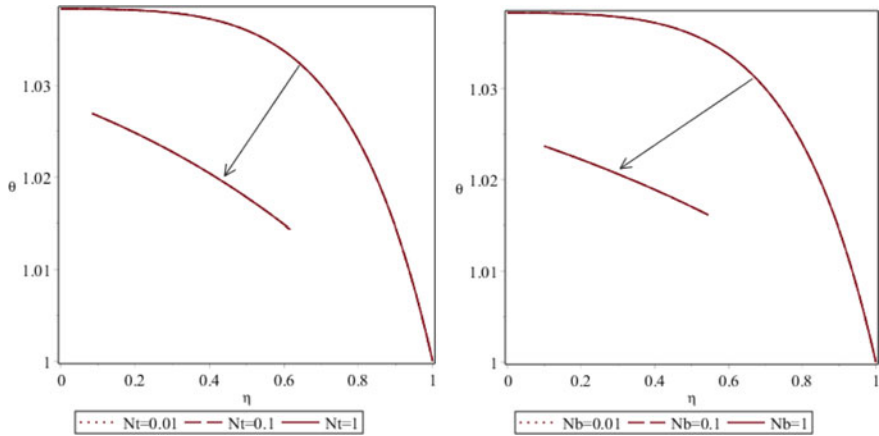
Fig. 5 Effect of thermophoresis parameter and Brownian motion parameter on temperature profile

5 Conclusion

In this work, unsteady squeezing nanofluid flow between parallel plates in the presence of variable magnetic field has been investigated analytically using the homotopy perturbation method. From the analysis, it is established that the results of HPM are in good agreement with the direct numerical solution NM. So, it can be used to solve this type of problem. Also, from the parametric study, it can be concluded that an increase in the Hartmann number and the squeeze number leads to a decrease in the velocity and temperature profiles. Also, thermophoresis parameter Nt and Brownian motion parameter Nb have significant effects on nanoparticle concentration profile,
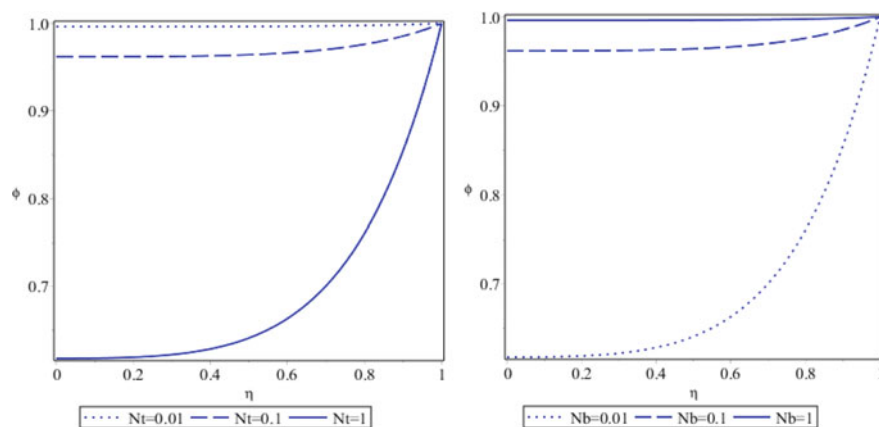
Fig. 6 Effect of thermophoresis parameter and Brownian motion parameter on concentration

and the concentration field is a decreasing function of thermophoresis parameter. However, it enhances by raising Brownian motion parameter.

References

1. Stefan MJ (1874) Versuch Uber die scheinbare adhesion. Sitzungsberichte der Academic der Wissenschaften in Wien. Mathematik-Naturwissen 69:713–721
2. Siddiqui AM, Irum S, Ansari AR (2008) Unsteady squeezing flow of a viscous MHD fluid between parallel plates, a solution using the homotopy perturbation method. Math Model Anal 13:565–576
3. Domairry G, Aziz A (2009) Approximate analysis of MHD squeeze flow between two parallel disks with suction or injection by homotopy perturbation method. Math Prob Eng 603–616
4. Hayat T, Yousaf A, Mustafa M, Obaidat S (2011) MHD squeezing flow of second grad fluid between parallel disks. Int J Num Meth Fluids 69:399–410
5. Mahmood M, Asghar S, Hossain MA (2007) Squeezed flow and heat transfer over a porous surface for viscous fluid. Heat Mass Transf 44:165–173
6. Khan SIU, Mohyud-Din ST, Mohsin BB (2016) Thermo-diffusion and diffuso-thermo effects on MHD squeezing flow between parallel disks. Surf Rev Lett. https://doi.org/10.1142/S0218625X17500226
7. El Harfouf A, Wakif A, Hayani Mounir S (2020) Analytical and numerical analysis of magneto hydrodynamic flow and heat transfer in a nanofluid via the Christov-Cattaneo heat flux theory. Sens Lett 18(8):643–657
8. Sheikholeslami M, Azimi M, Ganji DD (2015) Application of differential transformation method for nanofluid flow in a semi-permeable channel considering magnetic field effect. J Comput Meth Eng Sci Mech 16:246–255
9. Sheikholeslami M, Ganji DD (2015) Nanofluid flow and heat transfer between parallel plates considering Brownian motion using DTM. Comput Meth Appl Mech Eng 283:651–663
10. Jang MJ, Chen CL, Lin YC (2001) Two-dimensional differential transformation method for solving partial differential equation. Appl Math Comput 121:261–270
11. Mohyud-Din ST, Yildirim A (2010) Variation of parameter method for Fishers equation. Adv Appl Math Mech 2:379–388

12. Sheikholeslami M, Ellahi R, Ashorynejad HR, Hayat T (2014) Effects of heat transfer in flow of nanofluids over a permeable stretching wall in a porous medium. J Comput Theor Nanosci 11:486–496
13. Sheikholeslami M, Ganji DD, Ashorynejad HR (2013) Investigation of squeezing unsteady nanofluid flow using ADM. Powder Technol 239:259–265
14. Dogonchi AS, Ganji DD (2016) Investigation of MHD nanofluid flow and heat transfer in a stretching/shrinking convergent/divergent channel considering thermal radiation. J Mol Liq 220:592–603
15. Sheikholeslami M, Ganji DD (2013) Heat transfer of Cu-water nanofluid between parallel plates. Powder Technol 235:873–879
16. El Harfouf A, Wakif A, Mounir SH (2020) Heat transfer analysis on squeezing unsteady MHD nanofluid flow between two parallel plates considering thermal radiation, magnetic and viscous dissipations effects a solution by using Homotopy Perturbation method. Sens Lett 18(2):113–121
17. Buongiorno J (2006) Convective transport in nanofluids
18. He JH (2000) A new perturbation technique which is also valid for large parameters. J Sound Vib 229(5):1257–1263

Autonomous Solar Photovoltaic/Battery System for the Electrification of Wastewater Pumping Stations

Mohammed Chennaif, Mohamed Maaouane, Mohamed Larbi Elhafyani, Hassan Zahboune, Smail Zouggar, Jalal Blaacha, Mohammed El Fahssi, Omar Mommadi, and Jamal-Eddine Salhi

Abstract The wastewater treatment procedure generally requires pumping stations to allow the supply pipe to pass the topographical constraints of the site. Since these pumping stations are characterized by high energy consumption, and in parallel with the energy transition of our country, and national strategies for the integration of renewable energy systems, this paper presents an optimal sizing of an autonomous solar photovoltaic/battery system for supplying electricity to a wastewater pumping station. The Electric Systems Cascade Extended Analysis methodology, the LPSP as system reliability parameter, Life Cycle Cost, and Levelized Cost of Electricity are implemented together to obtain the optimal sizing of the proposed system. The hourly wastewater flow of the wastewater pumping station, the hourly solar irradiation of the site chosen for the case study, and the technical and economic data of the various components of the PV/Battery system are taken as inputs to the ESCEA method.

Keywords Photovoltaic/Battery · Wastewater pumping station · Optimal sizing · ESCEA

1 Introduction

Solar-based applications continue to gain popularity since they are a renewable, clean, and abundant source of energy that can reach most geographical regions, making them a perfect choice for powering remote places. Solar energy applications vary depending on uses and demands; photovoltaic/battery systems are one of the most

M. Chennaif (✉) · M. Maaouane · M. L. Elhafyani · H. Zahboune · S. Zouggar · J. Blaacha · M. El Fahssi · O. Mommadi · J.-E. Salhi
Université Mohammed Premier, Oujda, Morocco
e-mail: mohammed.chennaif@gmail.com

J.-E. Salhi
e-mail: j.salhi@ump.ac.ma

J.-E. Salhi
CREHEIO, (Centre de Recherche de l'Ecole des Hautes Etudes d'Ingénierie), Oujda, Morocco

861

essential, especially in energy-intensive sectors where high demand is enforced. The urban wastewater sector is presently being pushed to improve its energy consumption procedures in line with national energy efficiency and renewable energy initiatives, without sacrificing its quality requirements, because it is a system that consumes too much electricity.

Various interesting studies have been reported in the literature and have studied the wastewater pumping stations since the pumping is the largest consumer of electrical energy in a wastewater treatment plant [1]. The study [2] describes the design and implementation of a non-invasive, scalable, and resilient embedded-camera system for long-term and high-frequency automated monitoring of the accumulation of fat, oil, and grease (FOG) layer dynamics in the sump of wastewater pumping stations. Filipe et al. described in [1] an innovative predictive control policy for wastewater variable-frequency pumps that minimize electrical energy consumption, considering uncertainty forecasts for wastewater intake rate and information collected by sensors accessible through the supervisory control and data acquisition system. By combining information from data mining, mathematical modeling, and computational intelligence, a data-driven paradigm for increasing the performance of wastewater pumping systems has been established [3]. Odeh et al. used TRNSYS to create and test a transient simulation model for AC photovoltaic (PV) water pumping systems. The impact of mismatching pump and well system parameters on system performance, efficiencies, and average performance ratios for the system, subsystem, and PV array over time and under various operating situations is investigated [4]. The report details the results of an experimental investigation that looked into the performance of a basic, directly connected dc photovoltaic water pumping system [5]. PV array, DC motor, and centrifugal pump make up the system.

Through this work, we aim to propose an optimal sizing of an autonomous PV/Battery system for the electrification of a wastewater pumping station, using a simple and effective methodology called the Electric Systems Cascade Extended Analysis (ESCEA), on which we can find the optimal capacities of generation and stockage units of the said system under different technical and economic optimization criteria: the LPSP as system reliability parameter, Life Cycle Cost (LCC) and Levelized Cost of Electricity (LCOE) as economic indicators.

Section. 2 of this paper will describe the wastewater pumping system, the PV/Battery system and the modeling of its components. Section. 3 is devoted to the presentation of the ESCEA methodology adopted for the sizing of the PV/Battery system. A case study and all the findings are presented in the Sect. 4. Finally, we presented the most important conclusions in the Sect. 5.

2 Wastewater Pumping Station and PV/Battery System

2.1 Wastewater Pumping Station

When gravity cannot accomplish the job, a wastewater pumping station simply pumps the sewage from one area to another, horizontally and/or upstream.

Normally, the garbage from houses and businesses is drained into the sewage system by gravity. The garbage will have to be transferred via pumping stations in low locations when the main sewer is on higher ground than the connected sewer lines.

The wet well, pump, pipes with related valves and strainers, motor, power supply system, equipment control and alarm system, odor control system, and ventilation system are all important components of any pumping station. Pumping station systems and equipment are frequently housed in a closed building (Fig. 1).
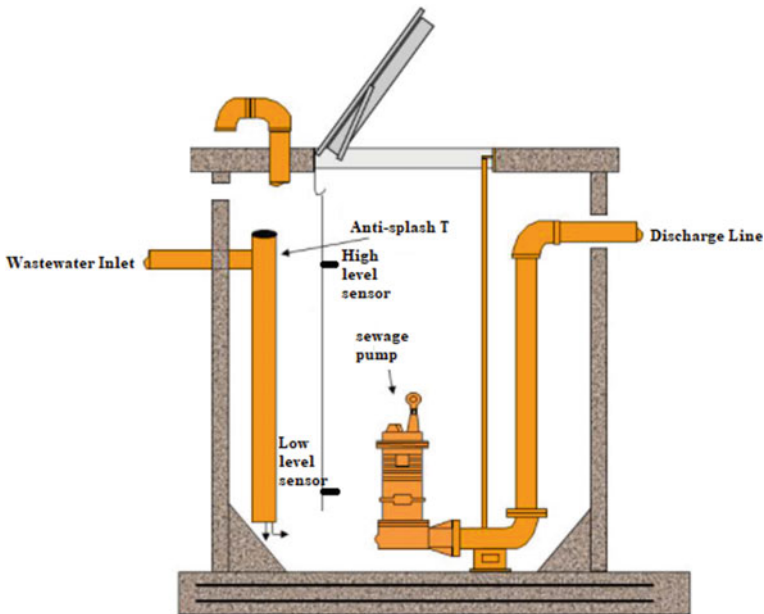
Fig. 1 Wastewater pumping station
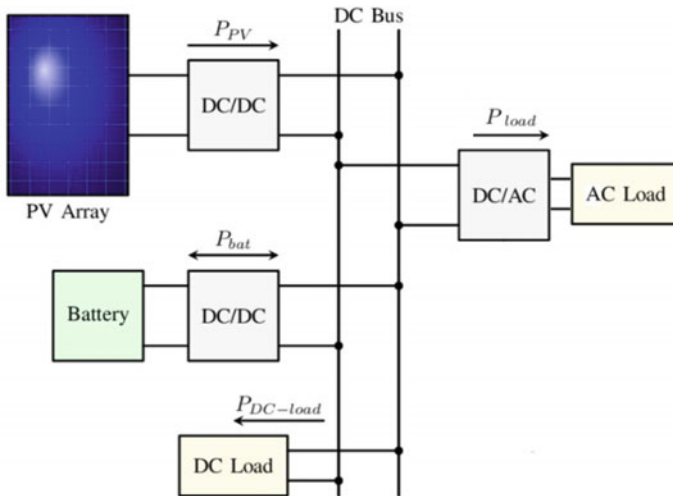
Fig. 2 Autonomous PV/Battery system

2.2 PV/Battery System

2.2.1 System Description

The PV/Battery combination is one of the most widely used renewable energy systems for electrification in recent years. Photovoltaic panels, batteries, and converters make up the PV/Battery system (Fig. 2).

The battery collects and stores the extra energy produced by the solar generators for usage at night or when there is no other source of energy. Pumps or motors may run intermittently because the battery might drain fast and create more current than the charging source can supply on its own.

2.2.2 Modeling of PV/Battery System

- **PV Array**

  Each PV array's efficiency varies depending on the kind of its PV panels and a variety of other factors such as temperature. PV array converts solar energy into electrical energy, which is then used to generate DC. During the present time step, the total hourly energy produced by the PV array may be expressed as follows in Eq. (1) [6, 7]:

$$E_{\text{PV}}(t) = A_{\text{PV}}(t) \times \eta_{\text{PV}} \times \text{GHI}(t) \tag{1}$$

A_{PV} is the PV array receiving surface (m^2) that will be determined by the ESCEA method and η_{PV} is the efficiency of PV array (%).

- **Load Profile**

The profile of the electrical load of the wastewater pumping station $E_L(t)$ depends on the hourly flow rate of the pump $Q(t)$, the total manometric head (H), the hydraulic constant C ($C =$ density of water $\times$ gravity constant), and the efficiency of the moto-pump group η_{M-P}, it can be calculated using Eq. 2 [8–10]:

$$E_L(t) = \frac{C \times Q(t) \times H}{\eta_{M-P}} \tag{2}$$

The net electrical excess/deficit $E_N(t)$ is the hourly difference between the solar system generation and the load demand taking into account the energy conversion losses, it is given by Eq. (3):

$$E_N(t) = E_{PV}(t) \times \eta_{DC/DC} - E_L(t) \times \eta_{DC/AC} \tag{3}$$

- **Battery Model**

The electricity produced by the PV array is used to satisfy the wastewater pumping station directly while the excess is used to charge the battery, and when the amount of electricity produced is insufficient, the battery can cover the difference. And so on, the battery allows us to store energy when there is an excess and covers deficit periods to guarantee the autonomy of the system. Then, the net amount of energy $E_N(t)$ calculated from Eq. (3), controls the hourly electricity charging CH(t) and discharging DCH(t) of the battery as shown in Eqs. (4, 5), respectively.

$E_N(t) > 0$; i.e., there is an excess of electricity that must be charged into the battery:

$$CH(t) = E_N(t) \times \eta_{CH} \tag{4}$$

where $E_N(t) < 0$, that means insufficient energy production, so the battery must discharge to cover the difference:

$$DCH(t) = \frac{E_N(t)}{\eta_{DCH}} \tag{5}$$

η_{CH} and η_{DCH} are the charging and discharging losses.

3 Electric Systems Cascade Extended Method (ESCEA)

For the optimal sizing of the PV/Battery system powering the wastewater pumping station, we used the algorithm of the ESCEA method. The method takes the power pinch analysis (PPA) a guideline technique for the sizing and the optimization of photovoltaic systems, to meet a specified load profile.

The sizing process followed by ESCEA algorithm (see Fig. 3) begins by extracting all relevant data from the location, including the hourly global horizontal irradiation (GHI(t)) and temperature (Temp (t)). Also, the analysis period T, the wastewater pumping station's hourly outlet flow via the discharge line ($Q(t)$), the total manometric head (H), and all technical and economic data of the PV/Battery system's components: type of photovoltaic panels, their characteristics, and cost; cost of the battery, its charge and discharge efficiencies, and depth of discharge (DOD); installation, operation and maintenance costs of the components, cost of the balance-of-system (BOS), cost of replacement, system lifetime, and the interest rate.

3.1 The ESCEA Algorithm for Sizing Autonomous PV/Battery System

The following requirements must be satisfied to guarantee proper execution of the ESCEA algorithm [11]:

- For each time step (1 h in this study), the load demand, temperature, and global horizontal sun irradiation must stay constant.
- All losses in energy conversion must be taken into account.
- The quantity of energy stored in the battery must be positive at all times.
- The quantity of energy stored in the battery at the conclusion of the time period ($t = T$) should be almost the same as it was at the start ($t = 0$).

3.2 Reliability and Economic Analysis

3.2.1 Loss of Power Supply Probability

The LPSP, which is defined as the proportion of power supply that cannot meet load demand, is now one of the most used indexes for evaluating the dependability of a renewable energy system and optimizing the size of power generating and storage unit [12].

$$\text{LPSP} = \frac{\sum_{t=0}^{T} E_{\text{defecit}}}{\sum_{t=0}^{T} E_{\text{L}}} \tag{6}$$
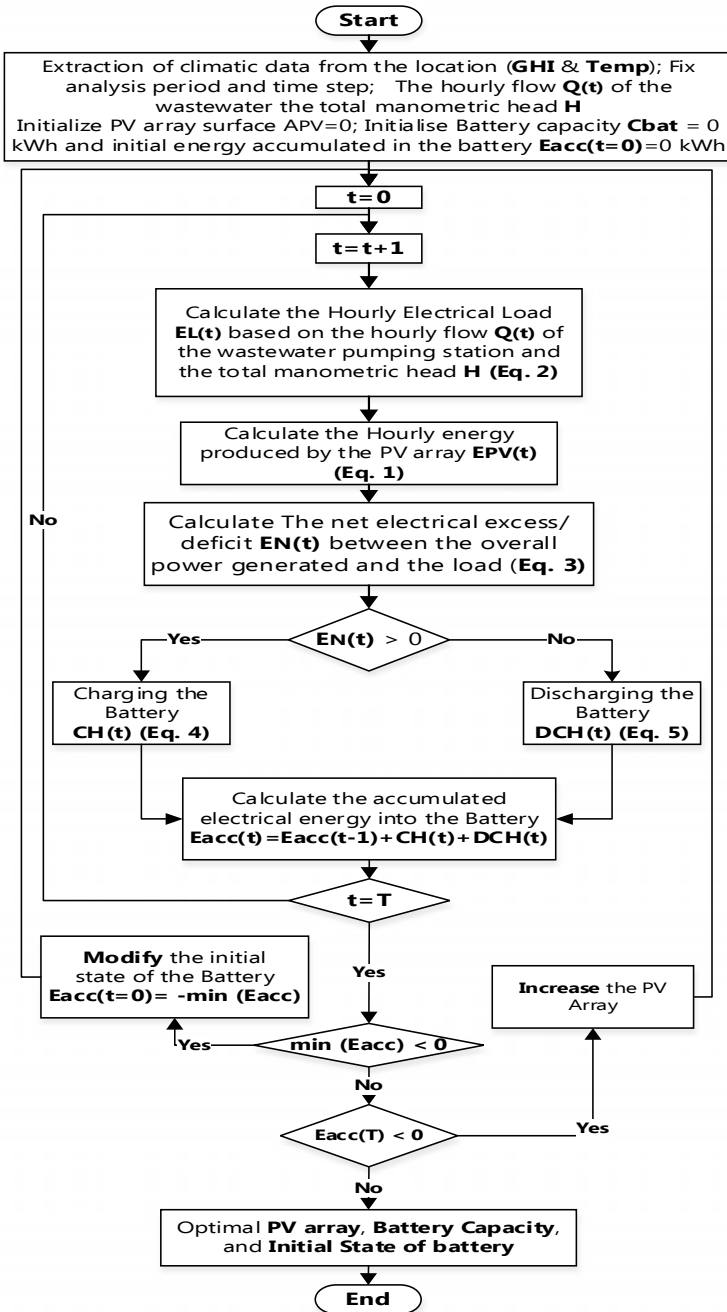
Fig. 3 ESCEA algorithm for sizing autonomous PV/Battery system

3.2.2 Levelized Cost of Electricity

The LCOE is an economic statistic that represents the whole cost of energy over the lifespan of the system that produces it. The LCOE is calculated in the following way:

$$\text{LCOE} = \frac{\text{LCC}}{\sum_{t=0}^{T} \frac{E_t}{(1+r)^t}} \tag{7}$$

4 Case Study

A case study is proposed in this section to apply the ESCEA method in order to obtain the optimal sizing of a PV/Battery system powering wastewater pumping stations. The algorithm of the ESCEA method for sizing is implemented for three load profiles of three different stations (Fig. 4), each characterized by its flow rate and its total manometric head H.

The sizing by the ESCEA method is based on the solar irradiation of the site. The GHI solar irradiation profile is shown in Fig. 5.

The economic and technical data of the PV/Battery system are presented in Table 1.

Fig. 4 Wastewater pumping stations' profiles
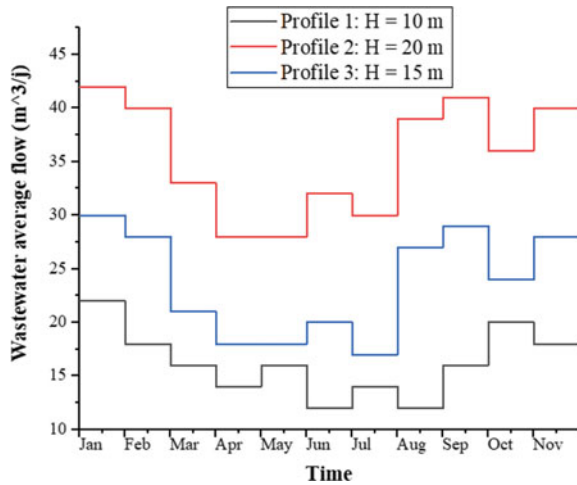
Fig. 5 GHI profile of the case study
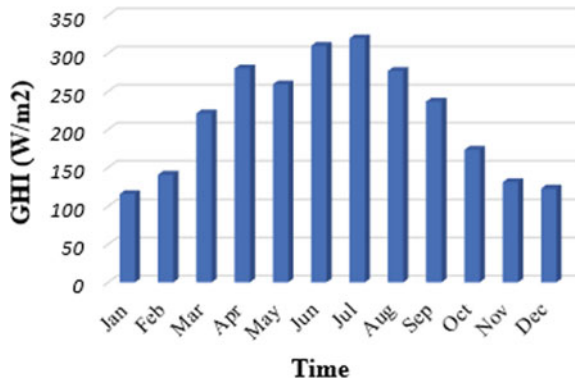
Table 1 Technical and economic data required for the PV/Battery system

Component	Characteristics	
PV array	Type	Mono-crystalline silicon
	Efficiency	15%
	Life time	25 years
	Cost	0.60 \$/m^2
	O&M cost	0.06 \$/m^2
Battery	Battery power	1.2 kW
	Cost	500 \$/kWh
	Replacement cost	0.82 \$/kWh
	Life time	5 years
	η_{CH}	90%
	η_{DCH}	90%
	Dept of discharge	70%
Inverter	Cost	100 \$/kWh
Balance of system	Cost	500 \$/kWh

5 Results and Discussions

5.1 Case study's Results and Discussions

On the basis of the Electric Systems Cascade Extended Analysis, the obtained sizing results of the PV/Battery system powering different wastewater pumping stations are as shown in this section. With reference to Table 2, the surface of PV array required for the first wastewater pumping station (Profile 1) is 33.80 m^2, the battery capacity is 3.05 kWh, and its initial state of charge at time $t = 0$ is 43%. 145.7 m^2 of PV array, a battery capacity of 11.43 kWh charged at 44.4%, and 74.15 m^2 of PV array,

Table 2 Sizing results of PV/Battery system powering wastewater pumping stations

Wastewater pumping station	ESCEA sizing results of PV/Battery system		
	PV array surface (m^2)	Battery's capacity (kWh)	Initial SOC (%)
Profile 1	33.80	3.05	43.00
Profile 2	145.70	11.43	44.42
Profile 3	74.15	6.45	33.84

Fig. 6 Accumulated energy in the battery
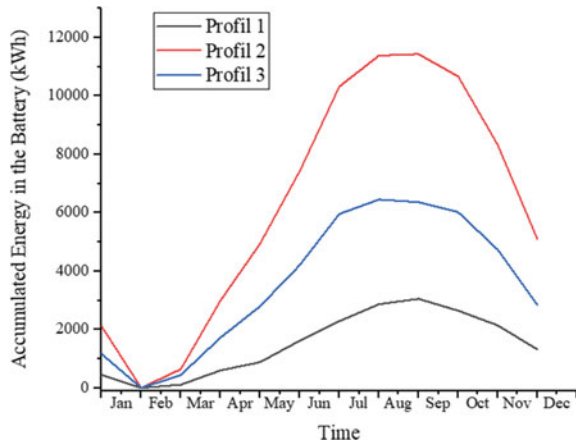
a battery capacity of 6.45 kWh charged at 33.84% are required for the optimal sizing of PV/Battery systems to fully satisfy the electrical needs of wastewater pumping stations 2 and 3 (Profiles 2 and 3), respectively.

To verify the sizing results of PV/Battery systems powering each wastewater pumping station, we will recall the conditions cited in the section that describes the methodology. As we mentioned above, the quantity of energy stored in the battery must be positive at all times and the quantity of energy accumulated in the battery at the end of the time period ($t = T = 8760$ h) should be almost the same as it was at the start ($t = 0$). Figure 6 shows that for the three wastewater pumping stations, the battery of the PV/Battery system respected the conditions of optimal sizing.

5.2 Techno-Economic Analysis

PV/Battery systems sizing results obtained by ESCEA methodology showed very viable signs for the supply of wastewater pumping systems. From Table 3, it is clear that the configurations of the PV/Battery systems are technically impeccable because the LPSP technical evaluation parameter is equal to 0 for the three profiles proposed

Table 3 Techno-economic analysis of the obtained results

Wastewater pumping station	Technical and economic parameters		
	LPSP (%)	LCC ($)	LCOE (¢$/kWh)
Profile 1	0.00	13,877,838	7.16
Profile 2	0.00	52,327,237	6.26
Profile 3	0.00	29,393,052	6.91

in the case study, which means a total and not intermittent satisfaction of the energy needs of stations without needing recourse to the electricity grid.

Also, the economic evaluation based on the Life Cycle Cost and the Levelized Electricity Cost of the results is very encouraging since the configuration of the PV/Battery systems presented minimized energy costs, around 0.06 $/kWh!

6 Conclusion

This study used the Electric Systems Cascade Extended Analysis technique, the LPSP technical assessor, and the LCC and LCOE energy-saving indicators to determine the best size of the different components of PV/Battery systems supplying wastewater pumping stations. The ESCEA technique takes as inputs the hourly wastewater flow of the wastewater pumping stations, the hourly solar irradiation of the case study location, and the technical and economic data of the different components of the PV/Battery system. In order to analyze the capabilities of the ESCEA approach in the sizing process, this research proposes a case study that incorporates three distinct profiles of wastewater pumping stations (different flow rate and total manometric head). The sizing findings obtained indicated extremely promising indicators for supplying wastewater pumping systems with a PV/Battery renewable energy system: technically, by ensuring complete station satisfaction during the investigation period, and economically, by showing very low energy costs, about 0.06 $/kWh.

References

1. Filipe J, Bessa RJ, Reis M, Alves R, Póvoa P (2019) Data-driven predictive energy optimization in a wastewater pumping station. Appl Energy 252:113423
2. Moreno-Rodenas AM, Duinmeijer A, Clemens FHLR (2021) Deep-learning based monitoring of FOG layer dynamics in wastewater pumping stations. Water Res 202:117482
3. Zhang Z, Kusiak A, Zeng Y, Wei X (2016) Modeling and optimization of a wastewater pumping system with data-mining methods. Appl Energy 164:303–311
4. Odeh I, Yohanis YG, Norton B (2006) Influence of pumping head, insolation and PV array size on PV water pumping system performance. Sol Energy 80(1):51–64
5. Mokeddem A, Midoun A, Kadri D, Hiadsi S, Raja IA (2011) Performance of a directly-coupled PV water pumping system. Energy Convers Manag 52(10):3089–3095

6. Chennaif M, Zahboune H, Elhafyani M, Zouggar S (2021) Electric system cascade extended analysis for optimal sizing of an autonomous hybrid CSP/PV/wind system with battery energy storage system and thermal energy storage. Energy 227:120444
7. Chennaif M, Maaouane M, Zahboune H, Elhafyani M (2022) Tri-objective techno-economic sizing optimization of Off-grid and On-grid renewable energy systems using Electric system Cascade Extended analysis and system Advisor Model. Appl Energy 305:117844
8. Chennaif M, Zahboune H, Elhafyani M, Zouggar S (2020) Techno-economic sizing of a stand-alone hybrid energy and storage for water pumping syste. In: Advances in smart technologies applications and case studies. Springer, Cham, pp 291–299
9. Chennaif M, Zahboune H, Elhafyani M, Zouggar S, Blaacha J (2021) Optimal sizing of a concentrated solar power system powering an autonomous water pumping system using the modified electric systems cascade analysis. In: The fourth edition of the international conference on materials and environmental science, pp 7747–7751
10. Chennaif M, Elhafyani ML, Zahboune H, Zouggar S (2020) Electric system cascade analysis for optimal sizing of an autonomous photovoltaic water pumping system. In: Advances in smart technologies applications and case studies. Springer, Cham, pp 282–290
11. Chennaif M, Elhafyani ML, Zahboune H, Zouggar S (2021) The Impact of the tilt angle on the sizing of autonomous photovoltaic systems using electric system cascade analysis. In: Proceedings of the 2nd international conference on electronic engineering and renewable energy systems. Springer, Singapore, pp 767–776
12. Khatib T, Ibrahim IA, Mohamed A (2016) A review on sizing methodologies of photovoltaic array and storage battery in a standalone photovoltaic system. Energy Convers Manag 120:430–448

Comparison of Fixed-Tilt and Tracking PV Plants Coupled to Reverse Osmosis Desalination System: Case of Agadir and Ouarzazate

H. El Mouden, M. Touba, M. Akhsassi, F. Ait Nouh, R. Benbba, M. Jabbour, L. Mandi, and A. Outzourhit

Abstract The water desalination by reverse osmosis (RO) requires a cheap and affordable source of electricity due to the relatively high specific energy consumption of this technology. The implementation of desalination systems working with renewable energies offers a solution to this situation. The aim of this study is to optimize the process of RO desalination by using photovoltaic (PV) solar energy. In this work, we studied three PV systems in two sites: Agadir and Ouarzazate. The compared systems are the Horizontal Single Axis Tracking (HSAT), the Dual-Axis Tracking (DAT), and the fixed plants. The comparison performance is based

H. El Mouden (✉) · M. Akhsassi · R. Benbba · A. Outzourhit
Materials Energy and Environment Laboratory, Faculty of Sciences Semlalia, Cadi Ayyad University, 40000 Marrakech, Morocco
e-mail: hasnae.elmouden@ced.uca.ma

M. Akhsassi
e-mail: m.akhsassi@uiz.ac.ma

R. Benbba
e-mail: rania.benbba@ced.uca.ma

A. Outzourhit
e-mail: aoutzour@uca.ma

H. El Mouden · M. Touba · F. Ait Nouh · L. Mandi
National Centre for Studies and Research On Water and Energy, Cadi Ayyad University, 40000 Marrakech, Morocco
e-mail: mandi@uca.ma

M. Akhsassi
Materials and Renewable Energy Laboratory, Faculty of Sciences, Ibno Zohr University, 80000 Agadir, Morocco

Faculty of Applied Sciences, Ibno Zohr University, 80000 Ait Melloul, Agadir, Morocco

Laboratory of Energy Engineering, Materials and Systems, National School of Applied Sciences, 80000 Agadir, Morocco

M. Jabbour
Materials and Biotechnology Laboratory, Faculty of Sciences, Ibno Zohr University, 80000 Agadir, Morocco

on performance ratio (PR) and levelized cost of energy (LCOE). The different simulation results by PVsyst software have shown that the fixed PV plant gives the best performance in Agadir, with a PR of 82.33% and a LCOE of 0.520 MAD.kWh^{-1}. While, in Ouarzazate, the HSAT system is the most optimal, with a PR of 82.83% and a low LCOE of 0.483 MAD.kWh^{-1}, which yields a desalinated water production cost of 14.24 MAD.m^{-3}.

1 Introduction

Nowadays, the problem of water shortage represents a great obsession, especially for developing countries, because of big and serious factors just like drought, industrial development, population growth, and overexploitation of groundwater and such like. 71% of the earth is covered with water; however, 98% of it has a really high salt concentration to be drinkable or even to be used in irrigation or industry. The need for drinking water in the world is constantly increasing, while underground reserves are constantly decreasing. To remedy this problem, Third-World countries should try to extract fresh water from seawater or brackish water, which is called water desalination. Desalination is becoming a very important solution to the survival of humanity; desalination systems require a large consumption of energy. Also, the increase in fossil fuel price and the increase in the percentage of greenhouse gases in the environment, renewable energies such as solar energy with desalination units, as a combination, will be a better investment. Especially in regions with a very high level of solar irradiation which, at the same time, suffers from the problem of fresh water, such as the Middle East and North Africa (MENA regions).

PV-powered desalination is a promising and non-competitive solution compared to conventional desalination; however, the diminishing of PV costs in recent years has changed many ideas. This document presents the current performance and costs analysis of RO desalination powered by PV systems. The purpose of our work consisted of a comparison between the PV systems in Agadir and Ouarzazate: fixed plan, Horizontal Single-Axis Tracking (HSAT), and Dual-Axis Tracker (DAT) [1].

2 Materials and Methodology

2.1 Geographical Location and Solar Radiation

The two chosen sites are located in Agadir and Ouarzazate Moroccan cities. The latitude, longitude, and altitude of these cities are given in Table 1. These sites are chosen for studies because they had great potential. They had also a good potential of

Table 1 Geographical coordinate details [2]

Cities	Latitude (°)	Longitude (°)	Altitude (m)
Agadir	30° 25′12″ N	9° 35′53″ W	31
Ouarzazate	30° 55′08″ N	6° 53′36″ W	1113

ambient temperature, solar radiation as well as wind resources, which are the main factors in PV system performances. The global and diffuse solar irradiation during the months in Ouarzazate and Agadir are given in Fig. 1.

2.2 Description of the Systems

The different systems under consideration are (see Fig. 2): PV panels, inverter, production counter, and MP20 RO pilot.

Description of the RO Desalination System. The RO MP20 pilot is a dedicated membrane desalination system to desalinate and purify saltwater (Fig. 3). This system's main objective is the elimination and extraction of salts in saltwater and in particular brackish water. Its principle consists in applying brackish water pressure higher than the osmotic pressure to pass the pure water through a semipermeable membrane of the thin-film composite polyamide type from the company SNTE ref TW30-2540 which has the characteristic of retaining dissolved salts in water with a permeability of approximately $2.64 \text{ L.h}^{-1}.\text{m}^{-2}.\text{bar}^{-1}$. In this principle, we necessarily need the electrical energy to power a 220 V high-pressure multistage centrifugal pump from the brand Lowara model 3SV19F022M Multistage Pump 1016L 2401 which supplies a circuit including an RO cartridge (Table 2). The result of the treatment by the pilot is assessed by means of a conductivity probe placed on the permeate circuit. The display associated with the probe makes it possible to work in temperature compensation if desired. Water can be used for drinking, irrigation, domestic, industrial, livestock, or ice production for fishing. The MP20 RO pilot has many advantages; it has low energy consumption and low investment cost compared to thermal processes, high freshwater recovery rate, and also there is the possibility of recovery of energy.

Description of solar PV Park

Selection of PV panels. The different photovoltaic technologies depend on different types of light-absorbing materials, namely silicon (Si), amorphous silicon (a-Si), crystalline silicon (c-Si), cadmium telluride (CdTe), cadmium sulfide (CdS), copper indium di-selenide (CIS), organic and polymer cells, hybrid photovoltaic cell, etc. The efficiency of c-Si cells is more as compared to a-Si for a small amount of material. c-Si cells are further classified as polycrystalline (p-Si) and monocrystalline (m-Si) [3]. m-Si PV panels have greater efficiency than p-Si. Therefore, m-Si PV panels are chosen for this rooftop solar power plant installation. PV panels used here is
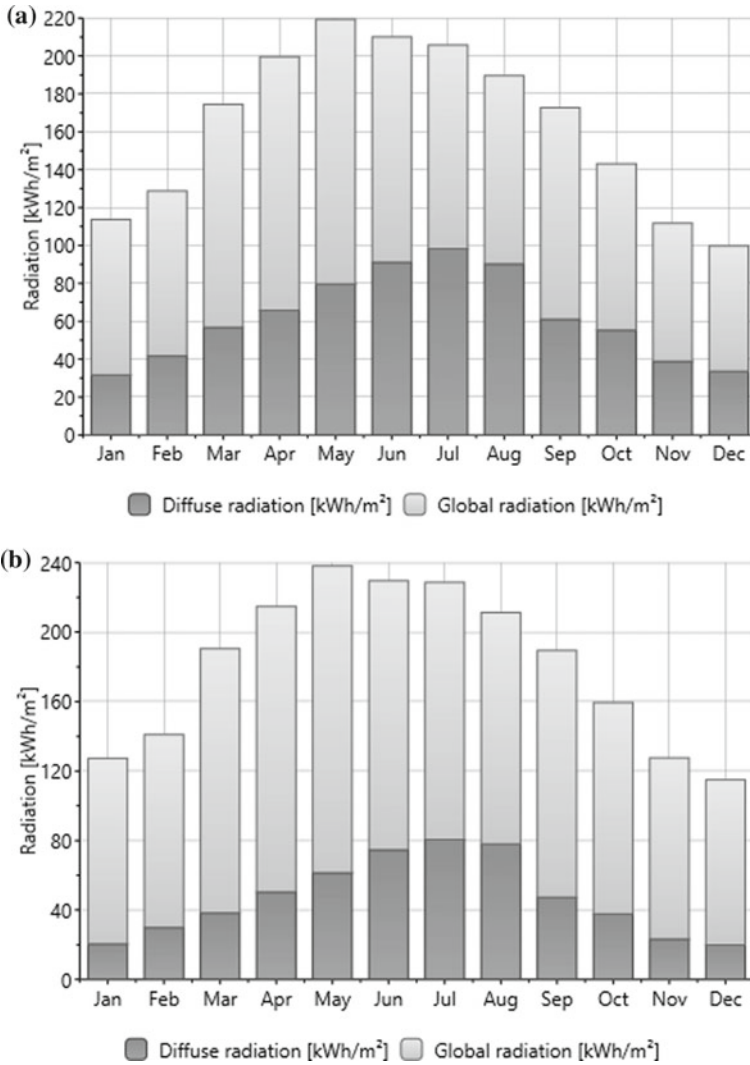
Fig. 1 Global and diffuse solar radiation in (kWh.m^{-2}) generated by Meteonorm.8; **a** in Agadir, **b** in Ouarzazate

of ratings 310 Wp of m-Si solar panels which comprise 60 cells. m-Si solar cells perform better than p-Si and thin film solar cells due to the purity of Si present in m-Si solar cells is high, with energy efficiency (>17% in STC) than the other variants [4]. The PV panels used here has a maximum voltage (Vmpp) = 28.5 V, current at maximum power (Impp) = 9.42 A, maximum module efficiency > 17% at STC, i.e., 1000 W.m^{-2} and 25 °C PV panels are as per the Ministry of New and Renewable
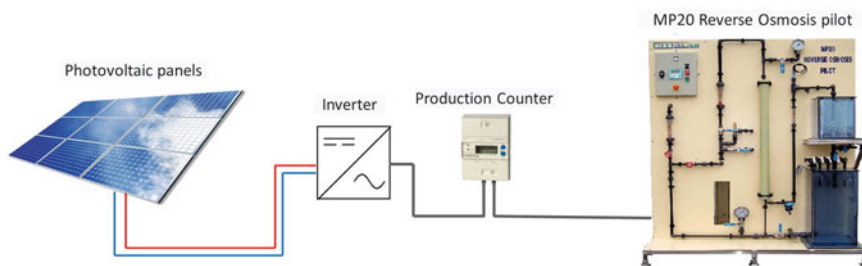
Fig. 2 PV systems coupled with the MP20 RO pilot

Fig. 3 Front and rear face of the MP20 RO pilot

Table 2 Characteristics of RO MP20 pilot

	Voltage amperage	Power type	Electrical power (kW)	Hours of operation per day	Electrical energy (kWh/d)
RO pilot MP20	220/12.5	AC 3.Phase	2.5	10	20

Energy (MNRE) specified IEC 61,215, 61,730, ISO 9001:2000 standard [4]. The properties of the boards used are listed in Table 3.

Table 3 Details of PV modules

PV module/Model	Si-mono/ SEAC60W_310
Manufacturer/Year	Almaden/2019
Module power	310Wp/21 V
Module area	1.645 m^2
Sizing voltage	Vmpp (60 °C) 28.5 V Voc (-10 °C) 44.2 V
Max. operating power at 1000 W.m^{-2} and 50 °C	2.8 kW

Design and selection of inverter. The inverter is designed and selected for the maximum power rating for the system. Since each PV system is designed for 3 kW, the inverter is also selected for 3 kW. The SMA make Sunny Boy 3000 U-240, three-phase inverter is chosen to inject the PV array DC power to the AC grid. The wide input voltage range included between 200 and 400 V.

2.3 PVsyst Simulation Approach

System performance simulations were performed using PVsyst software version 7.1.1. The selection of the simulation tool is based on the fact that PVsyst has a large meteorological database covering a wide range of geographic locations around the world and is also the oldest and most widely recognized software used to simulate solar photovoltaic systems' performance with a high degree of flexibility and reasonably achievable results. It can import weather information from a variety of sources. PVsyst is developed to estimate the performance of autonomous, grid-coupled, and pumping systems according to the type of module specified. The program accurately predicts the system outputs listed using detailed hourly simulation data. PVsyst also has a comprehensive database of solar photovoltaic system components for different modules and inverter manufacturers. NASA satellite and Meteonorm 7.3 station give access to the meteorological data for the installation site. In this study, PVsyst simulation software is used to investigate the performance of the autonomous solar PV plant [5].

3 Results and Discussion

By using the PVsyst software, the monthly values of horizontal global irradiation (GlobHor), diffused irradiation (DiffHor), temperature ($T\_$Amb), wind velocity ($W\_$Velocity), etc. have been described in Tables 4 and 5.

Table 4 Incident energy data by Mateo database in Agadir

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
GlobHor (kWh.m^{-2})	122.0	122.3	185.3	204.5	229.9	212.4	208.3	189.4	170.5	145.7	120.9	107.9
DiffHor (kWh.m^{-2})	25.0	42.9	46.1	61.3	74.1	90.1	97.7	93.8	64.8	52.6	28.8	25.4
T_Amb (°C)	13.5	15.6	18.2	18.3	19.9	21.7	23.9	23.6	22.4	21.5	17.6	15.1
$W\_Velocity$ (m.s^{-1})	3.0	3.4	3.4	3.5	3.5	3.2	3.0	3.0	3.1	3.0	2.9	2.9
Rel_Hum (%)	67.5	61.9	59.3	66.7	67.9	73.0	66.6	72.3	73.2	66.3	65.1	69.6

Table 5 Incident energy data by Mateo database in Ouarzazate

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
GlobHor (kWh.m^{-2})	132.2	136.0	201.6	224.9	250.4	240.8	241.0	225.0	189.7	163.3	130.2	118.9
DiffHor (kWh.m^{-2})	20.9	28.2	28.6	43.3	51.3	65.4	79.0	74.6	46.1	34.9	24.2	19.8
T_Amb (°C)	8.9	11.8	16.4	19.2	23.5	27.9	31.8	30.2	25.5	20.7	13.9	9.8
W_Velocity (m.s^{-1})	2.1	2.4	3.2	3.9	3.8	3.3	3.3	3.1	3.0	2.5	2.3	1.9
Rel_Hum (%)	43.4	38.8	29.5	27.1	23.4	20.3	18.6	21.1	29.8	34.2	41.8	49.9

Table 6 Comparison of the performances for two PV systems in Ouarzazate and Agadir

			Agadir	Ouarzazate
Fixed		LCOE (MAD.kWh^{-1})	0.520	0.490
		Energy produced (kWh/year)	5718	6077
		Performance Ratio PR (%)	*82.32*	*82.35*
Tracking	HSAT	LCOE (MAD.kWh^{-1})	0.525	0.483
		Energy produced (kWh/year)	6949	7547
		Performance Ratio PR (%)	*83.29*	*82.83*
	DAT	LCOE (MAD.kWh^{-1})	0.579	0.511
		Energy produced (kWh/year)	7624	8629
		Performance Ratio PR (%)	*82.20*	*80*

3.1 Comparison Study Using PVsyst

We have studied three PV systems in two sites: Agadir and Ouarzazate. The compared systems are the Horizontal Single-Axis Tracking (HSAT), the Dual-Axis Tracking (DAT), and the fixed PV plants. The results are obtained by using PVsyst software. The comparison performance is based on the performance ratio (PR), annual energy generation, and levelized cost of energy (LCOE). All the results generated through the simulation process in the two sites have been listed in Table 6. The comparison of results has shown that the fixed PV plant yields the best performance in Agadir, with a PR of 82.33% and an LCOE of 0.520 MAD.kWh^{-1}. While, the HSAT system is optimal, with a PR of 82.83% and a low LCOE of 0.483 MAD.kWh^{-1} in Ouarzazate. Hence, the tracking system is optimal in the regions where the DNI is high, as Ouarzazate. These results are close to those obtained by experimental experiments [6]; the small difference is due to the CAPEX (2020/2022), discount and degradation rates, and life-time period (25/30 years).

3.2 Experimental Part

Based on the experimental experiments performed to desalinate brackish water with an electrical conductivity equal to 4050 µS.cm^{-1} using the MP20 RO pilot, we have chosen a supply water flow of 900 L.h^{-1} and a pressure of 14 bars to feed our pilot, we obtain a conversion rate of 55.5% and a retention rate of 97.4%, and these conditions gave us a permeate flow rate of 500 L.h^{-1} and an electrical conductivity around of 105 µS.cm^{-1} (see Table 7).

Table 7 Basic input and
output parameters of the RO
pilot

Parameters	Supply water	Permeate
TDS (g.L^{-1})	2.142	0.054
Electrical conductivity (μS.cm^{-1})	4 050	105
Temperature (°C)	25	25
Flow (L.h^{-1})	900	500
Pressure (bars)	14	14

3.3 Economic Study

This section includes an economic study in which we focus on the comparison of the LCOW of different RO-PV (fixed, HSAT and DAT) and RO-Electricity from the grid installations. RO cost data includes the following [7]:

Direct capital cost (DC). The total investment cost for our RO unit is 55,991.17 MAD.

Plant capacity (_m_). According to the experiments carried out by the pilot of RO MP20, the permeate production is: 5 m^3.d^{-1}, during 10 h of operation.

Specific consumption of electric power (_w_). The permeate production is 0.5 m^3.h^{-1} and the electrical energy of our system equals 2.5 kWh then:

$$w = 5 \, \text{kWh. m}^{-3} \tag{1}$$

Cost of electricity per m3 (_c_). Is the multiplication of w times the cost of electricity used. The prices of electricity in Morocco are not uniform. They depend on the voltage and the consumer; they vary between 1.00 and 1.391 MAD.kWh^{-1} [8]. And for the coupling of the PV-RO, the LCOE is mentioned in Table 6.

Annual fixed charges. This cost defines the annual payments that cover the total direct and indirect costs. It is obtained by multiplying the total direct and indirect costs (DC) by the amortization factor, (_a_).

$$A_1 = a \times \text{DC} \tag{2}$$

With _a_ is defined by the following relation:

$$a = \frac{r(r + 1)^T}{(1 + r)^T - 1} \tag{3}$$

where **_r_** is the yearly interest rate and **_T_** is the lifespan of the plant. Experience from the desalination industry indicates that a 30-year amortization period is adequate. With respect to the interest rate, its average value is 5.

Table 8 Different production costs of desalinated water

	RO-PV			RO-Electricity from the grid	
	LCOW (MAD. kWh^{-1})				
	HSAT	DAT	Fixed	Min	Max
Agadir	15.34	15.47	15.21	27.21	36.99
Ouarzazate	14.24	14.99	14.46		

Annual electric power cost. Is a measure of the full cost of electricity for water production for a given system for a year.

$$A_2 = c \times w \times f \times m \times 365 \tag{4}$$

Total annual cost (A_t). Represents the sum of the two above-mentioned costs A_1 and A_2 which are contributed to the calculation of the unit cost of the product (LCOW).

$$A_t = A_1 + A_2 \tag{5}$$

With: Plant availability: $f = 0, 9$

$$\text{LCOW} = \frac{A_t}{f \times m \times 365} \tag{6}$$

The LCOW is between 27.21 and 36.99 MAD.m^{-3} for the RO connected to the grid; it depends on the electricity tariffs (Table 8). While, it will be decreased until 14.24 and 15.21 MAD.m^{-3}, respectively, for the HSAT system in Ouarzazate and the fixed PV plant in Agadir.

It should be noted that the LCOW for the RO coupled with the HSAT system depends on the desalination plant size [9]. Thus, in order to maximize the water produced, the generation profile of the HSAT system is optimal, despite the high value of LCOW, compared to the fixed PV plant. Therefore, HSAT is an optimal system for the implementation of this desalination technology that promotes attention to national water and energy problems.

4 Conclusion

The document presents an economic analysis of different concepts to associate desalination by the RO MP20 pilot with PV fixed and tracking (HSAT/DAT) in Agadir and Ouarzazate. The performance of this comparison is based on PR, LCOE, and LCOW. The various simulation results of the PVsyst software showed that the fixed PV installation gives the best performance in Agadir, with a PR of 82.33% and a LCOE of 0.520 MAD.kWh^{-1}, giving a production cost of desalinated water of 15.21

MAD.m^{-3}. Whereas in Ouarzazate, the HSAT system is the most optimal, with a PR of 82.83%, and a low LCOE of 0.483 MAD.kWh^{-1}, giving a production cost of desalinated water of 14.24 MAD.m^{-3}. We can conclude that it is better to use the fixed PV installation in Agadir and the HSAT system PV in Ouarzazate for the operation of our pilot since they give a low LCOW cost; all this shows the interest to look for alternative solutions such as coupling to achieve economic and environmental objectives at the same time.

References

1. Abdallah S, Abu-Hilal M, Mohsen MS (2005) Performance of a photovoltaic powered reverse osmosis system under local climatic conditions. Desalination 183(1–3):95–104
2. LatLong homepage, https://www.latlong.net/. Accessed 21 Jan 2022
3. Nogueira CEC, Bedin J, Niedzialkoski RK, de Souza SNM, das Neves JCM (2015) Performance of monocrystalline and polycrystalline solar panels in a water pumping system in Brazil. Renew Sustain Energy Rev 51:1610–1616
4. Shukla AK, Sudhakar K, Baredar P (2016) A comprehensive review on design of building integrated photovoltaic system. Energy Build 128:99–110. Elsevier
5. Kumar R, Rajoria CS, Sharma A, Suhag S (2021) Design and simulation of standalone solar PV system using PVsyst software: a case study. Mater Today Proc 46(11):5322–5328
6. Jabbour M, Darmane Y, Baala A, Akhsassi M, El Mouden H, Chaoui F, Bennouna A et al (2021) Energetic and economic evaluation of mono and poly c-Si grid-connected PV systems: In Ouarzazate, Morocco. In: 2021 9th international renewable and sustainable energy conference (IRSEC). IEEE, pp 1–6
7. El-dessouky HT, Ettouney HM (2002) Fundamentals of salt water desalination, 1st edn. Elsevier, Amsterdam
8. Akhsassi M (2019) Modélisation et optimisation de la production solaire photovoltaïque au Ma-roc dans le cadre de la Stratégie Energétique Nationale. Cadi Ayyad University
9. Monnot M, Carvajal GDM, Laborie S, Cabassud C, Lebrun R (2018) Integrated approach in eco-design strategy for small RO desalination plants powered by photovoltaic energy. Desalination 435:246–258

Evaluation of Energy Use Intensity and Energy Cost of a Residential Building in Morocco Using BIM Approach

Mohamed Maaouane, Mohammed Chennaif, Smail Zouggar, Mohammed El Arabi, Jalal Blaacha, and Mohammed El Fahssi

Abstract Nearly zero energy buildings (nZEBs) have been defined and standardised for certain developed countries. Most developing countries in the North Africa region lack building energy efficiency standards for their hot environments. With increased pressure on buildings to improve their energy and environmental performance, nZEB buildings are likely to gain popularity over the next few years. Their application in the MENA area can significantly reduce building energy consumption and CO_2 emissions. As a result, it is critical to maximise both the energy use intensity (EUI) and the energy cost of a building through Building Information Modelling (BIM) technology for energy analysis. Therefore, this study aims to develop an energy-efficient building by evaluating several design options in terms of EUI and energy cost following ASHRAE 90.1 and Architecture 2030 standards. The baseline scenario's energy consumption intensity was estimated at 400 kWh/m²/year. ASHRAE 90.1 standards' adoption has the potential to save 40% of energy demand. The most energy-efficient architecture was identified to be Architecture 2030 with PV integration (92% of energy savings compared to the baseline scenario). In terms of energy costs, 28.5 \$/m²/year was estimated to satisfy energy demand under the baseline scenario, 11.2 \$/m²/year under ASHRAE standards, and a 3.83 \$/m²/year payoff considering Architecture 2030.

Keywords Building information modelling · Building efficiency · Nearly zero energy buildings · Autodesk revit · Architecture 2030

1 Introduction

Nowadays, as energy crises and global warming continue to increase, energy consumption analysis has become one of the critical elements in building design, gaining prominence as a result [1].

M. Maaouane (✉) · M. Chennaif · S. Zouggar · M. El Arabi · J. Blaacha · M. El Fahssi
School of Technology, Laboratory of Electrical Engineering and Maintenance (LEEM), University Mohammed, BP: 473, 60000 Oujda, Morocco
e-mail: maaouane_mohamed1718@ump.ac.ma

It is vital to create structures with nearly zero energy efficiency and to mitigate global warming by manipulating different elements such as the environment in which a house is located, the building materials utilised, and the configurations [2]. Furthermore, significant design concerns include the amount and cost of energy used to create an environmentally friendly construction with minimum energy consumption that grows over time [3]. As a result, it is critical to implement and adopt techniques to reduce the building's energy use [4].

To analyse energy consumption, new advancements and simulation tools are being created daily to evaluate building performance, one of which is Building Information Modelling (BIM). BIM is a novel approach to project management based on a three-dimensional digital model incorporating reliable and structured data. This technique creates multiple plan options at the application stage of the assignment and structures that may be planned and visualised using BIM and other extension software. Numerous studies on various types of buildings have been conducted using BIM innovation. For instance, research on a building project was conducted to conduct energy analysis using Autodesk Insight and concluded that improving and anticipating building performance throughout its life cycle, as well as the reduction in the use of energy assets, is beneficial when combined with the adaptation of BIM for sustainability analysis [5–7].

To design more energy-efficient homes, various crucial elements, such as orientation, relative humidity, wall and roof materials, and heating, ventilation, and air-conditioning systems, among others, must be considered to determine their impact on the structure's annual energy budget [2]. For example, increasing the area of windows reduces the total annual cooling, heating, and lighting energy used by various building faces, hence reducing energy losses [8]. In addition, natural airflow through a building also contributes significantly to its energy efficiency. Numerous studies have been undertaken in North Africa to investigate and characterise the influence of operational conditions on the energy efficiency of building structures. Certain studies have focussed on the impact of a few design elements by employing simpler analytic techniques [9]. While sustainable buildings have been developed, debated, and implemented in a range of geographical areas, only a few studies have studied the cost-effectiveness of integrating various energy-saving techniques to construct low-energy structures in North Africa [10]. Typically, a non-depth analytical strategy is utilised to analyse a broad array of energy efficiency measures, or EEMs, through the use of sophisticated simulation tools and optimisation techniques [11–13]. Numerous countries in the North African region have made significant efforts over the last decade to improve the energy efficiency of their building stock, including the formulation of new legislation, the implementation of grading systems, and the construction of branding programmes [14]. However, there are no studies in the literature regarding the use of BIM technology in North African countries [15].

This study's aims were as follows:

- Conduct research on energy modelling software.
- Using available BIM tools, develop a verified model for energy use.

- Analyse various design solutions for reducing energy consumption and achieving the most energy-efficient building in terms of energy consumption intensity and cost.
- Calculate the building's energy cost–benefit analysis using the energy cost and the EUI.

2 Methodology

On a national level, energy requirements for buildings fall into two broad categories: normative and performance and both provide the potential for energy efficiency improvements on the path to zero energy buildings. Prescriptive codes establish defined minimum requirements for constructing a structure (e.g., minimum heat reflectivity [*R*-value] for insulation and installation and control requirements for HVAC systems).

ASHRAE 90.1 is the US model national energy code for commercial buildings. It provides minimum energy efficiency standards for new construction, substantial changes, and renovations to high-rise commercial and multifamily residential structures. On the other hand, Architecture 2030 was created in 2002 to lower building-related greenhouse gas emissions. The group made the ZERO code part of its 2030 challenge, which asks for all new construction, development, and significant renovations to be carbon neutral by 2030. The code corresponds to the most recent ASHRAE 90.1-2016 standards for its minimum energy efficiency requirements. In addition, the code permits both on-site and off-site renewable energy generation to get the zero energy label.

To test the accuracy of those recommendations, the energy analysis was based on a case study of a typical Moroccan residential building in the Moroccan City of Oujda. Autodesk Revit Architecture is used to create the 3D model of the commercial building, and Autodesk Insight is utilised to do the energy study. This study focuses primarily on a simplified energy simulation of an existing structure.

To begin, the building's design is modelled in Revit/Autocad using construction parts (walls, floors, roofs, etc.). During model setup, the location and weather data are specified. Next, a separate energy model consisting of surfaces and volumes is constructed directly from the design model. The energy model is automatically generated in Revit and may be visually verified for accuracy before submission to Insight for analysis. The energy parameters allow to specify which aspects should be considered when generating the energy model.

The calculation method in Revit to calculate cooling loads is based on the RTS approach is a two-stage process. To begin, the building's heat gains must be estimated or calculated. These gains include lighting energy, equipment, and human activity, conduction through the building envelope, convection caused by infiltration or ventilation, and solar gains transmitted or absorbed. Equations (1) and (2) summarise the relationship between conductive heat gains and relative magnitude:

$$f_{\mathrm{r}} = \frac{R_{\mathrm{c}}}{R_{\mathrm{r}} + R_{\mathrm{c}}} \tag{1}$$

$$f_{\mathrm{c}} = 1 - f_{\mathrm{r}} \tag{2}$$

with

f_{r}: radiative fraction of conductive heat gain.

f_{c}: convective fraction of conductive heat gain.

R_{r}: radiation film resistance ((m^2 °C)/W).

R_{c}: convection film resistance ((m^2 °C)/W).

The RTS technique evaluates their overall influence on outside surfaces rather than computing specific convection and radiation surface heat balances. Instead, conductive heat transfer is calculated using the sol–air temperature as indicated by Eq. (3):

$$q_\theta = \sum_{\theta=0}^{23} Y_p \left(t_{\mathrm{e},\theta} - t_{\mathrm{rc}} \right) \tag{3}$$

q_θ heat flux for the current hour (W/m^2).
Y_{p} air-to-air periodic response factors (W/(m^2 °C)).
t_{e} sol–air temperature (°C).
t_{rc} constant room temperature (°C).
θ: time step (h).

To analyse and optimise building performance, it is necessary to analyse and adjust the various variables affecting the energy plan (building orientation, window wall ratio, window shades, wall construction, roof construction, lighting efficiency, plug load efficiency, infiltration). Then, the various scenarios created must be compared based on EUI and energy cost.

3 Results and Discussion

To conduct an optimisation analysis and determine the EUI and energy cost of a selected range of energy efficiency measures and photovoltaic system sizes suitable for residential buildings in the city of Oujda, a simulated environment with a variety of parameters is evaluated. To begin, the fundamental characteristics of the residential prototype building used in the optimisation analysis are described. Table 1 summarises the fundamental characteristics of a prototypical single-family home as considered in the optimisation analysis [16]. Figure 1 depicts a three-dimensional view of a typical house (dubbed villa) created using the Revit software. Autodesk

Revit, a comprehensive energy simulation tool for the entire building, was used to conduct an energy analysis and optimisation of the prototype residential building in the city of Oujda.

As previously stated, the building's energy consumption is computed as an energy usage intensity (EUI) in kWh/m^2 per year based on the project's energy specifications. The EUI is determined by dividing the total energy consumed by the building over one year by the building's entire gross area.

In addition, the analysis report provides numerous design alternatives for modifying and controlling the building's energy usage, such as operating schedules, window-to-wall ratios, HVAC systems, building orientation, and lighting efficiency. Finally, the study report provides the project's energy consumption intensity and annual energy cost per square metre.

Table 1 The characteristic of the envelope materials and glazing

Building components	Material (layers)	Thickness (mm)	Thermal conductivity (W/m K)	Density (kg/m^3)	Thermal capacity (W h/kg K)	U-value (W/m^2 K)
Exterior wall	Cement plaster	20	1.153	1700	0.278	3.185
	Hollow brick	70	0.501	720	0.221	
	Hollow brick	70	0.501	720	0.221	
	Cement plaster	20	1.153	1700	0.278	
Floor	Tile	7	0.341	790	0.223	4.258
	Mortar	50	1.153	2000	0.233	
	Concrete	300	1.755	2300	0.256	
Roof	Cement plaster	20	1.153	1700	0.278	5.350
	Concrete block	160	1.09	1300	0.181	
	Concrete	40	1.755	2300	0.256	
Interior wall	Cement plaster	20	1.153	1700	0.278	5.737
	Hollow brick	70	0.501	720	0.221	
	Cement plaster	20	1.153	1700	0.278	
Glazing	*Thickness (mm)*	*U-value (W/m^2 K)*	*SHGC*	*Solar transmission*	*Solar reflectance*	*Visual transmission*
Single glazing	2.5	5.74	0.87	0.85	0.075	0.901
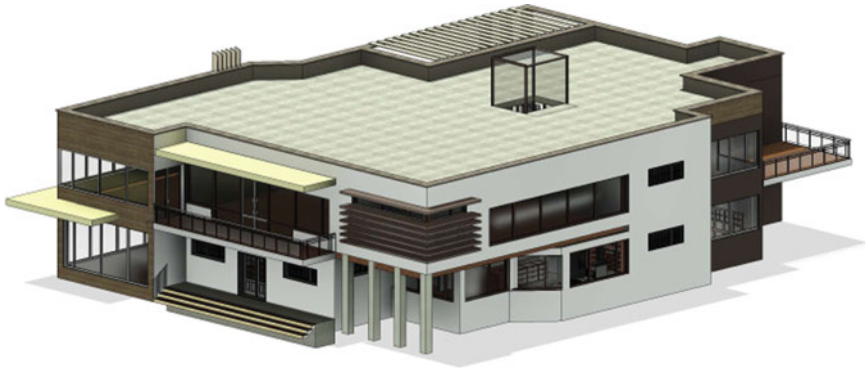
Fig. 1 Building modelling 3D view

This model is created using the data and characteristics of the model's constituent elements (walls, floors, doors, windows, etc.); it is based on the volume of the rooms; these spaces enable us to configure the areas in which the HVAC systems operate. Following the generation of the energy model, a simulation of its performance is available via the cloud. Using the energy model's data and the project's location, Revit enables specifying the exact location of the project and then selecting the nearest meteorological site to use its data during the simulation. This simulation site considers solar radiation, exterior temperature, the building's location, the building's operating hours (24/7, 12/5), and the building's typology (residential, office, school, or hotel). This is achievable as a result of Insight's integration.

As illustrated in the Figs. 2, 3 and 4, a comparison of three scenarios (Baseline, ASHRAE 90.1, and Architecture 2030) to ASHRAE 90.1 and ARCH 2030 standards is generated.

The figure compares the energy use and cost of energy in each scenario using the following benchmarks: The intensity of energy consumption under the baseline scenario before the design characteristics were modified was 28.5 \$/m^2/year; under
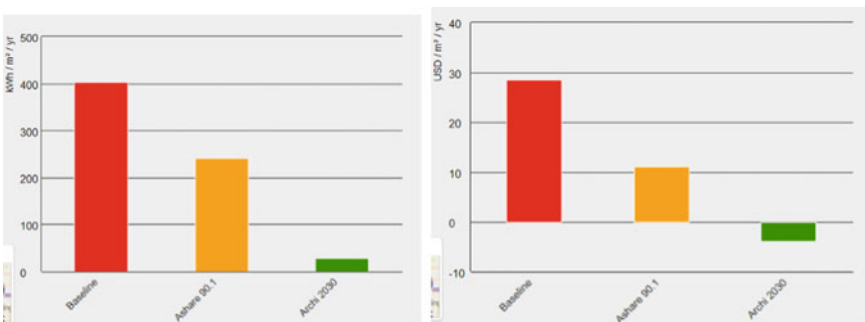
Fig. 2 Scenario's comparison Kwh/m^2/year and Usd/m^2/year
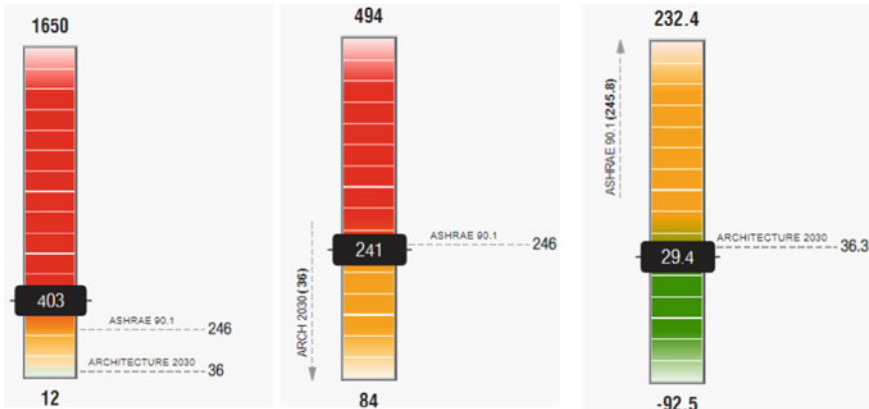
Fig. 3 Benchmark comparison Kwh/m^2/year (baseline, Ashrae90.1, Archi 2030)

Fig. 4 Benchmark comparison Usd/m^2/year (baseline, Ashrae90.1, Archi 2030)

the ASHRAE 90.1 scenario, the energy consumption was 241 Kwh/m^2/year, and the cost of energy was 11.2 \$/m^2/year; and under the Architecture 2030 with PV integration, the energy consumption was 29.4 Kwh/m^2/year, and the cost of energy was − 3.83 \$/m^2/year.

It can be seen that the building's EUI value decreased from 403 to 29.4 kWh/m^2/year after several design criteria were changed in the Insight platform. By adjusting each design criteria in the Insight analysis, the building's energy intensity and cost can be restored to the appropriate levels. Additionally, Autodesk Insight enables one to compare two energy models with varying shapes, orientations, and energy factors to choose the most appropriate building model for the lowest energy consumption and cost. Table 2 summarises the different suggestions to reach the requirements for each of the two scenarios (ASHRAE and Archi 2030).

Table 2 Summary of building design options

	Baseline	ASHRAE 90.1	Architechture 2030
WWR-eastern walls (%)	30	15	0
WWR-western walls (%)	30	15	7
WWR-southern walls (%)	30	30	30
WWR-nothern walls (%)	30	30	30
Roof construction	No insulation	R15	R60
Wall construction	No insulation	R13 wood	R38 wood
Plug load efficiency (W/m^2)	28	13.99	6.36
Lighting efficiency (W/m^2)	20.45	11.84	7.53
HVAC	Heat pump	ASHRAE VAV	High efficiency VAV
PV-surface coverage (%)	0	60	90

4 Conclusion

The energy analysis was based on a case study of a typical Moroccan residential structure in Oujda, Morocco. The commercial building's 3D model was created using Autodesk Revit Architecture, and the energy assessment was conducted using Autodesk Insight. The purpose of this study was to present a simplified energy simulation of an existing structure. It was vital to maximise both the energy use intensity (EUI) and the energy cost of a building through energy analysis using Building Information Modelling (BIM) technology. As such, the goal of this study was to produce an energy-efficient building by assessing numerous design options for EUI and energy cost following ASHRAE 90.1 and Architecture 2030 requirements.

The study evaluated alternative design solutions for energy conservation and achieving the most energy-efficient building regarding energy consumption intensity and cost. Additionally, the energy cost–benefit analysis of the building was performed using the energy cost and the EUI.

Significant primary energy consumption reductions have been achieved by reinforcing the building structure and adding solar photovoltaic (PV) energy into the basic model. Additionally, additional architectural innovations like improved thermal insulation, higher lighting efficiency, and photovoltaic (PV) panels have aided in developing positive energy buildings that generate more energy than they consume.

Future research will concentrate on the approach and cost–benefit analysis of net-zero energy residential building design in order to optimise residential building design in multiple locations throughout the MENA region in order to minimise life-cycle energy costs through the use of a variety of energy efficiency measures.

References

1. Ali Shoshan AA, Haseeb QS (2019) Analysis and reduction of energy consumption in educational buildings: a case study in the college of education for pure sciences at the University of kir-kuk. Int J Civ Eng Technol
2. Pathan G, Borkar MM, Bhimte KA, Ghorse VR, Ambalkar PB (2019) Comparative study of energy analysis on residen-tial building in Revit software. Int J Adv Res Innov Ideas Educ 2:1182–1189
3. Abdullah A, Cross B, Aksamija A, Will Perkins (2014) Whole building energy analysis: a comparative study of different simulation tools and applications in architectural design. Springer, New York, pp 1–12
4. Shaeri J, Habibi A, Yaghoubi M, Chokhachian A (2019) The optimum window-to-wall ratio in office buildings for hot-humid, hot-dry, and cold climates in Iran. Environ MDPI
5. Jangalve A, Kamble V, Gawandi S, Ramani N (2017) Energy analysis of residential building using BIM. Int J Emerg Eng Technol Sci 108:15–19
6. Yarramsetty S, Rohullah MS, Sivakumar MVN, Anand Raj P (2020) An investigation on energy consumption in residential building with different orientation: a BIM approach. Asian J Civ Eng 21(2):253–266
7. Lung L (2019) Using BIM to simulate the energy consumption and educe its cost. Int J Innov Manage Technol 10:21–26
8. Moakher PE, Pimplikar SS (2012) Building information modeling (BIM) and sustainability—using design technology in energy efficient modeling. IOSR J Mech Civ Eng 1(2):10–21
9. Tushar Q, Bhuiyan MA, Zhang G, Maqsood T (2021) An integrated approach of BIM-enabled LCA and energy simulation: the optimized solution towards sustainable development. J Clean Prod 289:125622. https://doi.org/10.1016/j.jclepro.2020.125622
10. Krarti M, Ihm P (2016) Evaluation of net-zero energy residential buildings in the MENA region. Sustain Cities Soc 22:116–125. https://doi.org/10.1016/j.scs.2016.02.007
11. El Sayary S, Omar O (2021) Designing a BIM energy-consumption template to calculate and achieve a net-zero-energy house. Sol Energy 216:315–320. https://doi.org/10.1016/j.solener.2021.01.003
12. Al-Saeed YW, Ahmed A (2018) Evaluating design strategies for nearly zero energy buildings in the middle east and North Africa regions. Designs 2(4):1–12. https://doi.org/10.3390/designs2040035
13. Abdou N, El Mghouchi Y, Hamdaoui S, El Asri N, Mouqallid M (2021) Multi-objective optimization of passive energy efficiency measures for net-zero energy building in Morocco. Build Environ 204:108141. https://doi.org/10.1016/j.buildenv.2021.108141
14. Sick F, Schade S, Mourtada A, Uh D, Grausam M (2014) Dynamic building simulations for the establishment of a Moroccan thermal regulation for buildings. J Green Build 9(1):145–165
15. Nägeli C, Jakob M, Catenazzi G, Ostermeyer Y (2020) Towards agent-based building stock modeling: bottom-up modeling of long-term stock dynamics affecting the energy and climate impact of building stocks. Energy Build 211. https://doi.org/10.1016/j.enbuild.2020.109763
16. Romani Z, Draoui A, Allard F (2015) Metamodeling the heating and cooling energy needs and simultaneous building envelope optimisation for low energy building design in Morocco. Energy Build 102:139–148

Numerical Modeling of Partial Dam-Break Over Mobile Bed: Very Fine Sand Case

Sanae Jelti and Abdelhafid Serghini

Abstract This work deals with the numerical modeling of dam-break problem over mobile bed. The governing equations are a combination of the coupled model and the non-capacity model. The mathematical model is solved by the finite-volume Roe scheme, associated with a new discretization of the source term. Attention is given to the behavior of the water flow and the bed rate change. The numerical scheme is applied on bottoms composed of very fine sand.

Keywords Roe scheme · Sediment transport · Unstructured grids · Mobile bed · Very fine sand

1 Introduction

Natural rivers are the center of interest of many researchers because it is linked to several domain like water resources management, hydropower production, and flood risks among others. The problem of dam-break is very studied in the literature; various mathematical models were proposed to modelize the problem. In this work, we applied the mathematical model proposed by Cao et al. [1], which link the water flow, the bed rate change, and the sediments transport.

In this work, we resolve numerically the dam-break problem over erodible bed, specially using very fine sand of which the goal is the discovery of a new behavior of the flow and the bed, in two-dimensional case using unstructured grids. The mathematical model consists of five equations; the shallow-water of the water-sediment mixture, the momentum conservation equation of the water-sediment mixture in the x and y directions, respectively, the transport diffusion equation of sediment particles and bed rate change equation. Roe scheme, introduced in [2], associated to the new

S. Jelti (✉)

M and E Laboratory, Faculty of science, Mohammed First University, Oujda, Morocco
e-mail: s.jelti@ump.ac.ma

A. Serghini

ANAA Research Team, ESTO, LANO Laboratory, FSO, Mohammed First University, Oujda, Morocco

discretization which satisfies the C-property is used to solve numerically the governing equations developed in [3]. The MUSCL method with generalized minmod limiter and the Runge–Kutta method is used in order to achieve a second-order accuracy in space and time, respectively. A local mesh refinement is implemented using sediment concentration as a monitoring function to achieve higher level of accuracy and gain on the computational cost.

Attention is given to the behavior of the flow and the bed rate change after dam-break. The numerical scheme is applied on partial dam-break over erodible bed. The obtained results are compared to existing similar works.

This paper is organized as follows. In Sect. 2, we introduce the governing equations. In Sect. 3, we present the numerical scheme. Discretization of the source term is given in Sect. 4. Numerical results are presented in Sect. 5. Finally, concluding remarks are summarized in Sect. 6.

2   Governing Equations

We study in this work the dam-break problem in open-channel hydraulic with rectangular cross-section of constant width, over a movable bed composed of uniform and noncohesive sediment particles. In this case, we treat specially bottoms containing very fine sand.

The considered physical problem is modeled by the system of equations presented below. The mathematical model is composed of the Saint–Venant equations of water-sediment mixture, the transport diffusion equation of sediments particles, and the bed rate change equation. This new concept was introduced in [1, 4] and resolved differently in different works as [5–7]. The governing equation are given by

$$\frac{\partial h}{\partial t} + \frac{\partial (hu)}{\partial x} + \frac{\partial (hv)}{\partial y} = \frac{E - D}{1 - p} \tag{1}$$

$$\frac{\partial (hu)}{\partial t} + \frac{\partial (hu^2 + \frac{1}{2}gh^2)}{\partial x} + \frac{\partial (huv)}{\partial y} = B_x \tag{2}$$

$$\frac{\partial (hv)}{\partial t} + \frac{\partial (huv)}{\partial x} + \frac{\partial (hv^2 + \frac{1}{2}gh^2)}{\partial y} = B_y \tag{3}$$

$$\frac{\partial (hc)}{\partial t} + \frac{\partial (huc)}{\partial x} + \frac{\partial (hvc)}{\partial y} = E - D \tag{4}$$

$$\frac{\partial z}{\partial t} = -\frac{E - D}{1 - p} \tag{5}$$

where B_x and B_y are the source terms defined by:

$$B_x = -gh\frac{\partial z}{\partial x} - \frac{\rho_s - \rho_w}{2\rho}gh^2\frac{\partial c}{\partial x} - ghS_{f_x} - \frac{\rho_0 - \rho}{\rho}\frac{E - D}{1 - p}u \qquad (6)$$

$$B_y = -gh\frac{\partial z}{\partial y} - \frac{\rho_s - \rho_w}{2\rho}gh^2\frac{\partial c}{\partial y} - ghS_{f_y} - \frac{\rho_0 - \rho}{\rho}\frac{E - D}{1 - p}v \qquad (7)$$

t is the time; x and y are horizontal coordinates; h is the flow depth; u and v are depth-averaged velocities in the x-, y-directions, respectively; z is the bed elevation; c is the flux-averaged volumetric sediment concentration; g is the gravitational acceleration; p is the bed sediment porosity. D and E are the sediment deposition and entrainment fluxes across the bottom boundary of flow; they represent the exchange between water column and bed. S_{fx} and S_{fy} are the friction slopes in the x and y directions, respectively; $\rho = \rho_w(1 - c) + \rho_s c$ is the density of water-sediment mixture; $\rho_0 = \rho_w p + \rho_s(1 - p)$ is the density of the saturated bed; ρ_w and ρ_s are the densities of water and sediment, respectively. We apply in this work the same empirical functions as [3].

3   Numerical Scheme

The mathematical model described earlier (1–5) can be arranged in the following conservative form:

$$\frac{\partial U}{\partial t} + \frac{\partial F(U)}{\partial x} + \frac{\partial G(U)}{\partial y} = S(U) + Q(U) \qquad (8)$$

where

$$U = \begin{pmatrix} h \\ hu \\ hv \\ hc \\ z \end{pmatrix}, \quad F = \begin{pmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \\ huv \\ huc \\ 0 \end{pmatrix}, \quad G = \begin{pmatrix} hv \\ huv \\ hv^2 + \frac{1}{2}gh^2 \\ hvc \\ 0 \end{pmatrix},$$

$$S = \begin{pmatrix} 0 \\ -gh\frac{\partial z}{\partial x} - \frac{(\rho_s - \rho_w)}{2\rho}gh^2\frac{\partial c}{\partial x} \\ -gh\frac{\partial z}{\partial y} - \frac{(\rho_s - \rho_w)}{2\rho}gh^2\frac{\partial c}{\partial y} \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} \frac{E - D}{1 - p} \\ -ghS_{f_x} - \frac{\rho_0 - \rho}{\rho}\frac{E - D}{1 - p}u \\ -ghS_{f_y} - \frac{\rho_0 - \rho}{\rho}\frac{E - D}{1 - p}v \\ E - D \\ -\frac{E - D}{1 - p} \end{pmatrix}$$

3.1 Finite-Volume Roe Scheme

The finite-volume Roe scheme is used to discretize Eq. (8) on unstructured triangular grids. The time interval is divided into sub-intervals $[t_n, t_{n+1}]$ with stepsize Δt, and the spatial domain is discretized in conforming triangular elements T_i. Each triangle represents a control volume, and the variables are located at the geometric centers of the cells. Hence, a finite-volume discretization of (8) yields

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{|T_i|} \sum_{j \in N(i)} \int_{\Gamma_{ij}} \mathcal{F}(U^n, \overrightarrow{\eta_{ij}}) d\Gamma + \frac{\Delta t}{|T_i|} \int_{Ti} S(U^n) d\Omega$$
$$+ \frac{\Delta t}{|T_i|} \int_{Ti} Q(U^n) d\Omega \tag{9}$$

where

$$\mathcal{F}(U^n, \overrightarrow{\eta_{ij}}) = F(U^n)\eta_x + G(U^n)\eta_y$$

$N(i)$ is the set of neighboring triangles of the cell T_i, U_i^n is an averaged value of the solution U in the cell T_i at time t_n:

$$U_i^n = \frac{1}{|T_i|} \int_{T_i} U^n d\Omega,$$

$|T_i|$ denote the area of T_i, Γ_{ij} is the interface between two control volumes T_i and T_j and $\overrightarrow{\eta_{ij}} = (n_x, n_y)^T$ denote the unit outward normal to Γ_{ij}.

The numerical flux is defined by Φ_{ij} such as:

$$\int_{\Gamma_{ij}} \mathcal{F}(U^n, \overrightarrow{\eta_{ij}}) d\Gamma = \Phi_{ij} |\Gamma_{ij}| \tag{10}$$

Equation (9) becomes:

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{|T_i|} \sum_{j \in N(i)} \Phi_{ij} |\Gamma_{ij}|$$
$$+ \frac{\Delta t}{|T_i|} \int_{T_i} S(U^n) d\Omega + \frac{\Delta t}{|T_i|} \int_{T_i} Q(U^n) d\Omega, \tag{11}$$

Using Roe scheme, the numerical flux is defined as

$$\Phi_{ij}(U_{ij}^L, U_{ij}^R) = \frac{1}{2}(\mathcal{F}(U_{ij}^L) + \mathcal{F}(U_{ij}^R)) - \frac{1}{2} \left| \mathcal{A}(\widetilde{U}_{ij}^n) \right| (U_{ij}^R - U_{ij}^L), \tag{12}$$

U_{ij}^R, U_{ij}^L are the right and left approximations of the solution U at the interface Γ_{ij} at time t_n. $\widetilde{U}_{ij}^n$ is Roe average and $\mathcal{A}_\eta(\mathcal{U})$ is Jacobian matrix of the system in use given in [3].

We refer the reader to the detailed calculation reported in [3].

To develop a second-order finite volume, we use a monotone upstream-centered scheme for conservation laws method MUSCL incorporating a slope limiters in the spatial approximation. Two-step Runge–Kutta TVD method is implemented for time integrating introduced in [6]. Concerning the boundary conditions, for open inflow and outflow, the flow variables on the boundary are set to the same values as at the interior of the flow. For solid walls, the value of the flow variables is simply the mirror image of that at the associated boundary point so that the normal velocity component is zero at the boundary.

4   Discretization of the Source Term

In this study, we implement the new discretization of the source term satisfying the conservation property developed in [3]. The numerical scheme in use satisfies the C-property if we approach the source term $\int_{T_i} S(U^n)\mathrm{d}\Omega$ given in Eq. (11) as following

$$\int_{T_i} S(U^n)\mathrm{d}\,\Omega = \frac{1}{2}(\mathcal{S}_i^R + \mathcal{S}_i^L)$$

where $\mathcal{S}_i^R$ and $\mathcal{S}_i^L$ are the right and left approximations of $\int_{T_i} S(U^n)\mathrm{d}\,\Omega$

$$\mathcal{S}_i^R = \begin{pmatrix} 0 \\ I_{xi}^R - \frac{\rho_0 - \rho_w}{2\rho} g \, (h_i^R)^2 \displaystyle\sum_{j \in N(i)} c_{ij}^R \, n_{xij} |\Gamma_{ij}| \\ I_{yi}^R - \frac{\rho_0 - \rho_w}{2\rho} g \, (h_i^R)^2 \displaystyle\sum_{j \in N(i)} c_{ij}^R \, n_{yij} |\Gamma_{ij}| \\ 0 \\ 0 \end{pmatrix}$$

and $\mathcal{S}_i^L$ is obtained by replacing in the expression of $\mathcal{S}_i^R$ the data right by the associated data left. The proposed decomposition of the source term $\int_{T_i} Q \, \mathrm{d}\Omega$ given in Eq. (11) is given as:

$$\int_{T_i} Q \, \mathrm{d}\Omega = \frac{1}{2}(\mathcal{Q}_i^R + \mathcal{Q}_i^L)$$

where $\mathcal{Q}_i^R$ and $\mathcal{Q}_i^L$ are the right and left approximations of the term $\int_{T_i} Q(U^n)\mathrm{d}\,\Omega$

$$Q_i^R = |T_i| \cdot \begin{pmatrix} \frac{E_i^R - D_i^R}{1-p} \\ -gh_i^R S_{fx,i}^R - \frac{\rho_0 - \rho}{\rho} \frac{E_i^R - D_i^R}{1-p} u_i^R \\ -gh_i^R S_{fy,i}^R - \frac{\rho_0 - \rho}{\rho} \frac{E_i^R - D_i^R}{1-p} v_i^R \\ E_i^R - D_i^R \\ -\frac{E_i^R - D_i^R}{1-p} \end{pmatrix}$$

Q_i^L is obtained by replacing in the expression of Q_i^R the data right by the associated data left.

5 Numerical Results

The partial dam-break problem is widely used in the literature as [3, 8] among others, except that it is never used over soil composed of very fine sand. Special attention is given to the behavior of the flow and bed rate change.

5.1 Partial Dam-Break: Bottom Composed of Fine Sand

In this part of study, we consider a square reservoir with a flat bottom, length, and width are equal to 200 m. The dam is 4 m thick, and the breach is assumed to be 75 m. In the present study, only, fine sand sediments are considered, precisely $d = 0.1$ mm, $d = 0.2$ mm and $d = 0.3$ mm. The initial conditions are given by

$$u(x, y, 0) = v(x, y, 0) = 0 \, \text{m/s}$$

$$c(x, y, 0) = \begin{cases} 0.01, & \text{if } x \le 100 \, \text{m} \\ 0, & \text{otherwise} \end{cases}$$

In order to study the behavior of the bed after dam-break and the performance of the numerical scheme with the proposed discretization, we supplied the bottom by three bumps defined as

$$Z(x, y, 0) = 10 \exp(-((xx - 2.5)^2 + (yy - 7.5)^2))$$
$$+ 15 \exp(-((xx - 5)^2 + (yy - 7.5)^2))$$
$$+ 12 \exp(-((xx - 5)^2 + (yy - 5)^2))$$
$$Z(x, y, 0) = Z(x, y, 0)/20$$

where $xx = \frac{x-100}{10}$ and $yy = \frac{y}{20}$. The flow depth is given by

$$h(x, y, 0) = 1 - Z(x, y, 0).$$
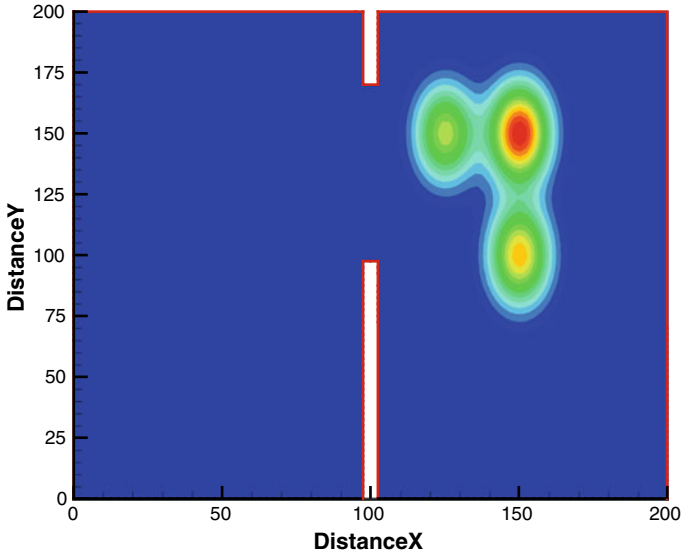
Fig. 1 Initial condition

Notice: All figures represent a cross-section on $y = 125$.

Figure 1 presents the initial conditions. We remark in Fig. 2 that the water-free surface sustained a depression marked by a stairs when the sediments are coarser. Sediment size has also a great effect on the bed rate change, more the sediment is finer more significant is the erosion. This is the reason why the higher concentration back to the finer sediments. We also remark that the wavefront undergoes a lateral expansion which is wider when the sediment is fine. The appearance of the profiles is in agreement with the results obtained in [3, 5, 7–10].
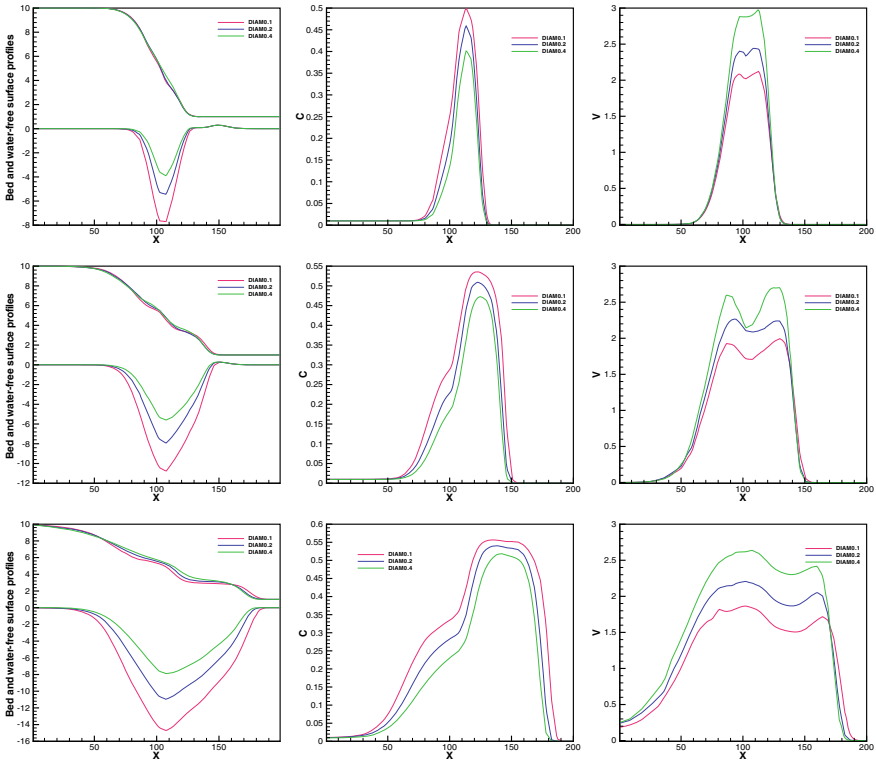
Fig. 2 Water free surface, bed, and concentration profiles at different times, for different sizes of sediment (from top to bottom $t = 2, 4, 6, 8$ s)
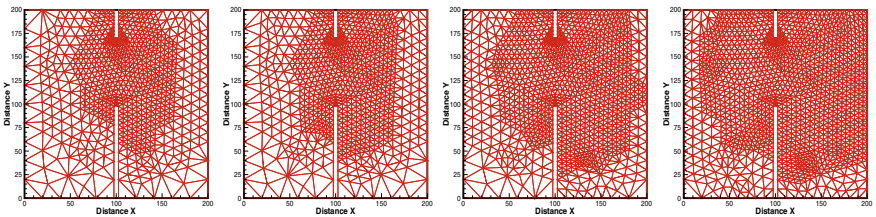
Fig. 3 Mesh adaptation at $t = 2, 4, 6, 8$ s min using fine sand of $d = 0.1$ mm

Comparing this results to those obtained in [3] of partial dam-break aver a bottom composed of medium sand, we remark that the bed rate change is more pronounced in this case which influence the concentration profiles, the concentration reached higher level when the soil is composed of fine sand.

Concerning the velocity, we remark in Fig. 2 that the velocity increases progressively when the sediment is coarser. The higher velocity back to the coarse sediments the same remark reported in works [1, 3, 5]. The numerical scheme is already applied

on medium and coarse sediments in [3]. We remark the same bed behavior, but the extent of bed rate change is relative to the sediment size and the velocity in use. This is in agreement with the results obtained in the benchmark test reported in [11].

Figure 3 represents the mesh adaptation in use; we observe that the refined surface follows the gradient concentration. After dam-break, we observe that the refinement was limited on the initial dam-break; then, it propagates on all the studied surface. We used a dynamic mesh to gain in computational time also in precision.

6 Conclusion

The present study is a numerical modeling of dam-break problem over mobile bed. The physical problem is modeled by the coupled model and the non-capacity model. Roe scheme with a discretization of the source term is used to solve the mathematical model. We focused in the behavior of the flow and the bed rate change after dam-break in the case of the finest sediments. Through the obtained results and comparing to other works, the numerical scheme in use presents a satisfying performance.

References

1. Cao Z, Pender G, Wallis S, Carling PA (2004) Computational dam-break hydraulics over erodible sediment bed. J Hydraul Eng 130:389–703
2. Roe PL (1981) Approximate Riemann solvers, parameter vectors and difference schemes. J Comput Phys 43:357–372
3. Jelti S, Boulerhcha M (2022) Numerical modeling of two dimensional non-capacity model for sediment transport by an unstructured finite volume method with a new discretization of the source term. Math Comput Simul 197(3):253–276
4. Cao Z, Day R, Egashira S (2002) Coupled and decoupled numerical modelling of flow and morphological evolution in alluvial rivers. J Hydraul Eng 128:306–321
5. Simpson G, Castelltort S (2006) Coupled model of surface water flow, sediment transport and morphological evolution. Comput Geosci 32:1600–1614
6. Gottlieb S, Chi-Wang S (1998) Total variation diminishing Runge-Kutta schemes. Math Comput 67:73–85
7. Yue Z, Cao Z, Li X, Che T (2008) Two-dimensional coupled mathematical modeling of fluvial processes with intense sediment transport and rapid bed evolution. Sci China Ser G Phys Mech Astron 51(9):1427–1438
8. Benkhaldoun F, Elmahi I, Seaid M (2013) An unstructured finite-volume method for coupled models of suspended sediment and bed load transport in shallow-waterflows. Comput Methods Appl Mech Eng 199:49–52
9. Capart H, Young DL (1998) Formation of a jum by the dam-break wave over a granular bed. J Fluid Mech 372:165–187
10. Jelti S, Mezouari M, Boulerhcha M (2018) Numerical modeling of dam-break flow over erodible bed by Roe scheme with an original discretization of source term. Int J Fluid Mech Res 45:21–36

11. Soares-Frazao S, Canelas R, Cao Z, Cea L, Chaudhry HM, Die Moran A, El Kadi K, Ferreira R, Cadorniga IF, Gonzalez-Ramirez N, Greco M, Huang W, Imran J, Le Co J, Marsooli R, Paquier A, Pender G, Pontillo M, Puertas J, Spinewine B, Swartenbroekx C, Tsubaki R, Villaret C, Wu W, Yue Z, Zech Y (2012) Dam-break flows over mobile beds: experiments and benchmark tests for numerical models. J Hydraul Res 50(4):364–375

Study Effect of Nanofluids on the Performance Enhancement of PV/T Collector

Safae Margoum, Chaimae El Fouas, Hajji Bekkay, Stefano Aneli, Antonio Gagliano, Giovanni Mannino, and Giuseppe M. Tina

Abstract Compared to the photovoltaic (PV) module, which only produces electricity, the photovoltaic/thermal (PV/T) system presents as a promising technology capable of simultaneously producing electrical and thermal energy. In addition, this technology considerably improves the photovoltaic efficiency by cooling the PV cells sensitive to the increase in temperature. This paper investigates the impact of various nanofluids on improving the performance of PV/T systems. Three different fluids, pure water, Al_2O_3/water, and Cu/water are studied as a coolant to reduce the temperature of the PV panel. The performances of the nanofluid-based PV/T system were simulated using a numerical model based on the energy balance equation, developed in MATLAB software. By analyzing the simulation results, it is observed that the use of Al_2O_3/water and Cu/water nanofluids improves the results of the PV/T

S. Margoum (✉) · C. El Fouas · H. Bekkay
Laboratory of Renewable Energy, Embedded System and Information Processing, National School of Applied Sciences, Mohammed First University, 60000 Oujda, Morocco
e-mail: safae.margoum@ump.ac.ma

C. El Fouas
e-mail: c.elfouas@ump.ac.ma

H. Bekkay
e-mail: hajji.bekkay@gmail.com

S. Aneli · A. Gagliano · G. M. Tina
University of Catania, Viale Andrea Doria 6, 95125 Catania, Italy
e-mail: stefano.aneli@unict.it

A. Gagliano
e-mail: agagliano@dii.unict.it

G. M. Tina
e-mail: giuseppe.tina@unict.it

G. Mannino
CNR–IMM, Strada VIII N. 5 Zona Industriale, Catania, CT, Italy
e-mail: giovanni.mannino@imm.cnr.it

system in terms of electrical and thermal efficiency compared to the pure water-based PV/T system. In particular, dispersing a 2% volume fraction of copper and alumina nanoparticles in water helps to increase the thermal and electrical efficiency of the PV/T by 26 and 1.24% for Cu/water and 10.33 and 0.99% for Al_2O_3/water, respectively.

Keywords PVT collector · Nanofluids · Electrical · Thermal efficiency

1 Introduction

In recent years, solar energy has been widely used for thermal energy generation by solar collectors and electrical energy by photovoltaic (PV) panels [1]. A novel innovative technology has been invented, able to produce simultaneously electrical and thermal energies in a single unit is known as a hybrid solar system (PV/T) [2, 3]. Due to different criteria and parameters, many types of (PV/T) systems can be identified. Regarding the cooling fluid, photovoltaic thermal systems (PVT) can be divided mainly into three types: water-based PVT, air-based PVT, and bi-fluid-based PV/T using both fluids simultaneously air and water [4]. Recently, a new generation of cooling fluids known as nanofluids is investigated as well as applied in the PV/T system. As a result, numerous research works have been conducted in order to predict the electrical and thermal behavior of this type of PV/T systems adopting nanofluids for the cooling effect. From literature, it is proved that nanofluids have better thermo-physical proprieties than basic fluids, which can potentially improve the PV/T system performances compared to the other systems (PV/T using air or water) [5]. An experimental study is conducted by [6], where the effect of the silica/water nanofluid (1 and 3 wt%) as a coolant on the thermal and electrical efficiencies of a PV/T sheet & tubes system is studied, based on the first and second thermodynamic laws. As a result, the electrical and thermal efficiencies were improved by 18.9% and 7.93%, respectively, by dispersing 3 wt% of SiO_2 in water. In the work performed by [7], the impact of applying 0.05% volume concentration of CuO/water nanofluid in a PV/T collector with and without glazing is investigated experimentally. It is concluded that under identical conditions, the electrical efficiency using nanofluid is inferior by 13%, while thermal efficiency is higher by 45% compared to water. In another study [8], a two-dimensional numerical model to study the effects of using (0.1, 0.2, and 0.4 wt%) of (Al_2O_3 and Cu) nanoparticles in water and glycol is developed. According to this study, using a Cu/water with 0.4 wt% as a coolant fluid on the uncovered PV/T system showed a better performance. Another numerical study is performed to examine the use of Ag/water and Al_2O_3/water nanofluids on PVT sheet & tube performances [9]. In comparison to pure water, they found that a 3% volume fraction of Ag and Al_2O_3 nanoparticles enhanced the thermal efficiency by 5.15% and 1.36%, respectively. Furthermore, the electrical efficiency of Ag/water and Al_2O_3/water improved by 1.88% and 0.95%, respectively. In order to evaluate the energy performances of the PV/T system, Al_2O_3, CuO, and SiC nanoparticles

have been applied in [10] with several volume fractions of 0.5, 1, 2, 3, and 4% in water. It is depicted that the SiC/water nanofluid is appropriate to improve the PV/T system production. In the work of [11], the impact of MWCNT/water nanofluid on PVT system performances was investigated. They found that by dispersing 1% of MWCNT in water, the electrical and thermal efficiencies were improved by 1.18% and 5%, respectively. The effectiveness of using nanofluids for the cooling of the PV cells is also proved through the results of [12]. They analyzed numerically the utilization of Cu/water end Al_2O_3/water nanofluid in PVT sheet & tube system. Moreover, by adding 2vol% of Cu/water, the thermal and electrical efficiency reached an enhancement of 4.1 and 1.9% while, those of Al_2O_3/water reached up to 2.7% and 1.2%, respectively.

In this work, the impact of dispersing 2% of copper and alumina nanoparticles in water as cooling fluids on the electrical and thermal efficiency of PVT sheet & tube system is examined numerically. Furthermore, the results were compared with those of water-based PVT. This work is organized into three main sections: Sect. 2 presents the mathematical model of nanofluid-based PVT. The results and discussion are illustrated in Sect. 3. Finally, the conclusion is shown in Sect. 4.

2 Mathematical Models of the Nanofluid-Based PVT Systems

This work investigates numerically the performance of a PV/T system using three types of cooling fluids: pure water, Al_2O_3/water, and Cu/water. Hence, The PV/T model is developed based on the energy balance established for each layer and the set of obtained Eqs. (1)–(8) is solved by RUNGE–KUTTA 4 (RK4) method in MATLAB software depending on [13]. The PV/T hybrid system studied in this research consists mainly of a single glass cover, a photovoltaic module, a Tedlar, a metal sheet absorber, a set of tubes, a cooling fluid, and an insulation layer for reducing the heat loss with the ambient. The whole design of the PV/T system and its components is presented in Fig. 1.
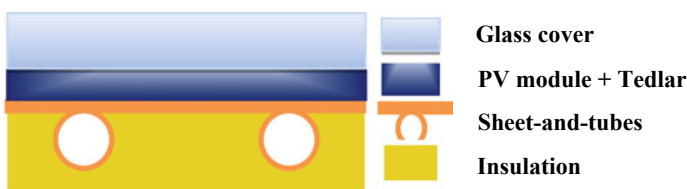
Fig. 1 PV/T "sheet & tubes" system design

2.1 Energy Balance Equation of the PV/T Systems

In this section, the energy balance for each layer of the PV/T system is established as shown in Table 1. Moreover, the set of obtained equations is written considering the different heat exchange by conduction, convection, and radiation and the effect of external excitation, e.g., solar radiation (see Fig. 2), ambient temperature. Furthermore, to simplify the calculation, the following assumptions are taken into account:

- The heat flux is only in one direction, from the top to the bottom of the PV/T system;
- The heat loss at the edges of the PV/T system is neglected;
- The material properties of each component are constant;
- The water and nanofluid mass flow rates in tubes are uniform.

2.2 Thermo-Physical Properties of Cooling Fluids

In this section, the common expressions adopted to determine the main key thermo-physical properties of cooling fluids influencing the PV/T system behavior are presented in Table 2.

3 Results and Discussion

3.1 Simulation Input Parameters

3.1.1 Effect of Volume Fraction on the PVT Performances

In this study, the thermal conductivity and heat capacity are analyzed in function of volume fraction in order to evaluate the impact of different nanofluids on the electrical and thermal behaviors of the PV/T system. Figure 3a and b show respectively the variations of the specific heat capacity Cp_{nf}/Cp_{bf} and of the thermal conductivity $\lambda_{nf}/\lambda_{bf}$ as a function of the volume fraction of Cu/water and Al_2O_3/water.

From Fig. 3a, it can be seen that the thermal conductivity of the two studied nanofluids varies linearly with the volume fraction and exhibits almost the same rate of variation. When the volume fraction varies from 0 to 2%, the thermal conductivity for Cu/water reaches 8.3%, while for Al_2O_3/water does not exceed 7.9%. So as a result, the thermal conductivity of the nanofluid increases as the nanoparticle volume fraction increases, which leads to a better cooling capacity. It is worth noting that adding nanoparticles to the base fluid in the tube enhances thermal conductivity and thickens the boundary layer along the tube's surfaces [9].

Table 1 Governing equation of the PV/T system layers

Layer	Equations	
Glass (upper layer)	$m_g C_g \dfrac{dT_{gext}}{dt} = \left[A_g G - A_{gg}\left(T_{gext}^4 - T_{sky}^4\right) - A_g h_{wind}\left(T_{gext} - T_{amb}\right) - A_g h_{cond,g}\left(T_{gext} - T_{gint}\right)\right]$	(1)
Glass (lower layer)	$m_g C_g \dfrac{dT_{gint}}{dt} = \left[A_g G_g + A_g h_{cond,g}\left(T_{gint} - T_{gext}\right) - A_g h_{cond,g\,pv}\left(T_{gext} - T_{pv}\right)\right]$	(2)
PV module	$m_{pv} C_{pv} \dfrac{dT_{pv}}{dt} = \left[A_g G_{gpv} + A_g h_{cond,gpv}\left(T_{gint} - T_{pv}\right) - A_{pv} h_{cond,pvted}\left(T_{pv} - T_{ted}\right) - A_g G_{g0}\left(1 - \left(T_{pv} - T_r\right)\right)\right]$	(3)
Tedlar	$m_{ted} C_{ted} \dfrac{dT_{ted}}{dt} = \left[A_{pv} h_{cond,pv\,ted}\left(T_{pv} - T_{ted}\right) - A_{ted} h_{cond,ted\,p}\left(T_{ted} - T_p\right)\right]$	(4)
Absorber	$m_p C_p \dfrac{dT_p}{dt} = \left[A_{ted} h_{cond,ted\,p}\left(T_{ted} - T_p\right) - A_{p\,tube} h_{cond,p\,tube}\left(T_p - T_{tube}\right) - A_{p\,iso} h_{cond,p\,iso}\left(T_p - T_{iso}\right)\right]$	(5)
Tubes	$m_{tube} C_{tube} \dfrac{dT_{tube}}{dt} = \left[A_{ptube} h_{cond,ptube}\left(T_p - T_{tube}\right) - A_{fluid} h_{conv,tubefluid}\left(T_{tube} - T_{fluid}\right) - A_{tubeiso} h_{cond,tubeiso}\left(T_{tube} - T_{iso}\right)\right]$	(6)
Cooling fluid	$m_{fluid} C_{fluid} \dfrac{dT_{fluid}}{dt} = \left[A_{fluide} h_{conv,tube\,fluid}\left(T_{tube} - T_{fluid}\right) - \dot{m} C_{fluide} T_{fluid}\right]$	(7)
Insulation	$m_{iso} C_{iso} \dfrac{dT_{iso}}{dt} = \left[A_{tube\,iso} h_{cond,tube\,iso}\left(T_{tube} - T_{iso}\right) + A_{p\,iso} h_{cond,p\,iso}\left(T_p - T_{iso}\right) - A_{iso} h_{wind}\left(T_{iso} - T_{amb}\right)\right]$	(8)
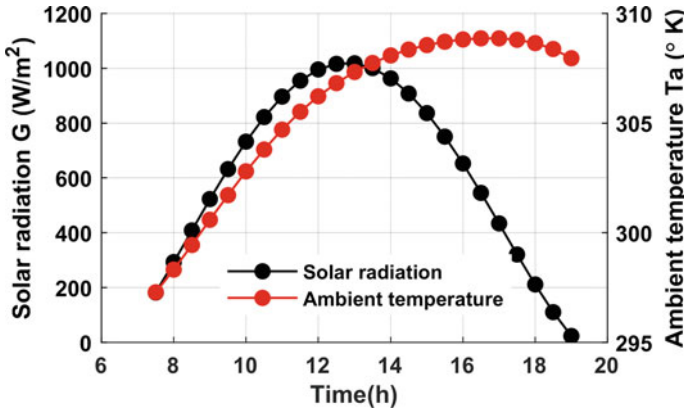
Fig. 2 Hourly variation of solar radiation and ambient temperature

Table 2 Thermo-physical properties of cooling fluids

Nanofluid		
Density (Kg/m^3) [14]	$\rho_{nf} = \rho_{np} + \phi\rho_{bf}$	(9)
Heat capacity (J/Kg.K) [14, 15]	$c_{p,nf} = c_{p,np} + (1-\phi)c_{p,bf}$	(10)
	$c_{p,nf} = \frac{c_{p,np}+(1-\phi)c_{p,bf}}{\rho_{nf}}$	(11)
Thermal conductivity (W/m.K) [16]	$k_{nf} = \frac{k_{np}+2k_{bf}-2(k_{bf}-k_{np})}{k_{np}+2k_{bf}+2(k_{bf}-k_{np})}k_{bf}$	(12)
Dynamic viscosity (Pa.s) [1]	$\phi < 0.05$, $\mu_{nf} = (1+2.5)_{bf}$ $0.05 < \phi < 0.1$, $0.05 < \phi < 0.1$, $\mu_{nf} = \left(1+2.5+6.5^2\right)\mu_{bf}$	(13) (14)
Nanofluid Nusselt number [17, 18]	Cu/Water [17]: $Nu_{nf} =$ $0.0059\left(1+7.628^{0.6886}Pe_{nf}^{0.001}\right)Re_{nf}^{0.9238}Pr_{nf}^{0.4}$ Al$_2$O$_3$ [18]: $Nu_{nf} = 0.085\,Re_{nf}^{0.71}Pr_{nf}^{0.35}$	(15) (16)
Pure water		
Dynamic viscosity (Pa.s) [19, 20]	$_{bf} = 2.701 10^{-10}T^2 - 1.849 10^{-7}T + 3.199 10^{-5}$	(17)
Nusselt number [19, 20]	$Nu_{water} = \frac{4.4+\left(0.00398\left(Re_{water}Pr_{water}\frac{Dint}{L}\right)\right)^{1.66}}{1+\left(0.0114\left(Re_{water}Pr_{water}\frac{Dint}{L}\right)\right)^{1.62}}$	(18)

Regarding the specific heat, it is clear from Fig. 3b that increasing the nanoparticles volume fraction from 0 to 4% produces a significant decrease of both nanofluids specific heat capacities; this is due to the low specific heat values of the nanoparticles heat capacity. At 2% of volume fraction, the specific heat capacity Cp_{nf}/Cp_{bf}
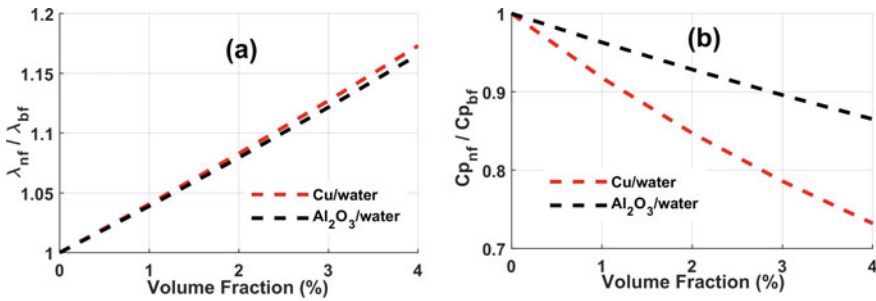
Fig. 3 a Thermal conductivity, **b** heat capacity of nanofluid studied

decreased to 0.92% for the Al_2O_3/water nanofluid and 0.84% for the Cu/water nanofluid. As known, the specific heat is defined as the amount of energy required to raise the temperature of a unit mass of a substance by one degree Kelvin. It is evident from the definition and at the same condition that any substance with lower specific heat must provide a higher temperature according to other fluids [21].

3.1.2 Effect of Mass Flow Rate on the PVT Performances

Figure 4a and b display the effect of the fluid mass flow rate on the electrical and thermal performance of the PV/T system using Cu/water, Al_2O_3/water, and pure water. It is observed that the electrical efficiency increases slightly as the mass flow rate rises, while thermal efficiencies increase rapidly with the mass flow rate rises until a maximum value of 62.5, 55, and 50% for Cu/water, Al_2O_3/water, and pure water, respectively, noted at a mass flow rate of 0.04 kg/s. After that, a slow increase of the efficiencies is observed by increasing the mass flow rate. In what follows, the mass flow rate value is taken as 0.04 kg/s.
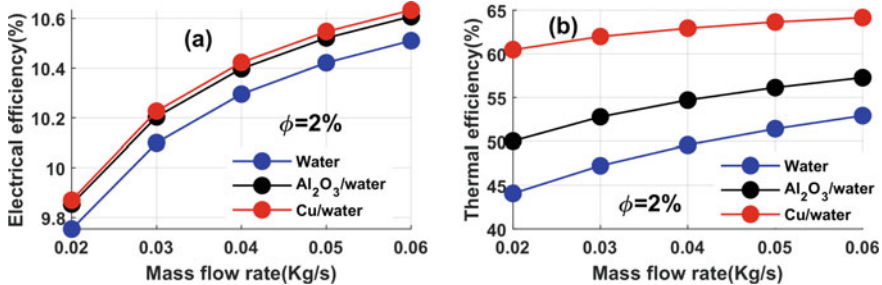
Fig. 4 Variation of **a** electrical efficiency, **b** thermal efficiency versus mass flow rate using pure water, Al_2O_3/water, and Cu/water nanofluids

3.2 PV/T System Performances

3.2.1 Electrical Performance

The electrical performances are calculated using the equations presented in [19]. The effect of the dispersing 2% of copper (Cu) and alumina (Al_2O_3) nanoparticles in pure water on the electrical performance of the PVT module is shown in Fig. 5. In particular, the variations of PV cell temperature (a), electrical efficiency (b), and electrical power (c), for three fluids: pure water, Cu/water, and Al_2O_3/water are depicted. According to Fig. 5a, at 13:00, the PV temperature of the PV/T panel reached the values of 54.2 °C, 54.67 °C, and 56.57 °C for Cu/water nanofluid, Al_2O_3/water nanofluid, and the pure water, respectively. According to this figure, it was noticed that the use of nanofluids leads to the decrease of the PV cell temperature by 4% in the case of Cu/water, and 3.3% in the case of Al_2O_3/water when compared with the pure water.

The PV panel temperature using Cu/water is the lowest. Moreover, Cu/water has a higher heat transfer coefficient than Al_2O_3/water. As a result, dispersing nanoparticles in the base fluid helps to increase the thermal conductivity of the substances that in turn enforce the heat transfer between the PVT collector layers. A diminution in PV temperature led to improve the PV electrical efficiency as proved in Fig. 5b. As observed, the electrical efficiencies are inversely proportional to the PV cell temperatures. As the PV temperature of the PV/T system increases, the electrical efficiency gain decreases and reaches a minimum value at 13:00 of about 10.42% and 10.39%, respectively, for Cu/water, Al_2O_3/water nanofluids, and 10.29% for pure water. Higher electrical efficiency is demonstrated by both nanofluids cases.

As shown in Fig. 6, the use of nanofluids slightly improves the electrical productions of the PV/T system, by about 1% at 13:00 the maximum electrical power of the PV/T reached 115.95 W, 115.66 W, and 114.53 W for Cu/water, Al_2O_3/water, and pure water, respectively. From an electrical viewpoint, the use of Cu/water nanofluids, which has the higher thermal conductivity in comparison with the other investigated cooling fluid, increases the heat transfer exchange in the PV/T system, and consequently, improving the system's electrical output.
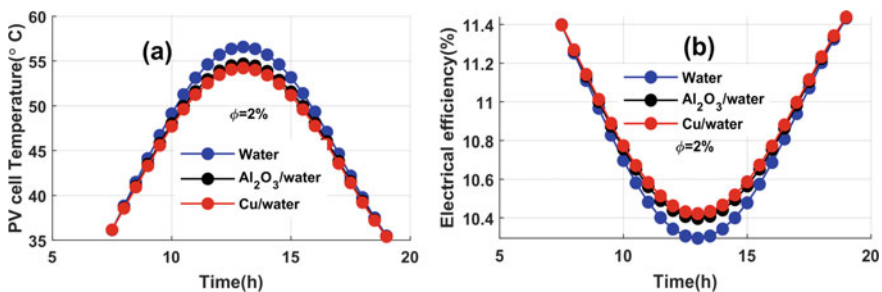
Fig. 5 Evolutions of PV temperature **a** electrical efficiency, **b** of PV/T systems

Fig. 6 Electrical power evolutions of PV/T systems
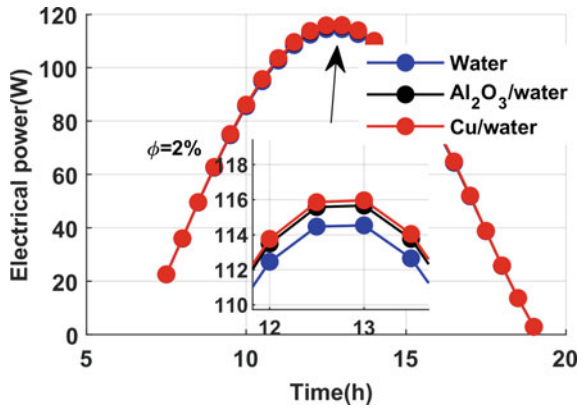
3.2.2 Thermal Performance

In this section, the impact of nanofluids on the thermal performances has been examined in which are calculated utilized the formula reported in [22]. Figure 7a presents the variation of the outlet temperature of the PVT system for the three working fluids: pure water, Cu/water, and Al_2O_3/water nanofluids.

At 13:00, it was noted that the working fluid outlet temperature reaches the higher value, (59.7 °C), for Cu/water nanofluid compared to the other working fluids, Al_2O_3/water (57.5 °C) and pure water (55.7 °C). A difference of 3.9 °C was recorded between Cu/water and pure water. This conclusion is in good agreement with the fact that the specific heat of the nanofluid Cu/water decreases by 15% for the volume fraction of 2% nanoparticles. Figure 6b depicts the thermal efficiency of the PV/T system for the three studied cooling fluids. Cu/water and Al_2O_3/water nanofluids significantly improve the thermal efficiency of the PV/T system. The thermal efficiency reaches its maximum value of 63.9 and 55.7% for Cu/water and Al_2O_3/water, respectively, against 49.58% in the case of pure water. As a result, an enhancement of 26.9 and 10.3%, was obtained by adding copper and alumina in the base fluid (water), respectively.
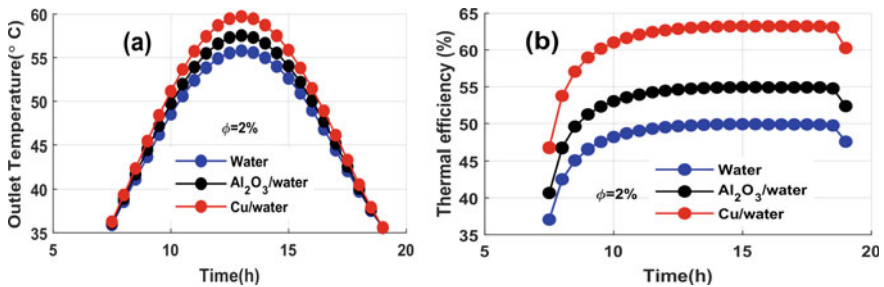
Fig. 7 Evolutions of fluid outlet temperature (**a**) thermal efficiency (**b**) of PV/T systems

Fig. 8 Thermal power evolutions of PV/T systems
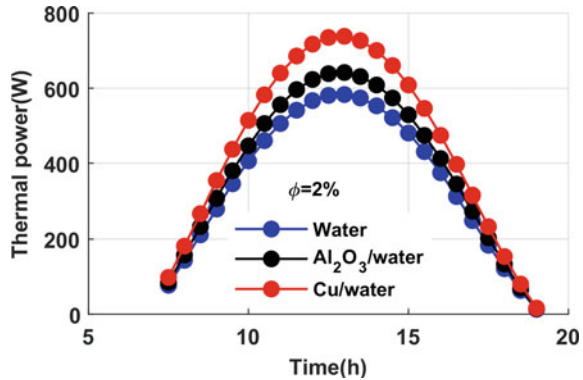
The daily thermal power PV/T for the investigated fluids is illustrated in Fig. 8. At 13:00, the thermal power reaches its maximum value of 737.01 W, 640.59 W, and 580.61 W for Cu/water, Al$_2$O$_3$/water, and pure water, respectively. The use of nanofluids allows to enhance the thermal performance of PV/T collector. Indeed, an increase in thermal power of 26.9% for Cu/water and 10.3% for Al$_2$O$_3$/water was recorded. These results were noted against the use of pure water.

4   Conclusion

In this work, a numerical study of the PV/T system using different cooling fluids, e.g., Cu/water, Al$_2$O$_3$/water, and pure water was carried out. According to the simulation results obtained, the following conclusions can be observed:

- Compared to pure water, Cu/water and Al$_2$O$_3$/water nanofluids provide better performance for the PVT system;
- Using Cu/water and Al$_2$O$_3$/water nanofluid as cooling fluids with a mass flowrate of 0.04 kg/s leads to reducing the PV temperature by 2.36 °C and 1.89 °C, respectively;
- The improvement of Cu/water and Al$_2$O$_3$/water nanofluids-based PV/T with 2% of nanoparticles reached by 1.24% and by nearly 0.99% at 13:00, respectively;
- The maximum electrical power produced by the PVT collector occurred at 13:00 is 115.95, 115.66, and 114.53 W for Cu/water, Al$_2$O$_3$/water, and water, respectively;
- The nanofluid increases the thermal efficiency of the PV/T system by 26% for Cu/water and by 10.33% for Al$_2$O$_3$/water;
- The thermal performance of the Cu/water nanofluid-based PVT system was found to be the highest compared to the Al$_2$O$_3$/water nanofluid-based PVT and the pure water-based PV/T, which are 737.01 W, 640.59 W, and 580.61 W, respectively.

References

1. Jia Y, Ran F, Zhu C, Fang G (2020) Numerical analysis of photovoltaic-thermal collector using nanofluid as a coolant. Sol Energy 196:625–636. https://doi.org/10.1016/j.solener.2019.12.069
2. El Fouas C, Hajji B, Gagliano A, Tina GM, Aneli S (2020) Numerical model and experimental validation of the electrical and thermal performances of photovoltaic/thermal plant. Energy Convers Manage 220:112939. https://doi.org/10.1016/j.enconman.2020.112939
3. Ventura C, Tina GM, Gagliano A, Aneli S (2021) Enhanced models for the evaluation of electrical efficiency of PV/T modules. Sol Energy 224:531–544. https://doi.org/10.1016/j.solener.2021.06.018
4. Abdulalah A, Misha S, Tamaldin N, Mohd Rosli MA, Sachit F (2018) Photovoltaic thermal/solar (PVT) collector (PVT) system based on fluid absorber design: a review. J Adv Res Fluid Mech Therm Sci 48:196–208
5. Aneli S, Gagliano A, Tina GM, Hajji B (2021) Analysis of the energy produced and energy quality of nanofluid impact on photovoltaic-thermal systems. In: Proceedings of the 2nd International conference on electronic engineering and renewable energy systems, Singapore, pp 739–745. https://doi.org/10.1007/978-981-15-6259-4_77
6. Sardarabadi M, Passandideh-Fard M, Zeinali Heris S (2014) Experimental investigation of the effects of silica/water nanofluid on PV/T (photovoltaic thermal units). Energy 66:264–272. https://doi.org/10.1016/j.energy.2014.01.102.
7. Michael JJ, Iniyan S (2015) Performance analysis of a copper sheet laminated photovoltaic thermal collector using copper oxide–water nanofluid. Sol Energy 119:439–451. https://doi.org/10.1016/j.solener.2015.06.028
8. Rejeb O, Sardarabadi M, Ménézo C, Passandideh-Fard M, Dhaou MH, Jemni A (2016) Numerical and model validation of uncovered nanofluid sheet and tube type photovoltaic thermal solar system. Energy Convers Manage 110:367–377. https://doi.org/10.1016/j.enconman.2015.11.063
9. Khanjari Y, Pourfayaz F, Kasaeian AB (2016) Numerical investigation on using of nanofluid in a water-cooled photovoltaic thermal system. Energy Convers Manage 122:263–278. https://doi.org/10.1016/j.enconman.2016.05.083
10. Al-Waeli AHA et al (2017) Evaluation of the nanofluid and nano-PCM based photovoltaic thermal (PVT) system: an experimental study. Energy Convers Manage 151:693–708. https://doi.org/10.1016/j.enconman.2017.09.032
11. Nasrin R, Rahim NA, Fayaz H, Hasanuzzaman M (2018) Water/MWCNT nanofluid based cooling system of PVT: experimental and numerical research. Renew Energy 121:286–300. https://doi.org/10.1016/j.renene.2018.01.014
12. Hissouf M, M'barek F, Najim M, Charef A (2020) Numerical study of a covered photovoltaic-thermal collector (PVT) enhancement using nanofluids. ScienceDirect, 15 Mar 2020. https://www.sciencedirect.com/science/article/pii/S0038092X20300906?via%3Dihub. Accessed 15 Jun 2021
13. El Fouas C, Chereches N-C, Hudisteanu SV, Hajji B, Turcanu F-E, Chereceș M (2022) Numerical and parametric analysis for enhancing performances of water photovoltaic/thermal system. Appl Sci 12:646. https://doi.org/10.3390/app12020646
14. Pak BC, Cho YI (1998) Hydrodynamic and heat transfer study of dispersed fluids with submicron metallic oxide particles. Exp Heat Transfer 11(2):151–170. https://doi.org/10.1080/08916159808946559
15. Xuan Y, Roetzel W (2000) Conceptions for heat transfer correlation of nanofluids. Int J Heat Mass Transf 43(19):3701–3707. https://doi.org/10.1016/S0017-9310(99)00369-5
16. Maxwell JC (1873) A treatise on electricity and magnetism. Clarendon Press
17. Xuan Y, Li Q (2003) Investigation on convective heat transfer and flow features of nanofluids. J Heat Transfer 125(1):151–155. https://doi.org/10.1115/1.1532008
18. El Bécaye Maïga S, Tam Nguyen C, Galanis N, Roy G, Maré T, Coqueux M (2006) Heat transfer enhancement in turbulent tube flow using Al_2O_3 nanoparticle suspension. Int J Numer Methods Heat Fluid Flow 16(3):275–292. https://doi.org/10.1108/09615530610649717

19. Fouas CE, Hajji M, Bouselham L, Hajji B, Mehdi AE, Bouali H (2017) Analysis and design of an energy system based on PV/T water collector for building application. In: 2017 International renewable and sustainable energy conference (IRSEC), Dec. 2017, pp 1–5. https://doi.org/10.1109/IRSEC.2017.8477361

20. Hajji M, Naimi SE, Hajji B, El Hafyani ML (2014) A comparative study between two structures of hybrid photovoltaic/thermal (PV/T) collectors for water pumping systems. In: 2014 International renewable and sustainable energy conference (IRSEC), Oct. 2014, pp. 235–240. https://doi.org/10.1109/IRSEC.2014.7059745

21. Said Z et al (2015) Performance enhancement of a flat plate solar collector using Titanium dioxide nanofluid and polyethylene glycol dispersant. J Clean Prod 92:343–353. https://doi.org/10.1016/j.jclepro.2015.01.007

22. Duffie JA, Beckman WA, Blair N (2020) Solar engineering of thermal processes, photovoltaics and wind. Wiley

Thermal Comfort Assessment of a Small House in Portugal Using EnergyPlus and Ansys Fluent

Inês M. Teixeira [iD]**, Diogo B. Esteves** [iD]**, Nelson J. Rodrigues** [iD]**,
Luís A. Martins** [iD]**, José C. Teixeira** [iD]**, Ana C. Ferreira** [iD]**,
and Senhorinha F. Teixeira** [iD]

Abstract According to the European Union Directive 2018/844, almost 50% of the final energy consumption is used in heating and cooling processes, whereas, of these, 80% are used in the buildings. Thus, to accomplish the proposed energy and climate objectives for 2050, priority has to be given to energy efficiency as well as the deployment of renewable energies in buildings. Most thermal modeling of buildings is carried out through computational dynamic simulation programs. The main objective of this paper is to compare the thermal analyses performed by two programs, EnergyPlus and Ansys Fluent. For this, the model of a simplified building was created, with its thermal surroundings, envelope characteristics, and technical systems. In the model, the insulation of the building was not considered, in order to better compare both software under study. The objective is to verify if two different philosophy software have similar results and validate them. Finally, the temperature

I. M. Teixeira (✉) · D. B. Esteves · A. C. Ferreira · S. F. Teixeira
ALGORITMI, Universidade do Minho, Guimarães, Portugal
e-mail: ines.teixeira@dps.uminho.pt

D. B. Esteves
e-mail: a70937@alunos.uminho.pt

A. C. Ferreira
e-mail: acferreira@dps.uminho.pt

S. F. Teixeira
e-mail: st@dps.uminho.pt

N. J. Rodrigues · L. A. Martins · J. C. Teixeira · A. C. Ferreira
Metrics, Universidade do Minho, Guimarães, Portugal
e-mail: nrodrigues@dem.uminho.pt

L. A. Martins
e-mail: lmartins@dem.uminho.pt

J. C. Teixeira
e-mail: jt@dem.uminho.pt

A. C. Ferreira
Centro de Investigação em Organizações, Mercados e Gestão Industrial (COMEGI), Universidade Lusíada Norte, Porto, Portugal

and PMV results obtained from EnergyPlus are compared with the results of Ansys Fluent. There was a small deviation for both software, with a PMV value of -1.10 for the EnergyPlus and for -1.12 Fluent. The PMV distribution was compared with K–S test that showed no significant statistical differences.

1 Introduction

Buildings currently account for 40% of total final energy consumption in the European Union. As the sector is expanding, its energy consumption can be expected to increase. Almost 50% of the final energy consumed in the EU is used for heating and cooling purposes, and about 80% of this is used in buildings. This accounts for about 36% of all CO_2 emissions in the European Union [1].

Portugal experienced a huge growth in the buildings sector, both residential, commercial, and services buildings, after integration in the European Union, and the total final energy consumption increased until 2007. From 2010 to 2017, the energy consumption of the residential sector decreased by 13%, linked to a decrease in the consumption of petroleum products, mainly heating oil and liquefied petroleum gas (LPG), in which the consumption of the latter fell by 38%. In the same period, natural gas and electricity consumption decreased by 16% and 13%, respectively. On the other hand, the use of energy from renewable sources, such as biomass and solar thermal, grew by 11% in the same period [2]. Following the same trend, energy consumption in the service sector has fallen by 10%. Petroleum products went from a 27% contribution in 2007 to 8% in 2017 [2]. In the opposite direction, natural gas evolved from a contribution of 8% to 13%. Meanwhile, electricity in 2017 accounted for 73% of all energy consumption in this sector. The contribution of other energy forms evolved from 1% in 2007 to 6% in 2017 [2]. These reductions in building energy consumption were influenced by improvements in energy efficiency, but were mostly due to the effects of the prolonged economic crisis from 2010 onward. The efficiency improvements have been ensured by the use of better building materials together with the development of methodologies used in the design phase to predict the energy consumption, such as spreadsheets and simulation software. The Portuguese Decree-Law nº118/2013 defines building dynamic simulation as the prediction of the annual energy consumption based on occupancy, HVAC system, and the local typical Meteorological year (TMY) on an hourly basis [3, 4]. The most commonly used dynamic simulation programs, certified by the ASHRAE140 standard, include: EnergyPlus, Trace 700, Carrier HAP, IES-VE, Design Builder, eQuest, and TRNSYS [3, 4]. Moreover, thermal comfort is essential for the well-being of users and also a mandatory need for certification of buildings and constructions. In addition, it has a major impact on occupant health and performance [8]. In this paper, the EnergyPlus, was used to model the energy

consumption for heating, cooling, ventilation, lighting, and other internal loads in a mono-zone building and compared with the Ansys Fluent, an industry-leading fluid simulation software known for its advanced physics modeling capabilities and industry-leading accuracy [5]. The main objective is to compare the mean temperature and predicted mean vote (PMV), considering the results from both tools. PMV is an index that predicts the average thermal sensation vote of a large group of people on a 7-point scale, based on the thermal balance of the human body [6, 7]. The assessment is performed inside the building without insulation during the heating season, by analyzing the similarities of the results and exploring the reasons and explanations for the differences. These two parameters were chosen because indoor air temperature has the greatest weight in thermal comfort, and PMV is the best approximation for determining and quantifying thermal comfort.

2   Methods and Materials

In this section, the test building is defined, with its geographic location, in addition to the characterization of its envelope, which is essential for the calculation of the energy needs for heating and cooling.

2.1   Study Case Description

The building consists of a rectangular prism with the dimensions of 10 m × 6 m × 3 m. The south facade includes 12 m^2 of glazed area, and the floor area equals 60 m^2 with a ceiling height of 3 m. On the ceiling wall, three boundary conditions were defined, two inlets with 0.3 m × 0.15 m (standard measurement) and one extraction as a pressure outlet with 0.15 × 4 m. Figure 1 shows the model representation. The location chosen was Viana de Castelo, located in the northwest of Portugal.

Because of the increasingly demanding level of regulation, the constructive solutions, without thermal insulation, have been progressively abandoned as they are practically excluded from all regulatory application contexts. Nevertheless, this scenario is the simplest case to compare both software. Table 1 summarizes the definition of

Fig. 1 Schematic model of the building geometry in the study
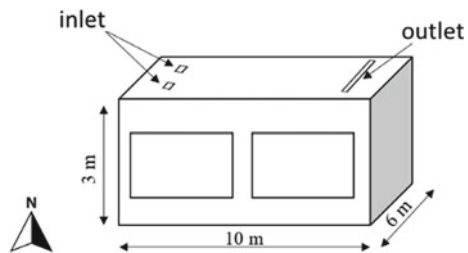
Table 1 Constructive solutions of the building without insulation [9]

Constructive solution type	Material	Thickness (m)	Conductivity (W/mK)	Density (kg/m^3)	Heat capacity (J/kgK)	U^* (W/m^{2}K)
Simple exterior wall	Plaster	0.02	1.333	1900	1046	2
	Light concrete block	0.20	0.408	2000	840	
	Traditional beton plaster	0.02	1.333	1900	1046	
Ceiling	Gravel	0.05	2	1700	0.8	1.61
	Bitumen	0.232	0.043	1000	185	
	Autoclaved beton	0.05	0.160	450	1000	
	Structural beton	0.2	2	2300	1000	
	Traditional beton plaster	0.02	1.333	1900	1046	
Pavement	Glazed ceramics	0.01	1.25	2300	920	2.06
	Grout	0.11	1.294	1900	1046	
	Structural beton	0.2	2	2300	1046	

* U—global heat transfer coefficient

the building envelop characteristics, which are existing constructive solutions, but devoid of any insulation, thus not fulfilling the current regulatory requirements.

The glazed windows were considered as a simple 6 mm colorless glass with a density equal to 2200 kg/m^3 [9], with a heat capacity equal to 750 J/kg.K and a conductivity of 1.10 W/m.K, corresponding to a global heat transfer coefficient U of 6 W/m^2.K.

2.2 Ansys Fluent—CFD Model

The computational fluid dynamics (CFD) simulation was developed using FLUENT 2020 R2, from ANSYS®. In addition to the mass and momentum balances, the simulation includes the energy equation, which accounts for thermal exchanges between boundaries and as well as within the fluid, the radiation model, and the species model for humidity calculations. The turbulence model chosen was the transition $k - kl - \omega$, with 3 equations, developed to focus on the boundary layer development and can be used to effectively address the transition of the boundary layer from a laminar to turbulent regimes. The radiant heat transfer was evaluated with the surface-to-surface

model (S2S), which considers the radiant heat exchange between surfaces in a field of view. In this model, the solar calculator was also activated, which is an automatic way to consider the thermal gains from solar radiation exposure by simply indicating the hour of the TMY dataset.

The mesh was generated considering 212,160 elements, presenting very good quality metrics, such as average skewness of 0.05 and average orthogonal quality of 0.973. The near-wall mesh was refined to address the near-wall flow.

All the walls, (except the pavement defined only as a convection thermal condition), are considered mixed for the thermal condition combining the convection and radiation boundary conditions. The free stream temperature is used to calculate the convective losses or heat gains. In this case, free stream and radiation temperature were defined by a user-defined function based on the temperature profile for January 23 of the TMY. As for the outside temperature profile, it ranges between 2 ℃ and 9.72 ℃ with a curve approximated by a polynomial model of degree 4. The temperature starts to increase at 8:00, reaching the peak at 16 h. After this hour, the temperature gradually decreases, achieving 7.35 ℃ at 21:00.

The building has an HVAC system set to 1 ren/h, which corresponds to a velocity-inlet of 1.11 m/s at 30 ℃, to guarantee an internal comfortable temperature of 20 ℃.

2.3 EnergyPlus—Dynamic Simulation

The EnergyPlus software is a collection of integrated modules to calculate the energy required for heating or cooling a building, using different active systems and energy sources. The software can model, with excellent precision, radiant, and convective heat fluxes between indoor and outdoor, HVAC systems performance, heat exchange with the ground, thermal comfort, natural, artificial, and hybrid systems [4, 10].

The simulation process started by defining the variables presented in the previous subsection (geometry, study zone, period of time of the TMY, materials, thermal loads, and set point temperature) [10]. After this process, the simulations are performed to obtain the heating and cooling demands of the building, assuming a temperature set point of 20 ℃. The simulation is based on the solution of the heat balance equation through the surfaces of the building [10]. It is essential to point out that the 3D model was created in SketchUp2017, linked to OpenStudio via the appropriate plugin.

All the building surfaces were defined as well as their boundary conditions. The building envelope has outdoor conditions, and only one thermal zone was defined for the interior. After defining the 3D model, the model was imported to the OpenStudio that interfaces with the EnergyPlus.

3   Results and Discussion

Figure 2 shows the variation of the air temperature inside the building for a winter day, January 23 considering the results from EnergyPlus and Ansys Fluent simulations. This specific day was selected because it is, statistically, the coldest of the year. The analysis was done for the occupancy period, i.e., from 8:00 to 21:00.

As can be seen in Fig. 2a, until 14:00, the temperature obtained from Ansys Fluent is slightly lower than that from EnergyPlus. After 14:00, the reverse is observed. The slower rate of cooling after 15:00 observed with Ansys Fluent may occur due to different inertial considerations on both software and also due to differences on how Ansys Fluent considers the solar thermal radiation and near-wall heat transfer. Overall, the average indoor temperature is 19.44 °C with Ansys Fluent, which is
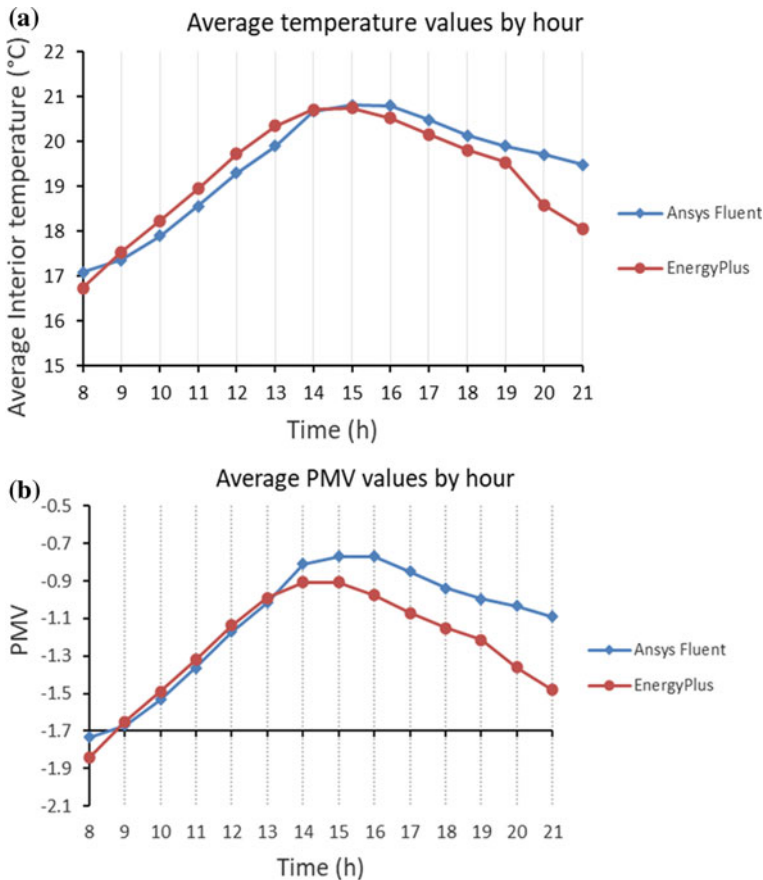
Fig. 2   a Indoor air temperature and **b** mean PMV in the building for January 23

slightly colder than that obtained with EnergyPlus, 19.26 °C. Nonetheless, both software were in close agreement.

Figure 2b shows that the variation of the PMV over time tends to follow the temperature profile, demonstrating the greater influence of this variable on comfort. Nevertheless, the average PMV in Ansys Fluent is −1.12 and in EnergyPlus is −1.25. Both results correspond to a medium cold sensation with 32–38% of people dissatisfied with the thermal conditions inside the building, which is mainly due to the lower surface temperatures of the non-insulated envelope. In both cases, the dissatisfaction rate is considerable and the feeling of cold in the space is predominant. For a better comparison of both software distributions, the Kolmogorov–Smirnov statistic test (K–S test) was used, which proved that there are no relevant statistical discrepancies between them.

4 Conclusions

This work provided a direct comparison between two different software programs. Ansys Fluent is a R&D oriented numerical simulation with greater accuracy of the phenomena, whereas EnergyPlus is more focused on building certification and uses simplified models for quick and averaged results.

The results show that, despite their differences, both software yield realistic results and are in close agreement. On the other hand, the use of both software can be of great value for both academic and business purposes. Although EnergyPlus is more cost effective, taking less computing effort, Ansys Fluent can complement the study with a better understanding of a particular 3D space by providing detailed information about the local heat fluxes and temperatures.

Both software did show that a building without insulation is greatly influenced by the outside temperature that affects the thermal comfort due to the flow average radiation temperature and despite an air temperature near 20 °C. This is represented by the simulated temperature oscillations that follow the temperature profile of the day. Based on K–S test, it was verified that the results from Ansys Fluent and EnergyPlus are not statistically different.

References

1. Rodrigues A, Canha B (2009) Térmica de Edifícios, Lisboa, Edições OR. Lisbon
2. DGEG ADENE—Agência para a Energia (2020) Energia em Números
3. Roriz L (2007) Climatização, Concepção, Instalação e Condução de Sistemas, 2nd ed

4. US Department of Energy (2021) EnergyPlus essentials, 58
5. Manual A (2000) Ansys Inc. Canonsburg, PA 15317:724–746
6. Standardization FOR, Normalisation DE (1987) International Standard ISO
7. Rodrigues N, Oliveira R, Teixeira S et al (2015) Thermal comfort assessment of a surgical room through computational fluid dynamics using local PMV index. Work 51:445–456. https://doi.org/10.3233/WOR-141882
8. Teixeira I, Rodrigues N, Teixeira S (2022) Energy, thermal comfort and pathologies—a current concern. Occupational and environmental safety and health III. Springer International Publishing, Cham, pp 273–281
9. Santos C, Matias L (2006) Coeficiente de Transmissão Térmica de Elementos da Envolvente de Edifícios 171
10. Esteves D, Silva J, Martins L et al (2021) Building energy performance: comparison between energyplus and other certified tools. Lect Notes Comput Sci 12949:493–503. https://doi.org/10.1007/978-3-030-86653-2_36

Influence of Typical Meteorological Years on the Optimization of Incident Solar Radiation for PV Applications in Portugal

Ana C. Ferreira 🆔**, Nuno Menezes, Inês M. Teixeira** 🆔**,**
Senhorinha F. Teixeira 🆔**, and Luís A. Martins** 🆔

Abstract In Portugal, the values used for forecasting the production of solar thermal systems, photovoltaic and thermal simulation of buildings are periodically updated due to climate changes. There is a special interest in optimizing these systems since Portugal has high insolation levels and the sector has benefited from technological developments that improved its efficiency and costs. The main objective is the analysis of typical meteorological years influence on the optimization of incident solar radiation, applied to the municipality of Braga (north of Portugal). A study was carried out on the fundamental concepts related to solar geometry, and the best methods for data processing were identified to develop an analysis tool, in MS Excel, to compare the selected climatic databases. Different solar tracking configurations were also considered: biaxial, polar axis, horizontal axis and fixed tracking systems. The analysis shows that the climatic database update resulted in the increase of total incident solar radiation when comparing both typical meteorological years. The polar axis present values very close to the orthogonal incidence biaxial system, which maximizes the incident solar radiation in the panel. The application of simple atmospheric models to estimate/predict the total radiation incidence presents more conservative values when compared to the more complex models.

A. C. Ferreira (✉) · N. Menezes · L. A. Martins
MEtRICs, University of Minho, Guimaraes, Portugal
e-mail: acferreira@dps.uminho.pt

L. A. Martins
e-mail: lmartins@dem.uminho.pt

A. C. Ferreira · I. M. Teixeira · S. F. Teixeira
ALGORITMI, University of Minho, Guimaraes, Portugal
e-mail: ines.teixeira@dps.uminho.pt

S. F. Teixeira
e-mail: st@dps.uminho.pt

A. C. Ferreira
Centro de Investigação em Organizações, Mercados e Gestão Industrial (COMEGI), Universidade Lusíada Norte, Porto, Portugal

925

1 Introduction

Portugal submitted to the United Nations the Roadmap for Carbon Neutrality 2050, which represents the country's long-term development strategy regarding the Greenhouse Gas (GHG) emissions, in compliance with the Paris Agreement targets. Considering that one of the outlined objectives is the total decarbonization of the electricity production sector and urban mobility, Portugal has committed to a reduction between 45 and 55% of GHG by 2030, between 65 and 75% by 2040 and between 85 and 90% by 2050, in comparison with 1990 [1]. A special focus is given to the building sector, which represents about 36% of total GHG emissions and 40% of energy consumption. Within the scope of the National Energy and Climate Plan 2030 (NECP2030), the energy strategy for building renovations sets the goals for achieving a zero-emission building stock by 2050 and to orient choices toward decarbonized solutions [2]. This energy strategy predicts the intensive implementation of solar PV systems.

Typical and representative climatic data facilitate the design and long-term evaluation of systems as weather conditions can significantly vary from year to year [3, 4]. Typical Meteorological Year (TMY) corresponds to a set of meteorological data defined for a specific location, listing hourly values of solar radiation and other meteorological parameters over the year. The most used method to generate a TMY is the Finkelstein–Schafer statistical method. Ohunakin et al. [5], Jiang [6], Skeiker and Ghani [7] applied this method to generate TMYs for different locations in their studies. It is an empirical approximation that selects data from different years of the period under study, considering variables, such as ambient temperature, precipitation, relative humidity, wind velocity, cloudiness rates, global solar radiation, diffuse, and direct solar radiation. The data of each meteorological parameter is grouped and the cumulative distribution functions (CDF) are calculated [5, 8]. These methodologies have been applied either for flat panels or concentered solar systems [9]. Portugal has a huge solar energy production potential when compared to Northern European countries, due to a higher average annual value of insolation and lower latitude [10]. Thus, photovoltaic (PV) technology can be used to convert the most abundant renewable energy source, solar energy, directly into electricity. The associated technologies have been investigated at a scientific and commercial level, resulting in continuously efficiency improvements and lower prices over the last few years [11, 12].

From 2013, the energy certification system introduced an update to the TMY used as reference, so as to take into account Climate change effects. Also, the building's energy certification and the design of solar PV systems are based on yearly climate forecasting. Thus, the improvement of the solar incidence estimation presents itself as an opportunity to increase the efficiency of solar energy conversion systems. The

objective of this paper is to study the influence of TMY on the optimization of incident solar radiation in Portugal in flat-panel systems. Thus, the former TMY corresponding to the period 1961–1990, applicable to the old energy legislation Regulation of the Characteristics of Thermal Behavior of Buildings, (RCCTE) is analyzed and compared with the TMY based in the period of 1971–2000 and applicable to the current Energy Certification System of Buildings (SCE) (Decree-Law nº118/2013). The calculation methodology was developed in MS Excel. With this comparison, the impact of the TMY update on the energy production forecast is evaluated, through the variation of the incident solar radiation in a flat surface and for different solar tracking systems.

2　Methodology for Solar Radiation Calculation

A methodology was defined to develop a tool for predicting the incident solar radiation on inclined flat panels, by assessing the effect of TMY on the total incident energy. The TMY database includes data for the 8760 h of a reference year, considered as representative of a specific location. In this study, the considered location was the Braga municipality (north of Portugal), with a centered latitude of $\phi = 41.54°$ and reference altitude of 170 m. The two data sets were then obtained using the *Solterm* software:

(1)　TMY 2006—data period 1961–1990 applicable to RCCTE up to 2012;
(2)　TMY 2013—data period of 1971–2000 applicable to SCE from 2013 onwards.

　　The calculation of radiation components was performed considering two isotropic atmospheric models: simple and diffuse isotropic models (Table 1). For the simple isotropic model, the total radiation on an inclined plane corresponds to the sum of the direct beam radiation on inclined plane with the diffuse horizontal irradiance. For the diffuse isotropic model, total radiation on an inclined plane includes the direct component, the diffuse solar sky radiation and the ground reflected diffuse radiation.
　　Both models considered three particularities:

Table 1 Simple and diffuse isotropic model assumptions

Atmospheric model	Considerations
Simple isotropic model	Diffuse incident radiation on the inclined plane is considered equal to that incident on a horizontal plane: $G_t = G_{bt} + DHI$ G_{bt}—Direct beam radiation on the inclined plane DHI—Diffuse Horizontal Irradiance
Diffuse isotropic model	Separation of diffuse radiation into two components: $G_t = G_{bt} + G_{dt} + G_{rt}$ G_{dt}—Diffuse solar sky radiation G_{rt}—Ground reflected diffuse radiation ($\rho = 0.2$)
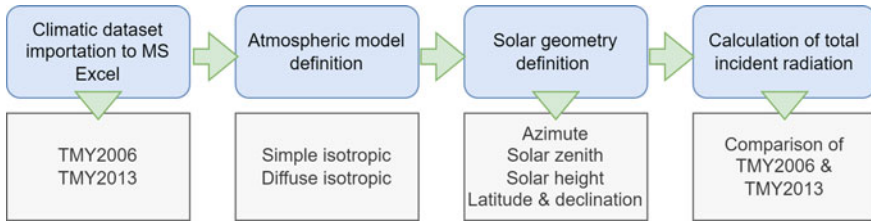
Fig. 1 Methodology implemented in MS Excel for solar radiation calculation

- True solar time was used in all calculations on an hourly basis;
- The direction of the sun radiation was evaluated at the half-hour;
- Day-night transitions correction coefficient.

The direction of the sun rays was evaluated at the half-hour because in the TMY files, the solar radiation value for each hour represents the total energy received in the previous hour. The determination of incidence angles on inclined surfaces is based on usual solar geometry, equations evaluated at the sun exact time position. [13]. The methodology implemented in MS Excel for solar radiation calculation can be calculated as in Fig. 1.

In the day-night transitions, the solar zenith angle (θ_z) is close to 90º which leads to the overestimation of direct radiation because the TMY data correspond to the values of global and diffuse radiation on a horizontal surface. A coefficient was considered to avoid incorrect values in day-night transition hours.

3   Results and Discussion

For the radiation analysis, the monthly sum of the main components of solar radiation was calculated as well as the monthly mean temperature evolution, as shown in Fig. 2 for TMY 2006 and TMY 2013, respectively. All the components of solar radiation are maximum in the summer period due to the higher apparent position of the sun, the greater number of daylight hours and to the reduced cloudiness. Comparing both TMYs, the mean annual temperature is lower for TMY 2006 (14.03 °C) when compared to TYM 2013 (15.01 °C) in what appears to be a clear sign of climate change. Upon receiving solar radiation, the solid surfaces store heat and its release is not immediate. This thermal inertia effect is reflected in the mean temperature variation that is kept high. Considering the values for the municipality of Braga, it can be seen that the mean radiation values drop significantly in September and October, but it is only from November that there is a significant drop in the mean air temperature. The global horizontal radiation reaches a peak in July with 206 kWh/m^2 for the TMY 2006 and a higher 225 kWh/m^2 for TMY 2013. It should be noted that, in the winter season, the direct normal radiation exceeds the values of the global

horizontal radiation, mostly because in winter the apparent position of the sun is lower and the incident solar radiation is spread over a larger horizontal area.

Calculations were made to determine the total annual incident radiation for both simple and diffuse isotropic models and also considering different tracking system
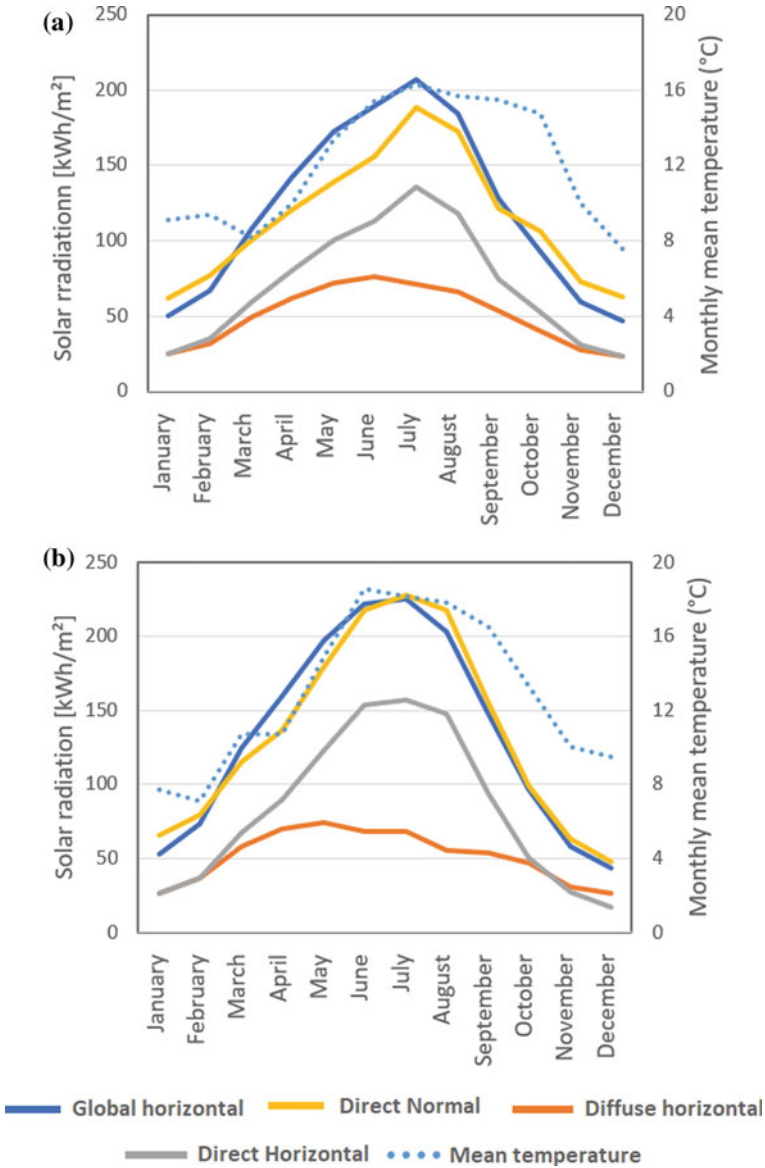
Fig. 2 Components of solar radiation and monthly mean temperature throughout the reference year in the municipality of Braga for: **a** TMY2006 and **b** TMY 2013

Table 2 Comparison of the incident solar radiation between TMY 2006 and TMY 2013, for fixed and different tracking system configurations (simple and diffuse isotropic model)

Atmospheric model	Tracking system configuration	Annual incident solar radiation (kWh/m^2)		Comparison (%)
		TMY 2006	TMY 2013	
Simple isotropic model	Biaxial	1977	2217	12.1
	Polar axis	1924	2151	11.7
	Horizontal axis N-S$_C$	1770	2007	13.3
	Horizontal axis E-W$_D$	1671	1794	7.4
	Fixed (South, $\beta = \phi$)	1601	1736	8.4
Diffuse isotropic model	Biaxial	1907	2152	12.8
	Horizontal axis E-W$_D$	1637	1740	6.3
	Fixed (South, $\beta = \phi$)	1562	1674	7.2

Notes β—surface tilt angle from the horizontal; $\beta = \phi$—tilt angle equal to latitude; N-S$_C$—North–South orientation with continuous adjustment; E-W$_D$—East–West orientation with daily adjustment

configurations. The results are presented in Table 2. Note that, calculations for the polar axis and N-S$_C$ horizontal axis tracking system were only applied to the simple isotropic atmospheric model due to methodology limitations.

When TMY 2013 is compared to TMY 2006, for the same tracking configuration and atmospheric model, it is verified that there is an increase in the forecast of the incident yearly solar energy for the TMY 2013. The increase is more significant with the biaxial, polar axis and horizontal axis N-S$_c$ tracking systems (approx. $+ 12$ to $+ 13\%$), regardless of the atmospheric model used. Regarding fixed surface systems, for an interval of up to 10° in tilt angle, the effect of the tilt angle variation is smaller than the effect of the atmospheric model and to maximize the energy production over a full year the optimal value is 30 to 35°. This means that, for annual optimization, it is preferable to decrease the tilt angle of the fixed solar panels, in order to obtain a higher diffuse radiation values and by comparison to the values normally used in solar thermal systems.

In addition to this analysis, both TMY were compared by forecasting the electricity production for a fixed panel (a 3kWp polycrystalline panel, model AXIpower 270P/60S (year of 2020), south oriented, $\beta = 35°$) and considering the *Hay-Davies*, a robust model that provides good results even when the diffuse radiation component is not fully known. The annual electricity production obtained for TMY 2006 and TMY 2013 corresponded to 229 kWh/m^2 and 246 kWh/m^2, respectively.

4 Final Considerations

The climatic annual database update resulted in the increase of the incident solar energy for any surface orientation. In particular, during the summer months, the peak values for the global horizontal solar radiation are significantly higher for the TMY 2013 when compared with the TMY 2006. The annual mean air temperature is also higher for TMY 2013. When applying the annual production models to the updated and more realistic radiation conditions of the TMY2013, the electricity production is expected to improve by at least 7% with a fixed system and by 11–13% with the best tracking systems.

If the polar tracking system is compared with the biaxial orthogonal incidence tracking system, there difference is only of about 4%. Therefore, polar axis is more economically interesting, given the greater complexity of the biaxial solar tracking system. As future work, it is intended to study the influence of the optimal fixed tilt angle for seasonal applications (e.g., photovoltaic irrigation) and the development of the methodology considering anisotropic atmospheric models.

References

1. IRENA (2020) Renewable energy statistics 2020
2. European Parliament, The Council of the European Union (2021) Directive of the European Parliament and of the Council on the energy performance of buildings (recast), vol 0426, pp 1–79
3. Cebecauer T, Suri M (2015) Typical meteorological year data: SolarGIS approach. Energy Procedia 69:1958–1969. https://doi.org/10.1016/j.egypro.2015.03.195
4. Siu CY, Liao Z (2020) Is building energy simulation based on TMY representative: a comparative simulation study on doe reference buildings in Toronto with typical year and historical year type weather files. Energy Build 211. https://doi.org/10.1016/j.enbuild.2020.109760
5. Ohunakin OS, Adaramola MS, Oyewola OM, Fagbenle RO (2013) Generation of a typical meteorological year for north-east, Nigeria. Appl Energy 112:152–159. https://doi.org/10.1016/j.apenergy.2013.05.072
6. Jiang Y (2010) Generation of typical meteorological year for different climates of China. Energy 35:1946–1953. https://doi.org/10.1016/j.energy.2010.01.009
7. Skeiker K, Ghani BA (2008) Advanced software tool for the creation of a typical meteorological year. Energy Convers Manage 49:2581–2587. https://doi.org/10.1016/j.enconman.2008.05.013
8. Lhendup T, Lhundup S (2007) Comparison of methodologies for generating a typical meteorological year (TMY). Energy Sustain Dev 11:5–10. https://doi.org/10.1016/S0973-0826(08)60571-2
9. Realpe AM, Vernay C, Pitaval S et al (2016) Benchmarking of five typical meteorological year datasets dedicated to concentrated-PV systems. Energy Procedia 97:108–115. https://doi.org/10.1016/j.egypro.2016.10.031

10. Abreu E, Canhoto P, Prior V, Melicio R (2018) Solar resource assessment through long-term statistical analysis & typical data generation with different time resolutions using GHI measurements. Renew Energy 127:398–411. https://doi.org/10.1016/j.renene.2018.04.068
11. Martinopoulos G, Tsalikis G (2018) Diffusion and adoption of solar energy conversion systems: Greece. Energy 144:800–807. https://doi.org/10.1016/j.energy.2017.12.093
12. Trzmiel G, Głuchy D, Kurz D (2020) The impact of shading on the exploitation of photovoltaic installations. Renew Energy 153:480–498. https://doi.org/10.1016/j.renene.2020.02.010
13. Duffie J, Beckman W (2013) Solar engineering of thermal processes, 4th edn. Wiley, New Jersey

Time Series Forecasting of a Photovoltaic Panel Energy Production

Abdelaziz El Aouni and Salah Eddine Naimi

Abstract Predicting the energy production for few days horizon is the key for best managements of photovoltaic residential installations. This paper compares two methods for predicting the power output of solar PV system. We first create a database of around 36,000 points by simulating a PV module using a real weather data (hourly measurements of the temperature and the solar irradiance). To make predictions on the PV panel energy production, two techniques were compared: the combination of the discrete Fourier transform (DFT) and an artificial neural network (ANN), and a well new technique, the long short-term memory (LSTM) time series forecasting with a neural network. The results show an accuracy of the DFT-ANN model around 90% of the energy produced for three days horizon (85% as a minimal sporadic value). For the LSTM method, the accuracy is around 92% for 5 days forecasting.

Keywords PV energy forecasting · Discrete fourier transform · LSTM

1 Introduction

The energy is becoming more and more important and being an indispensable element in human's life. Because of global warming and limited amount of fossil fuel energy stock in the world, it is urgent to find alternative solutions to these sources of energy. Photovoltaic energy is one of them, but unlike conventional power sources, solar energy supply cannot be precisely planned beforehand. This is due to the fact that solar energy is highly dependent on weather conditions, especially cloud structure and day/night cycles. To successfully integrate increased levels of PV power production while maintaining reliability, it is therefore very important to be able to predict the power output. Power output forecasts on multiple time horizons play a fundamental role in storage management of PV systems, control systems in buildings, as well as

A. El Aouni (✉) · S. E. Naimi
Laboratory of Energy, Embedded System and Information Processing, ENSA, Mohammed First University, Oujda, Morocco
e-mail: abdelaziz.elaouni@ump.ac.ma

for the grid regulation and power scheduling. It allows grid operators to adapt the load in order to optimize the energy, transport, allocate the needed balance energy from other sources if no solar energy is available, plan maintenance activities at the production sites, and take necessary measures to protect the production from extreme events.

The PV power forecasting can be performed by several methods, but the most promising is the machine learning models [1]. The choice for the method to be used depends mainly on the prediction horizon; actually, all the models have not the same accuracy in terms of the horizon used [2, 3].

In this paper, we evaluate two techniques for time forecasting of a photovoltaic panel energy production. The first method is a combination of Fourier transformation of sample data and an artificial neural network (DFT-ANN), while the second method uses a well new LSTM network for time series forecasting.

2 Time Series Forecasting of PV Power Output Using Machine Learning

The machine learning models can be used in forecasting problems in three different ways [4]:

- Structural models which are based on other meteorological parameters;
- Time series models which only consider the historically observed data of PV power as input features;
- Hybrid models which consider both, PV power and other variables as exogenous variables.

In this work, we focus on the combination of the DFT with ANN to perform the time series forecasting. For evaluating this technique, we use the LSTM algorithm as reference because it is widely used in such problems.

In fact, it is difficult to compare different methods under different conditions; a small change in input data or models parameters may result in substantial variations in the prediction accuracy of the models. Results of several forecasting methods are proposed in the literature based on ML for solar PV generation as shown in Table 1.

2.1 Generation of a Realistic PV Power Database

To make the best prediction of the PV power output, we need large historical data to train the model on it. This section describes the process flow for creating the database (Power output versus the environmental temperature and solar irradiance). The chosen model, of a PV cell, is the single diode model with both series and parallel resistors for greater accuracy. Modeling this device, necessarily, requires taking weather data

Table 1 A summary of the literature studies

Publications	Models	Time horizon	Error metrics	Results
Lin et al. [5]	Pattern sequence NN	24 h-ahead	MAP, RMSE	77.37, 106.33
Gensler et al. [6]	Autoencoder, LSTM	24 h-ahead	MAP, RMSE	0.0366, 0.713
Wang et al. [7]	CNN, ANN, LSTM	24 h-ahead	RMSE	0.343–1.434
Gao et al. [8]	LSTM-NN	2 days ahead	MAD	1.41% and 3.97%

Error Metrics: Mean Absolute Percentage (MAP), Mean Absolute Deviation (MAD), and Root Mean Squared Error (RMSE)

(irradiance and temperature as a minimal parameters) as input variables. The purpose of the simulations is to get $I(V)$ curve and $P(V)$ curve that can permit to generate the power output of the module. We consider that the global system (PV-load) operates in ideal condition (maximum power point). For the temperature and solar irradiance, we used a real set of data from the NREL Solar Radiation Research Laboratory located at Colorado (USA-39.7423, -105.1785) (irradiance: the Global Normal CMP22 instrument, Temperature: Dray Bulb Temp (Tower) instrument, The data can be freely accessed via https://nrel.gov). All data are collected, theoretically, each minute for several years. The modeled PV module was from JFSOLAR (JFS2-144-545M panel) with the following characteristics: Maximum Power = 545 W, Open-Circuit Voltage = 49.65 V, Short-Circuit Current = 13.92 A, Number of Cells = 6× 24.

2.2 DFT-ANN for PV Power Forecasting

For a digital data, a discrete Fourier transform (DFT) is defined as a method that allows to decompose functions depending on time into functions depending on frequency. The most intuitive application of the Fourier transform is a time series decomposition. Any extrapolation in frequency domain can be interpreted as time prediction but with some risks. However, it is new that the combination of the FT technique with artificial neural network can be used for time forecasting of PV power production [9]. The variation dynamic of the PV power output is controlled by day and night cycles. As the zero power at every night does not provide any real useful informations, we deliberately generate a symmetrical curve by replacing all nights data with data from the previous day. With this transformation, the signal spectrum is less complicated as we will show in the results section. Once this transformation is performed, the DFT of the signal is calculated. Figure 1 shows the algorithm workflow applied to the training dataset. The DFT is estimated each step for N samples and $N + M$ samples. The window is moved each time with a predefined step (P in the figure). As the Fourier transform characteristic is symmetric, we used this property to reduce the number of coefficients (real and imaginary parts). Consequently, two matrices are obtained; the first matrix generated from N samples is used as an input for an
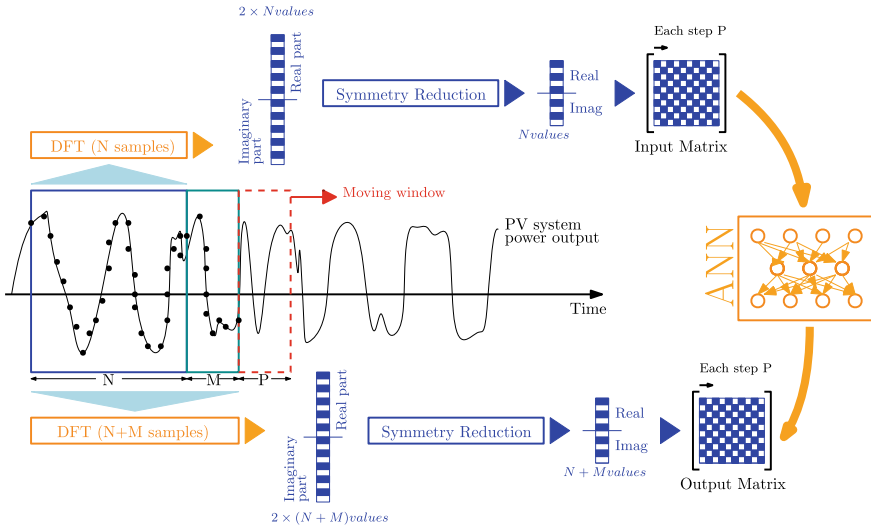
Fig. 1 Prediction steps using an artificial neural network (ANN) trained with data from Fourier transform calculus of a temporal sample

artificial neural network, while the second matrix (obtained from $N + M$ samples) represents the target for the feedforward ANN. Thus, the trained network can be used to estimate each time M values. We need to compute the inverse Fourier transform to be able to predict the next M temporal values of the curve. After testing different configurations, the best neural network model is chosen considering the combination of the RMSE mean errors. This best ANN model is comprised of an input layer of 200 neurons and a single hidden layer of 130 neurons, in addition to 270 neurons for the output layer. For the training algorithm, we use Levenberg–Marquardt algorithm.

2.3 Time Series Forecasting Using LSTM Method

In order to estimate future values of the energy production, a time series forecasting techniques can be applied. The algorithms are generally based on some historical records of a variable to predict the next temporal value. The most significant of these models are the linear autoregressive integrated moving average (ARIMA) model or the autoregressive-moving average (ARMA) model. On the other hand, the nonlinear model such as a recurrent neural network (RNN) architecture is more powerful for complicated series.

Thus, forecasting with long short-term memory (LSTM), which is type of RNN, is well-suited for making prediction. The LSTM network was trained on the same database as the previous algorithm. The LSTM method will not be described further in this section as it is explained in previous studies [10–12].

3 Results and Discussion

Figure 2 shows an example of the simulated characteristics *I–V* of solar panel. We verified that the value of short-circuit current (13.92 A) and the value of open-circuit voltage (48.38 V) are in agreement with the manufacturers' data. After this first validation of the PV model, the weather parameters were used to generate the output power of the panel. Figure 2 shows the real temperature as well as the irradiance variations during a week (1–7 June 2020) as well as the PV power output simulated during the same amount of time. The simulated output power during 1–7 years (with a sampling period of 1 h) is used as a training and validation database for the two methods of power time forecasting. In Fig. 3a, we plot the predicted and actual power target for 11 days (200 points at a sampling period of 1 h); as already mentioned, a symmetric curve is generated by replacing all night's data by those of the previous day. The corresponding signal frequency spectrum is presented in Fig. 3b. The real and imaginary parts of the computed DFT of the signal are used to train the neural network (ANN input). The sequence inputs–outputs are, obviously, normalized for the training, and the outputs de-normalized after the validation step. To predict the
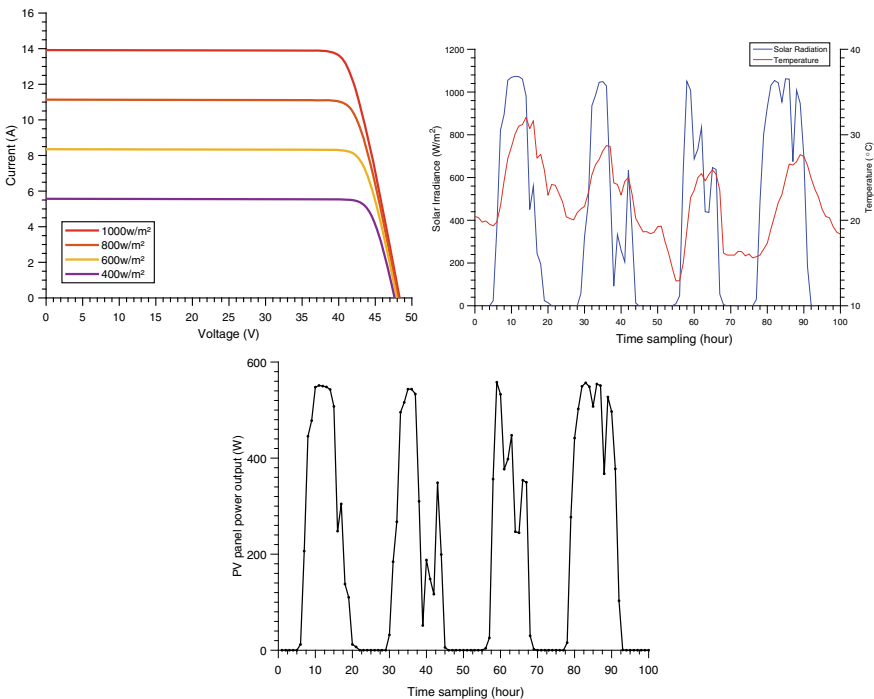
Fig. 2 Results of the PV panel modeling. (Top-left) the *I–V* characteristics at different irradiance, (Top-right) example of weather data from NREL Laboratory, (Bottom) the simulated PV maximal power output for the weather conditions
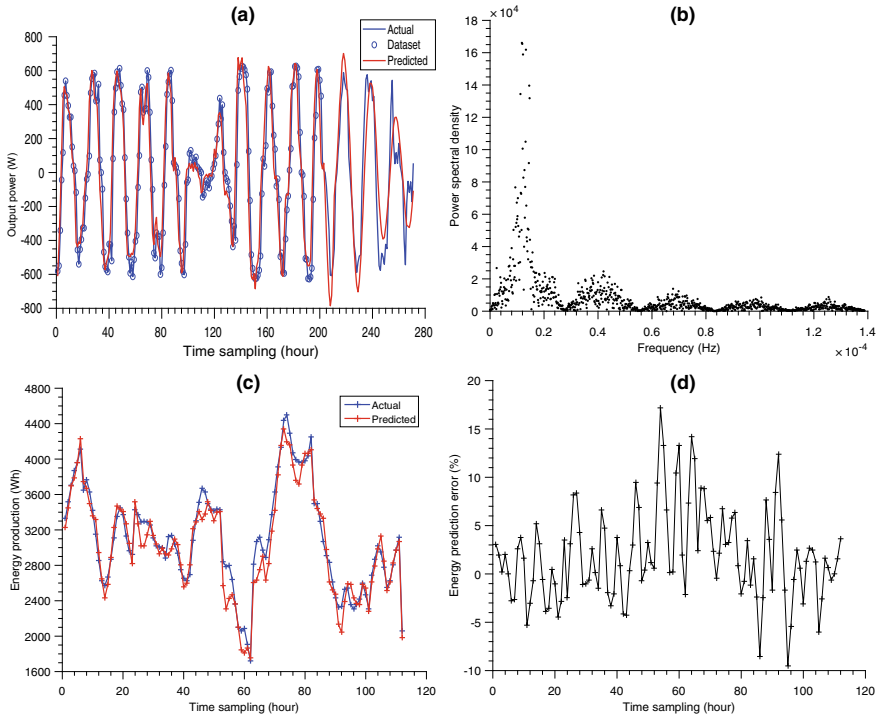
Fig. 3 DFT-ANN results. **a** (Blue curve) the PV power dataset, (Blue circles) the 200 data used at the input of the ANN, (Red curve) the predicted power output (between 200 and 270). **b** An example of the calculated DFT for 200 samples. **c** The actual and predicted energy produced by the PV for 3 days horizon estimated each hour. **d** The relative error of the energy forecasted for the following three days

power of the next hours, the DFT of approximately 14 days (270 sample) is used as a target for the ANN. The dataset is processed as windows of 270 samples that moves in all data length. An example of power time forecasting is shown in Fig. 3a for three successive days. As shown in Fig. 3d, the error in prediction of PV energy production for 3 days horizon is less than 20% (10% in mean). In Fig. 3c, we plot the predicted and actual power target; each point represents the sum of 3 days of PV power production estimated each hour. To evaluate the performance of the LSTM algorithm to forecast PV power, we performed various simulations according to the number of hidden layers, sampling data interval, and the time steps of a deep RNN-based forecasting model. The LSTM network structure with one input layer, one hidden layer of 500 neurons, and one fully connected output layer was the most efficient. In Fig. 4, the graphics represent the predictions for 8 days horizon using (Fig. 4a) 7 years hourly sampling dataset and (Fig. 4b) 2 years dataset for training the network. From Fig. 4, we were able to calculate the relative error of the estimated energy produced after 3 days and 5 days. Therefore, if 7 years dataset is used, the
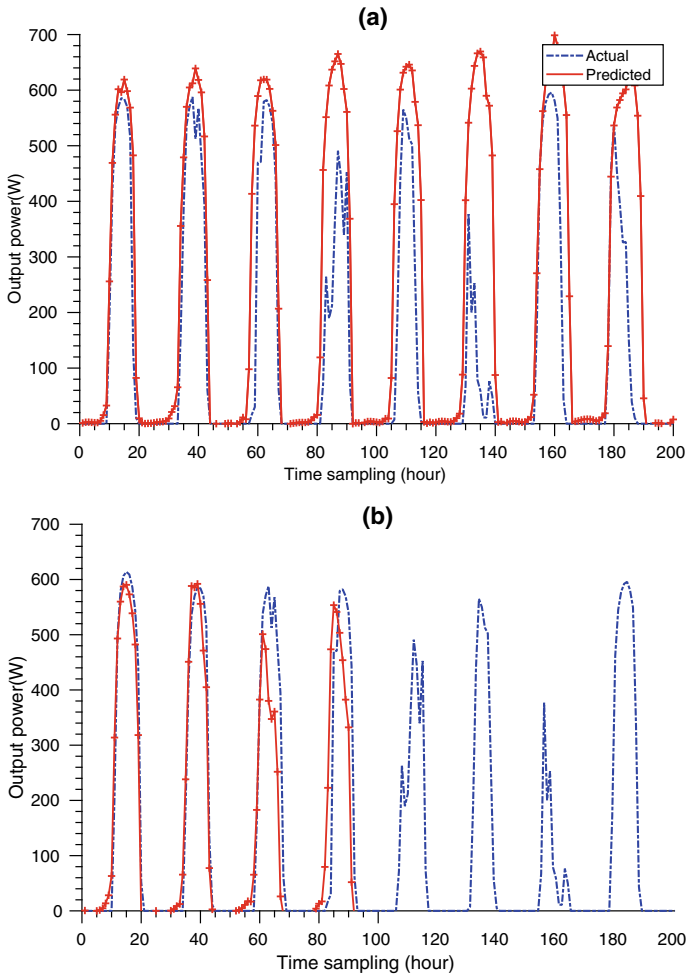
Fig. 4 PV maximal power prediction using LSTM method. **a** Results for 7 years training dataset. **b** Results for 2 years trading dataset

relative error (of the produced energy) is around 30% for 3 days horizon and 50% for 5 days horizon (in cloudy days). In fact, the LSTM network was unable to predict correctly the duration of daylight. However, if we use a 2 years dataset for training the LSTM network, the relative error of the estimated energy drop to around 12% for 5 days forecasting and less than 8% for 3 days forecasting.

4 Conclusions

In this paper, we compare the RNN-based LSTM forecast algorithm to the combined DFT-ANN for PV power generation forecasting. We used on-site weather information and a one diode model of the PV cell to simulate a realistic hourly power production of a commercial panel. The simulation results showed that, for the DFT-ANN algorithm, the prediction accuracy was the best when it consists of one hidden layer with 130 neurons. The input layer accepts a set of 200 data corresponding to Fourier transform of a sampling period of 10 days every hour and produce 270 values in the frequency domain that is equivalent to 3–4 days prediction horizon. The energy produced by the PV module was estimated hourly for the next 3 days with relative error less than 8% (the increase of the error was sporadic only). For the LSTM algorithm, when applied to daily steps PV production, the method was inefficient, while for hourly time forecasting, the LSTM algorithm was able to predict the energy produced by the panel for the following 5 days with a relative error less than 8%.

References

1. Mellit A, Pavan AM, Ogliari E, Leva S, Lughi V (2020) Advanced methods for photovoltaic output power forecasting: a review. Appl Sci 10:487. https://doi.org/10.3390/app10020487
2. Khan W, Walker S, Zeiler W (2022) Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. Energy 240. https://doi.org/10.1016/j.energy.2021.122812
3. Behera MK, Nayak N (2020) A comparative study on short-term PV power forecasting using decomposition based optimized extreme learning machine algorithm. Eng Sci Technol Int J 23(1):156–167. https://doi.org/10.1016/j.jestch.2019.03.006
4. Voyant C, Notton G, Kalogirou S, Nivet ML, Paoli C, Motte F, Fouilloy A (2017) Machine learning methods for solar radiation forecasting: a review. Renew Energy 105:569–582. https://doi.org/10.1016/j.renene.2016.12.095
5. Lin Y, Koprinska I, Rana M, Troncoso A (2019) Pattern sequence neural network for solar power forecasting. Commun Comput Inf Sci. https://doi.org/10.1007/978-3-030-36802-9_77
6. Gensler A, Henze J, Sick B, Raabe N (2016) Deep learning for solar power forecasting an approach using AutoEncoder and LSTM neural networks. In: 2016 IEEE international conference on systems, man and cybernetics, SMC 2016-conference proceedings. https://doi.org/10.1109/SMC.2016.7844673
7. Wang K, Qi X, Liu H (2019) A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. Appl Energy 251(C):1. https://doi.org/10.1016/j.apenergy.2019.113315
8. Gao M, Li J, Hong F, Long D (2019) Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM. Energy 187:115–838. https://doi.org/10.1016/j.energy.2019.07.168
9. Birlasekaran S, Ledwich G (1998) Use of FFT and ANN techniques in monitoring of transformer fault gases. In: 30th symposium on electrical insulating Ma, pp 75–78. https://doi.org/10.1109/ISEIM.1998.741688
10. Abdel-Nasser M, Mahmoud K (2019) Accurate photovoltaic power forecasting models using deep LSTM-RNN. Neural Comput Appl 31:2727–2740. https://doi.org/10.1007/s00521-017-3225-z

11. Chen H, Chang X (2021) Photovoltaic power prediction of LSTM model based on Pearson feature selection. Energy Rep 7(7):1047–1054. ISSN 2352-4847. https://doi.org/10.1016/j.egyr.2021.09.167
12. Hyung Keun A, Neungsoo P (2021) Deep RNN-based photovoltaic power short-term forecast using power IoT sensors. Energies 14(2). https://doi.org/10.3390/en14020436

Study of the Hybrid Solar Energy Supply of a Mobile Service Unit (MSU)

Abdelkrim Laabid, A. Laamimi, A. Saad, and M. Mazouzi

Abstract Mobile Service Units (MSUs) are mobile structures (on truck, semi-trailer, bus, etc.) equipped with material and human resources to relocate complete services and bring them closer to populations far from the official centers (peri-urban and rural areas). These services can be administrative, banking, training, employment, social assistance, support or medical. These services obviously require the deployment of equipment (office automation, IT, equipment specific to the service delivered) which need a reliable and permanently available power supply, something that is not always guaranteed, particularly in the case of rural sites isolated from the electricity network or provided with a low-power network. The quality of services provided by MSU, the reduction in operating costs and the reduction in pollution generated by conventional sources of energy have led to the recent development of alternative sources of clean energy. Our work falls within this context and aims to develop MSU power supply solutions based on hybrid energy sources (PV/batteries). The network (if present on the site) and/or a generating set (GE) provide backup against the intermittent of the PV. We worked on a first version of MSU produced in 2018 and equipped with a 2120 Wp solar roof associated with batteries totaling a capacity of 19,200 Wh to meet the consumption needs of the MSU estimated at around 10,000 Wh per day. We collected and analyzed data from the deployment in the field (solar production, consumption, SOC batteries). We proceeded to a modeling of the energy chain and simulated the behavior on all days of the year. The measurement results are validated, and several PV/battery combination scenarios are proposed and analyzed for better

A. Laabid (✉) · A. Laamimi · A. Saad · M. Mazouzi
LESE, ENSEM, HASSAN II University, Casablanca, Morocco
e-mail: abdelkrim.laabid-etu@etu.univh2c.ma; laabidk1@gmail.com

A. Laamimi
e-mail: laamimi@yahoo.fr

A. Saad
e-mail: a.saad@ensem.ac.ma

M. Mazouzi
e-mail: m.mazouzi@ensem.ac.ma

optimization of the hybrid system produced. The results obtained made it possible to propose an improved solution for the second version of the MSU, which has been completed. It will also be subject to energy monitoring during its deployment in the field.

1   Introduction

The Kingdom of Morocco has given itself, among other things, the priority of reducing social inequalities, in particular for the middle class and the rural areas [1], improving the healthcare system [2] and raising the bar in terms of renewable energies which are thus targeting to exceed the current goal of 52% of the national electricity mix by 2030 [3]. The use of renewable energies concerns all sectors of activity (industry, transport, buildings, agriculture, etc.). In our research and development, we are particularly interested in the integration of renewable energies in Mobile Service Units (MSUs). The MSU is a service delivery unit for populations with no or difficult access to the said services, made up of a mobile team traveling with a vehicle adapted to the geographical and climatic conditions to provide the services. A significant part of the Moroccan population is made up of sedentary populations established in areas far from urban centers (rural and peri-urban world).

This article presents a case study of the integration of photovoltaic energy in a MSU. This work is tackling the concerns part of our research, namely the study with a view to improving the performance of Mobile Service Units (MSUs) manufactured in Morocco.

After a brief review of the scientific literature and industrial concerns on this subject, a state of the art was established and served as a reference for our study. Starting from the established state of the art, we first contributed to the realization of a first version of MSU equipped with a hybrid energy source (PV + battery + emergency generator) (PV field of 2 kWp and battery totaling approximately 20 kWh). We then collected and analyzed the operating data of these units (PV production, MSU consumption, battery SOC, etc.). We proposed a modeling of the hybrid source and developed an algorithm that scrutinizes the energy operation of the MSU on every day of the year. Several PV-battery combination scenarios are presented, analyzed and compared in order to find the optimal solution to meet the load requirement while guaranteeing the longevity of the batteries.

The study carried out and presented will be continued to further optimize the components of the on-board energy source and also to implement an intelligent management approach.

2   Overview of the Literature and the Industrial Context

The issue we are dealing with is current and subject to many developments and advances. Isolated sites and sites connected to networks, unsecured and exposed to power outages and instability, have been powered by hybrid sources generally associated with one or more renewable resources (PV and/or wind power and storage (battery); a diesel generator or the network if it exists is connected in parallel. For example, we can cite the references for inaccessible and away from the distribution network (Houses, Dispensaries, Administration and others) [4–9] which develop hybrid solutions for use in rural or remote area's environment. The MSU (mobile clinics, training units, laboratories, bank branches, etc.) serving these areas is also increasingly equipped with renewable energy, like the sites mentioned. Here, we are talking about on-board energy sources [10–12].

Embedded renewable energies are also developing in the industrial field of transport, in particular in van trucks, commercial vehicles [13], ambulances [14] and caravans [15]. These energies now make it possible to meet the minimum consumption needs of the loads transported, in particular refrigeration [16, 17], air conditioning and the supply of electrical equipment. In general, the electricity produced by PV is used to:

- Reduce diesel consumption through the use of PV panel on the vehicle.
- Reduce CO_2 emissions from diesel generators by, for example, supplying energy for refrigeration when vehicles are parked or on the road.
- Take charge of the air conditioning of buses and utility vehicles.
- Take care of secondary applications such as thawing ice and snow on truck roofs.
- Serve as a power source for electric vans [18].

To further improve the performance achieved and the levels of safety required, research efforts are continuing and are aimed at:

- The realization of light and efficient PV modules, resistant to vibrations and to shearing and bending forces.
- Ease of installation at low heights.
- PV high-voltage systems that meet the safety requirements of PV and vehicle standards.
- Larger storage capacity with minimum weight [19–21].

Presentation of the studied MSU and their energy needs.

2.1   General

In its development strategy, the National Agency for the Promotion of Employment and Skills (ANAPEC) acquired and put into service in 2019 the first MSU which are real mobile agencies intended for rural and peri-urban populations. The core values

of this development strategy are equity and proximity to populations. A flexible and tailor-made service is provided to young job seekers by mobile advisors made available to them by ANAPEC.

Figures 1 and 2 show the exploded view of the interior of the truck-type agency and its solar roof.

The power demand is calculated from the power balance of the various equipment and components of the MSU considered (agency on truck). It is rated at 4.3 kW. The daily energy needs are also estimated. The average consumption is estimated at 10 kWh.

Fig. 1 Equipment for MSU—interior
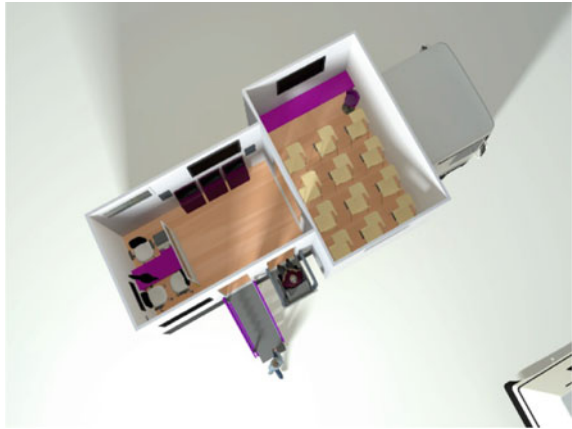
Fig. 2 Solar photovoltaic array mounted on the roof of the truck

The surface available on the roof of the vehicle to accommodate the PV panels is equal to 17.5 m². This surface made it possible to install 8 panels with a unit power of 265 Wp, i.e., a total of 2120 Wp (Fig. 2). Based on an average producible of 1700 kWh/kWp, the expected average daily production is estimated at 10 kWh/day.

2.2 Assessment of Battery Needs

It is based on daily operation of 6 h with an average power of 60% of the estimated nominal power, i.e., a daily consumption of 0.6 * 4340 * 6 = 15.6 kWh. Then, the capacity of the necessary accumulators is estimated by the formula:

$$C = (Ec * N)/(D * U)$$

C: battery capacity in ampere-hours (Ah).
Ec: energy consumed per day (Wh/d) (according to the power balance).
N: number of days of autonomy.
D: maximum permissible discharge (0.8 for lead batteries).
U: battery voltage (V).

By choosing $U = 24$ V with a depth of discharge of 80%, one day of autonomy would lead to a capacity of 15,600 * 1/(0.8 * 24) = 812.5 Ah or 19,500 Wh (Fig. 3).

The hybrid system presented has been installed on the MSU and tested before deployment in the field since 2019. It has been subject to regular monitoring, the results of which we have reported and analyzed with a view to proposing solutions for improvement.

Skirt 1
Diesel Fuel Tank Generator Set.
Truck Battery.

Skirt 2
Generator Set.
Air Conditioning Condenser.

Skirt 3
Solar Battery.
2 Inverters.

Skirt 4
Solar Battery – 1 Inverter.
1 Regulator – Electrical Box.

Skirt 5
Side Staircase.

Skirt 6
Extinguisher – 2 Water Tanks – AdBlue.
Power Supply Cable – Grounded Spike.
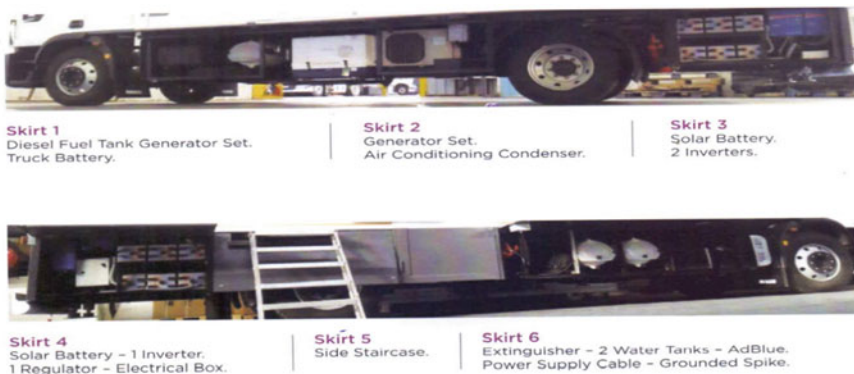
Fig. 3 Diesel generator/solar battery/inverter/regulator are mounted in the skirts of the truck
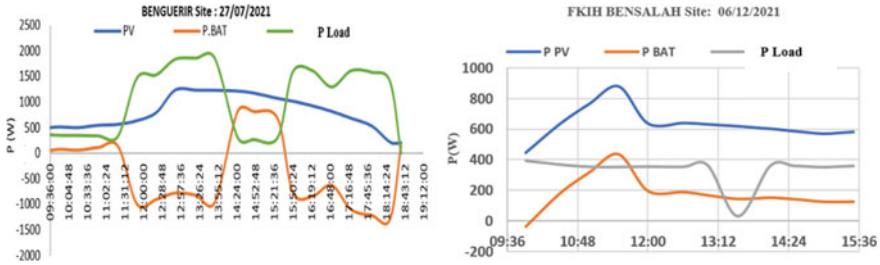
Fig. 4 Examples of results obtained in the Marrakech-Safi region and the Beni Mellal-Khénifra region

3   Analysis of MSU Energy Monitoring Results

The monitoring focused mainly on the powers produced by the PV source and that consumed by the MSU as well as on the SOC of the battery, through recordings and measurements taken at time intervals of a few minutes during the operation of the MSU. We have processed and analyzed the data collected and those in the different geographical areas served by the MSU.

Below are two examples of results obtained in the Marrakech-Safi region and the Beni Mellal-Khénifra region (Fig. 4).

We were able to follow the MSU over almost a year in order to analyze the impact of sunshine which affects PV production and the impact of the HVAC system which affects the load. We noted that over the entire period explored, it was not necessary to use the network or the GE. The hybrid power system only ran on PV backed up by the batteries.

4   Assessment of the Performance of the PV Source on Board the MSU

To assess the performance of the PV modules, we compared the results of measurements carried out on the MSU during its operation in Ben Curer on 07/27/2021 and those generated by simulation using the PVGIS software [22]. The coordinates of the site (32.14 N latitude; 7.96 E longitude, height of 474 m) are entered into the calculation software, as well as the power of the PV source. The theoretical production curve is finally compared with the measurements of the day in question.

Figure 5 shows the measured production profile and that calculated with PVGIS. The observed difference is attributed to local weather conditions which would not be identical to those of the same day given by PVGIS [22], to aging and to dirt deposited on the PV modules.

We also compared the SOC calculated from the initial battery charge and the production given by PVGIS to the SOC measured (Fig. 6).
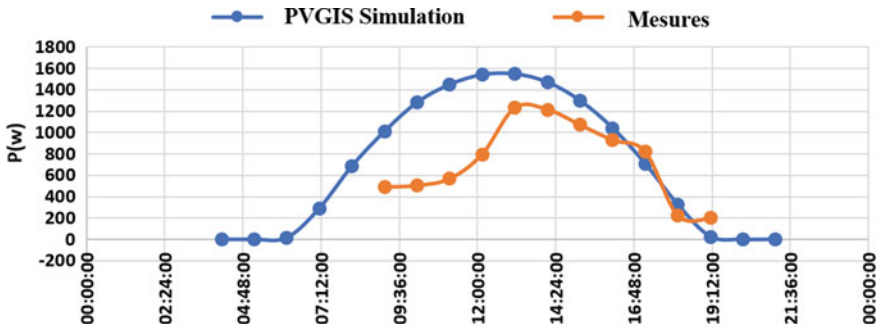
Fig. 5 PV production measured versus production given by simulation with PVGIS on 27/07/2021
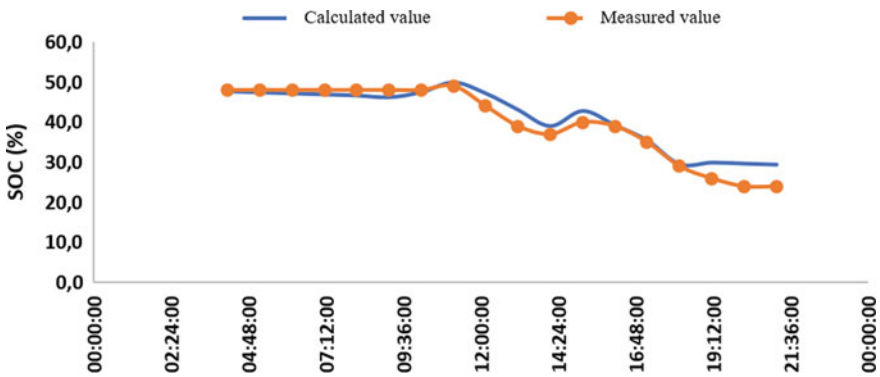
Fig. 6 Evolution the SOC battery during the day of 07/27/2021

During the day of observation, the battery charge was initially at a level slightly below 50%. At the end of the day, it went almost to the SOC limit of 20%. We will come back to this situation later.

5 Modeling of the MSU Hybrid Energy Supply Source

The design of the MSU on-board power source solution is based on an average annual model. Daily, weekly and monthly variations are not taken into consideration as well as the impact on the charge and discharge cycles of the battery. In order to achieve an optimal solution, we proposed a modeling of the hybrid system and analyzed several PV/battery combinations in order to better reduce the call on the network and/or generator.

5.1   Presentation of the Studied System

The hybrid energy source that equips the studied MSU is composed of:

- PV modules covering the roof of the vehicle.
- The energy storage battery.
- The hybrid inverter.
- The generator.
- The cable reel that can be connected to the electrical network when possible.

The source is schematized in Fig. 7.
PPV: Power supplied by the PV source.
PR: Power supplied by the network or the generator (GE).
PB: Power exchanged with the battery.
PC: Power called by the load.
At any moment:

$$PC = PPV + PR + PB$$

The load is supplied primarily by the PV source with the following scenarios:
If the need is satisfied, the surplus is used to recharge the battery until its maximum level is reached, and there, the residual is not recovered (energy not recovered).
If the need is not satisfied by the PV, the battery provides the rest as long as its charge EB (expressed in Wh) is greater than the minimum threshold EBm associated with the battery.
If the need is not satisfied by the PV and the battery reaches EBM, the power grid or the generator set comes into action to satisfy the use and charge the battery.
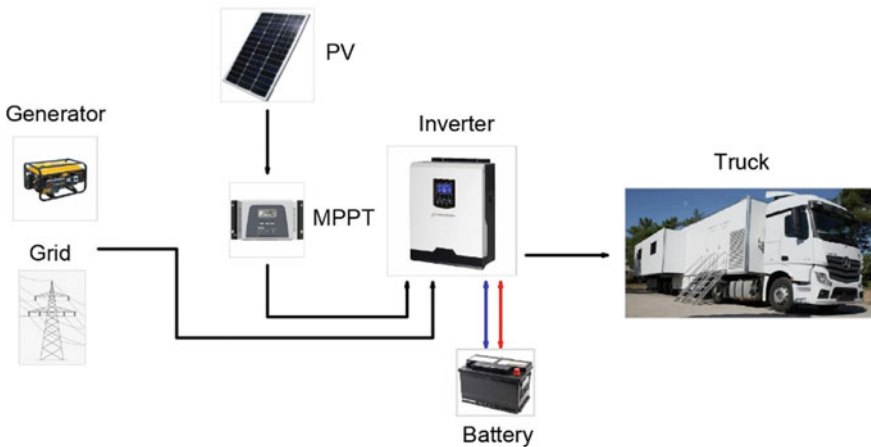
Fig. 7 PV/diesel generator/grid/battery hybrid system configuration

5.2 Functional Modeling

- The operation of the system is modeled over an entire year on the basis of:
- Daily solar production (hourly data: from midnight to 11 p.m. with a step of $h = 1$ h) provided by the PVGIS simulation (free tool for simulating solar production anywhere in the world). The PV power is given in the form of a column matrix of $326 \times 24 = 8760$ elements.
- The load profile constituted by the MSU (hourly data) given with the same number of points. PC is a column matrix of 8760 elements.
- The battery initially charged to its maximum energy EBM (expressed in Wh).
- The time matrix T also contains 8760 points, successively 1, 2, 3, …, 24, 25, …, 8760 (Fig. 8).

5.3 Operation Simulation

The simulation is based on:

- The production of a 2120 Wp solar field placed on the roof of the vehicle, provided by PVGIS, operating in the geographical area of BENGUERIR.
- The average daily consumption profile of the MSU.
- Storage batteries totaling energy of 19,200 Wh.

The objective is to evaluate the performance of the solution and explore more advantageous PV-storage association scenarios.

For this purpose, we considered three storage capacities: installed (19,200 (Wh)), 40,000 (Wh) and 31,680 (Wh).

We compared the power produced to the power consumed throughout the year. Figure 9 illustrates the results obtained.

It is clear that throughout the year, the need for consumption is largely met by PV production. However, the seasonal fluctuation of PV production generates either a surplus of production (case of April and July) or a deficit (case of December). The adjustment between production and consumption must be best fulfilled by the battery.

The battery must ensure the assigned function under the best conditions, in particular:

- Limit the use of the network or the generator.
- Limit the number of charge–discharge cycles for better battery life.

The impact of battery capacity on reducing the number of charge/discharge cycles, and therefore on battery life, is studied by comparing three battery capacities: 40,000, 31,680 and 19,200 Wh (the one installed). The power of the PV field is maintained at 2120 Wp in the three situations. Figures 10, 11, and 12 show the balance over the year of the hybrid system with the three battery capacities (Table 1).

The following observations emerge from the summary table:
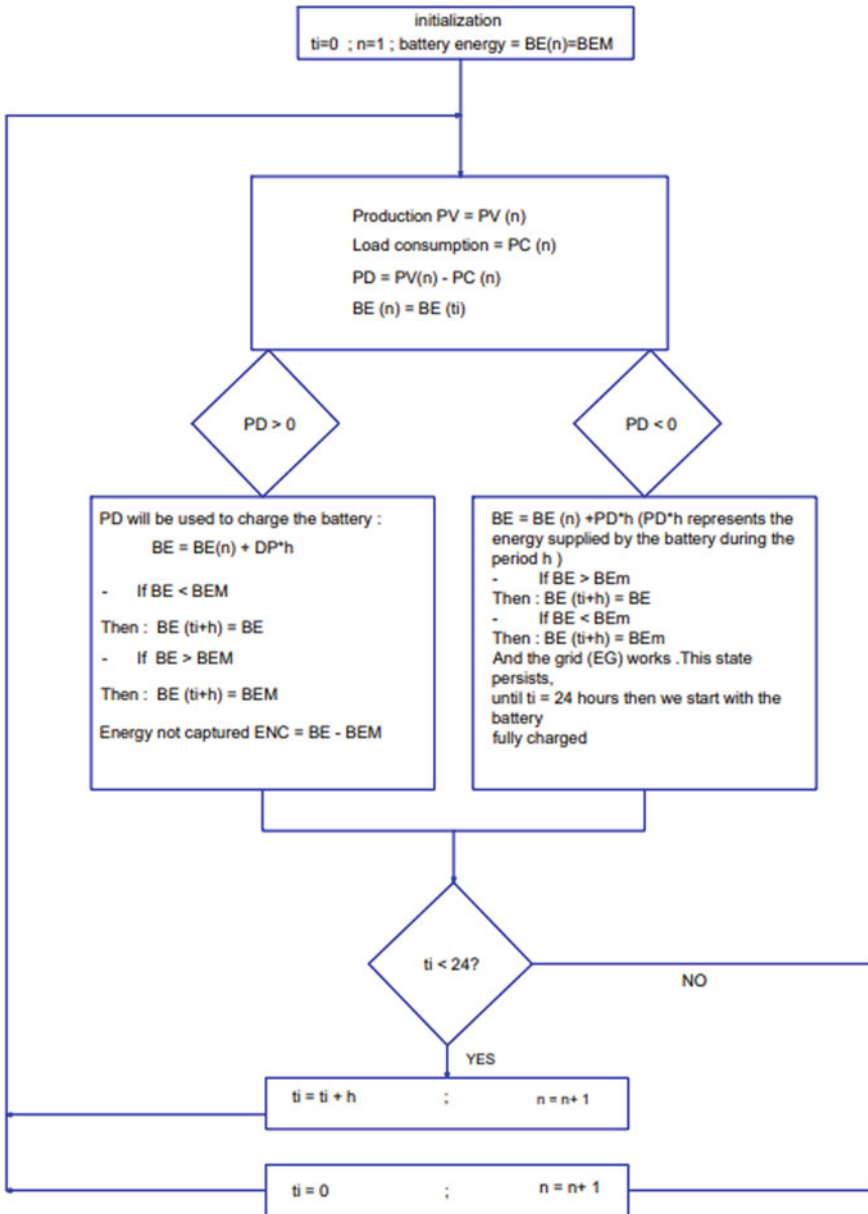
Fig. 8 Flowchart used for system modelization and simulation of the energy management system
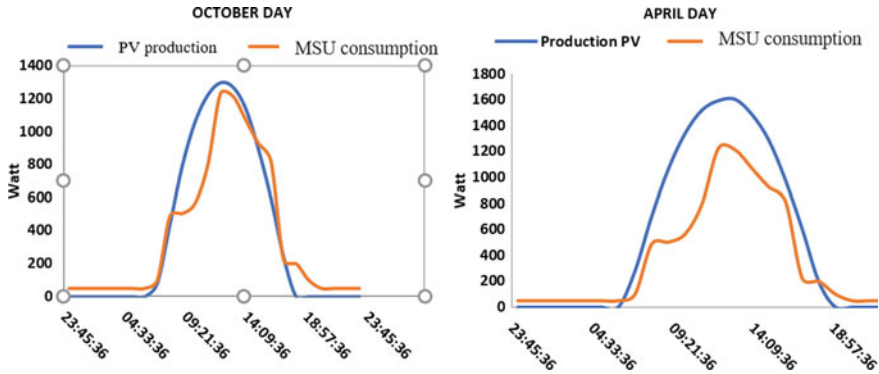
Fig. 9 Daily profiles of PV production and MSU consumption

Fig. 10 Annual operating of the hybrid PV-storage system 19,200 Wh battery
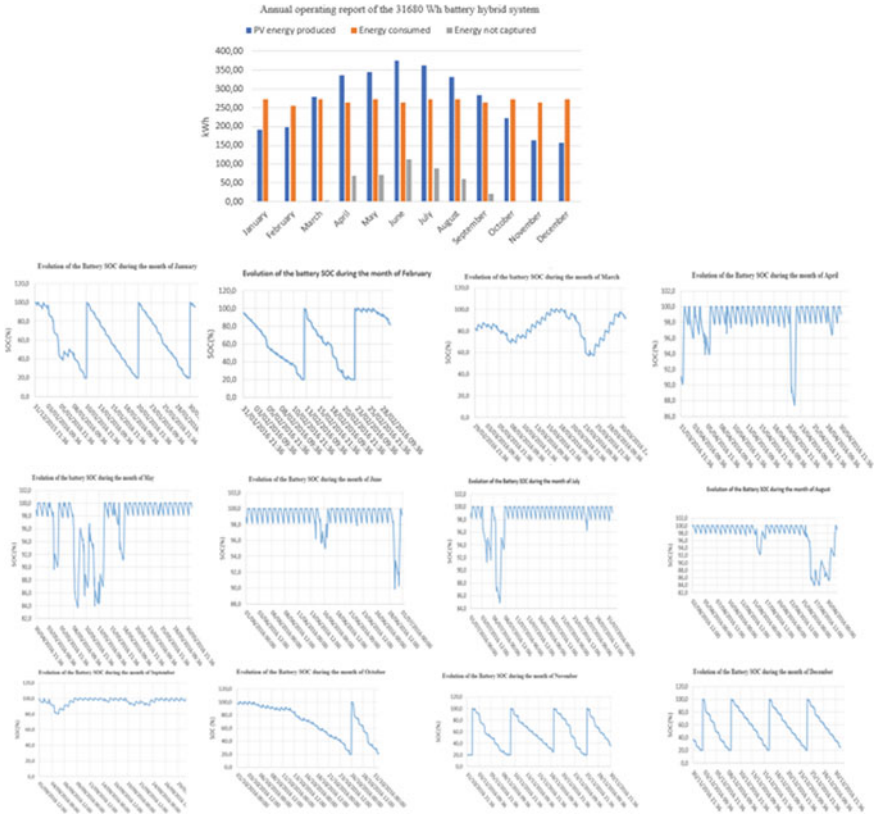
Fig. 11 Annual operating of the hybrid PV-storage system 31,680 Wh battery

- The increase in battery capacity leads to a reduction in the number of charge/discharge cycles as well as the depths of discharge. We see that when we double the capacity, we almost halve the number of cycles.
- On the other hand, the capacity does not significantly affect the uncaptured energy which remained around 435 kWh/year in the three cases.
- This approach helps to choose the optimal battery capacity based on the manufacturer's specifications and the desired lifetime of the batteries.

6 Conclusion and Prospects

Following the production of a first MSU version equipped with a PV/battery hybrid source backed up by a generator, we were able to validate the concept developed in the field. Operational monitoring in the field confirmed the expected performance. The slight discrepancies observed between the measurements and the theoretical
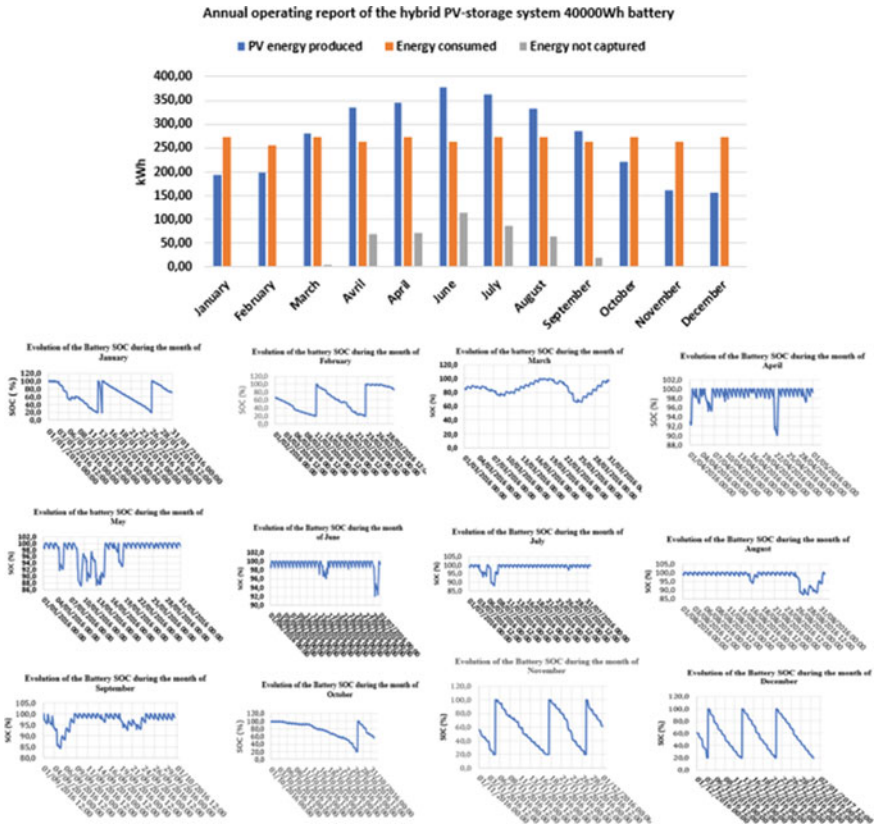
Fig. 12 Annual operating of the hybrid PV-storage system 40,000 Wh battery

Table 1 Summary of the results obtained

Power (Wc)	2120		
Battery capacity (Wh)	40,000	31,680	19,200
Energy produced per year (kWh)	3248.4	3248.4	3248.4
Energy consumed per year (kWh)	3219.7	3219.7	3219.7
Energy not captured per year (kWh)	434.3	433.09	443.99
Percentage of energy not captured (%)	10.3	13.33	13.66
Number of deep charge/discharge cycles	12	13	21

simulation are mainly due to the cleanliness of the panels and the weather conditions. The modeling of the operation of the hybrid source and the simulation of several PV/battery combination scenarios provide information on the optimal solutions to be achieved with the best compromise between the size of the PV field, the capacity of the batteries and the minimization of unrecovered energy. In addition to the financial

constraints, the weight of the batteries and the limited reception surfaces of the PV modules must be the subject of an optimization study planned in the continuation of this work. The new MSU completed at the end of 2021 has already benefited from feedback from previous versions. Equipped with systems sized with a PV power of 3540 Wp, they are monitored and the results will be presented in a future article. The same will apply to MMUs (Mobile Medical Units) completed in 2020 whose operation is delayed following the COVID-19 pandemic. The performance of their hybrid energy source will be evaluated and optimized.

References

1. Royal speech of August 20, 2019, Vidéo. Discours du 20 août: le roi détaille la feuille de route du nouveau modèle de développement | le360.ma. Accessed 30 Aug 2019
2. Rapport_General.pdf (csmd.ma) 19
3. https://www.medias24.com/MAROC/Les-plus-de-Medias-24/187243-Le-Roi-Mohammed-VI-preside-une-seance-de-travail-sur-les-energies-renouvelables.html. Accessed 19 Dec 2021
4. Das BK, Zaman F (2018) Performance analysis of a PV/diesel hybrid system for a remote area in Bangladesh: effects of dispatch strategies, batteries, and generator selection. Energy. https://doi.org/10.1016/j.energy.2018.12.014
5. Rad MAV, Vaghar MP (2022) Techno-economic evaluation of stand-alone energy supply to a health clinic considering pandemic diseases (COVID-19) challenge. Sustain Energy Technol Assess 51:101909
6. Kamal MM, Ashraf I (2022) Planning and optimization of hybrid microgrid for reliable electrification of rural region. J Inst Eng (India): Ser B 103(1):173–188. https://doi.org/10.1007/s40031-021-00631-4
7. Rout KC, PS Kulkarni (2020) Design and performance evaluation of proposed 2 kW solar PV rooftop on grid system in Odisha using PVsyst. In: 2020 IEEE international students' conference on electrical, electronics and computer science
8. Haffaf A, Lakdja F (2019) Electrification d'une charge isolée d'agriculture par hybridation énergétique. Revue des Energies Renouvelables 22(1):1–17
9. Das BK, Hasan M (2021) Optimal sizing of a grid-independent PV/diesel/pump-hydro hybrid system: a case study in Bangladesh. Sustain Energy Technol Assess 44:100997
10. https://timm-sante.com/equipements/camion-derniere-generation/
11. https://www.victronenergy.com/markets/automotive/mobile-clinic-ghana
12. https://www.financialtechnologyafrica.com/2017/08/22/inlaks-designs-builds-worlds-first-solar-powered-mobile-bank-branch-for-wema-and-stanbic-ibtc-banks/
13. https://www.transportealdia.es/dhl-trailar/
14. https://www.ems1.com/ems-products/ambulances/articles/ambulance-solar-panels-save-money-and-reduce-environmental-impact-dYD23JAXFCp3HcMU/
15. https://newatlas.com/outdoors/retreat-all-electric-e-rv-caravan/
16. https://interestingengineering.com/solar-powered-trucks-can-tackle-refrigeration-emissions
17. https://www.tipeurope.com/knowledge-and-news/knowledge/are-solar-powered-trailers-the-next-new-thing
18. https://www.ise.fraunhofer.de/en/business-reas/photovoltaics/photovoltaic-modules-and-power-plants/integrated-photovoltaics/vehicle-integrated-photovoltaics/vehicle-integrated-trucks.html
19. https://ecoinventos.com/paneles-solares-en-la-cubierta-de-un-camion/
20. https://www.pveurope.eu/solar-modules/vehicle-integrated-pv-electrical-commercial-vehicles

21. https://energy.soton.ac.uk/keeping-food-chilled-by-solar-power
22. https://ec.europa.eu/jrc/en/pvgis. Accessed Dec 2021

Performances Comparison of PV/T Solar Plants with Roll-Bond and Sheet-and-Tube Absorbers

Chaimae El Fouas, Mohamed Hajji, Oussama El Manssouri, Hajji Bekkay, Antonio Gagliano, and Giuseppe Marco Tina

Abstract Photovoltaic/thermal unit (PV/T) is an innovative technology that allows the renewable energy production by cogenerating electricity and heat simultaneously. This technology is seemed promising for the energy supply in many applications, especially for building envelope. In the present work, the performances of two different PV/T solar plants with a Roll-bond and PV/T sheet-and-tube absorbers are compared under the operating conditions of Catane, Italy. The performance assessment of theses PV/T solar plant is ensured by the dynamic modeling of two PV/T in MATLAB software. Moreover, the PV/T energy yields for suppling the energy need of a simple building are evaluated. From results, the comparison of the two PV/T systems proves that the "Roll-bond" configuration provides a higher thermal performance with a maximum fluid outlet temperature of 42.13 °C with respect to the standard one of 40 °C "Sheet-and-tube" configuration). In terms of the PV/T energy yields used for the energy supply of a standard building, it's observed that PV/T" Roll-bond" system is very promising. It can fully cover the energy needs of the building with a total coverage factor of 76.79%.

Keywords PV/T solar plant · Absorbers · Roll-bond · Sheet-and-tube · Energy yields · building needs

1 Introduction

PV/T technology offers many attractive options since it merges a PV and solar thermal technologies in the same unit, for producing electricity and heat simultaneously. This

C. El Fouas (✉) · M. Hajji · O. El Manssouri · H. Bekkay
Laboratory of Renewable Energy, Embedded System and Information Processing, National School of Applied Sciences, Mohammed First University, Oujda, Morocco
e-mail: c.elfouas@ump.ac.ma

A. Gagliano · G. M. Tina
Department of Electric, Electronics and Computer Engineering, University of Catania, Catania, Italy

hybrid unit is seen promising, in particular for building applications, due to its multiples advantages [1]. In the PV/T system, different fluids can be used for the cooling effect [2]. Besides, an effective cooling can be offered by the suitable choice of the absorber configuration. The most conventional PV/T system configuration is with Sheet-and-tubes one, where a parallel tubes are welded with metal plate. In recent years, a developed similar PV/T with Roll-band configuration is obviously adopted. It's characterized by two metal absorbers combined by a rolling process. These two PV/T systems configurations are widely investigated in literature, for further improve their performances. For the PV/T Sheet-and-tubes, a dynamic numerical model and a new bi-dimensional finite difference numerical model have been developed and validated experimentally by [3, 4]. The obtained results show a very good agreement with the experimental data in case of stable and transient weather conditions. For the work of [5], a new 3D dynamic model for the Sheet-and-tube PV/T water system is developed to evaluate the electrical energy production and the supply of hot water under a real climatic condition. This model fits well with the measured data published by the Eurofins laboratory. For the PV/T Roll-bond, the performance of a glazed Roll-bond PV/T system is evaluated by [6]. Hence, excellent performances are noted for this PV/T system. The work of [7] investigates a one-dimensional model of an unglazed and uninsulated Roll-bond PV/T system, using the numerical finite volume approach. A good agreement is noted with the experimental data. Another study of unglazed Roll-bond PV/T system, whose model is implemented in the TRNSYS environment is conducted by [8], considering the data of three sites in Europe (Paris, Milan and Athens). Moreover, the experimental results obtained prove the trends observed in the simulation of PV/T model.

A few studies are made on the comparison of these PV/T systems: Roll –bond and Sheet-and-tube configurations. In fact, it is more challenging, due to the very different geometric characteristics of the channel's configurations. The work of [9] have experimentally evaluated the performance of three different configurations of PV/T system, under real operating conditions at the SolarTech LAB laboratory (Milan, Italy): PV/T Sheet-and-tubes and PV/T Roll-bond with a different insulation. The results show that the PV/T Roll-bond is the most effective one. New prototypes of PV/T Roll-bond systems were presented and compared with a standards PV/T Sheet-and-tubes systems as reported by [10]. The thermal performances of four PV/T systems are investigated theoretically and experimentally, in the steady-state and quasi-dynamic regime. A higher thermal efficiency of about 82% was noted for the PV/T Roll-bond system with a black coating. The present study aims to determine the best PV/T solar plant configuration (with Roll-bond or Sheet-and-tubes absorber) allowing a good performance. It is underlined the advantage of integrating the best PV/T configuration on a simple Italian building to cover its energy needs.

2 PV/T Solar Plants Modeling

In this section, the considered PV/T solar plants are characterized and described. Both PV/T solar plants are equipped with: two unglazed PV/T panels connected in series with an inclination of 45° [11], hydronic circuit including the thermal storage tank, the pump and pipe connection. The difference between the two solar plants is the PV/T panel (the absorber) as shown in Fig. 1. In first case, the thermal absorber is designed with a conventional Sheet-and-tube configuration. While in second one, the absorber is in Roll-bond design.

2.1 *PV/T Thermal Models*

In this study, the PV/T models are developed similarly to previous literature models, based on the energy balance established in MATLAB. The PV/T Roll-bond model is built and validated as reported in [11, 12]. While the PV/T Sheet-and-tubes model consists an extension of the model presented in [13–17] and validated in [14, 15].

The energy balance for each PV/T design includes the different heat exchanges by radiation, convection and conduction. Figure 2 presents an equivalent electrical circuit of the heat flux trough the studied PV/T panels. In case of the Roll-bond PV/T design (a), the considered layers are: the glass cover, PV cells, Tedlar, Roll-bond absorber, and the insulation. While in the PV/T Sheet-and-tube one (b), the Roll-bond absorber is replaced by a metal plate and a set of tubes. The meteorological variables, e.g., the solar irradiance, the ambient temperature and the wind speed are integrated in the developed PV/T models as well some hypotheses are used for the
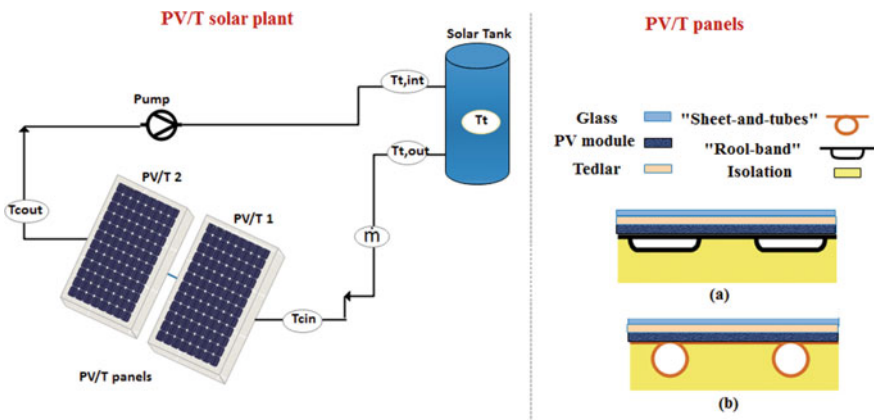
Fig. 1 Schematic diagram of the whole PVT solar plant and the PV/T panels: **a** with roll-bond configuration, **b** with sheet-and-tubes configuration

(a)

(b)
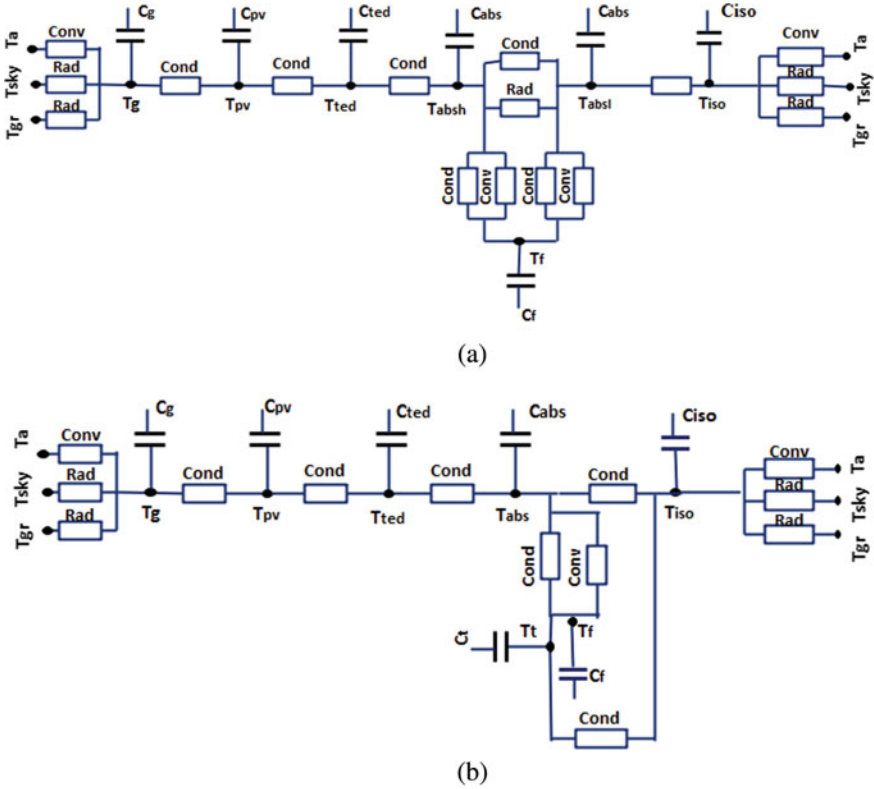
Fig. 2 Equivalent electrical circuit of the PV/T panels: **a** with roll-bond absorber, **b** with sheet-and-tubes absorber

calculation as shown in [13–15]. Figure 3 shows the solar irradiance (G) and ambient temperature (T_a) profiles used in the simulation for the period 7–10 Mars 2019.

2.2 PV/T Energies Production

The electrical power production (P_e) of the PV/T solar plant is calculated from the solar irradiance at the PV/T panel surface (G), the total surface of the PV cells (A_{PV}) the reference electrical efficiency η_{ref} and the PV cells characteristic β as presented in [1]

$$P_e = \eta_{ref}[1 - \beta(T_{PV} - 25)]A_{PV}G \tag{1}$$

The net energy produced $E_{el,ne}$ depends on the efficiency of the inverter device η_{inv} considered equal to 95%, electrical efficiency of the panel η_e and P_{pump} power
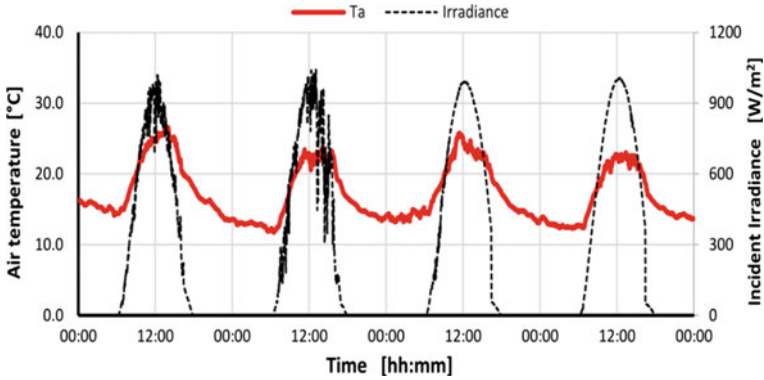
Fig. 3 Meteorological data: ambient temperature (T_a) and solar irradiance (G)

required for pumping the coolant fluid.

$$E_{el,ne} = \int (\eta_{inv} P_e - P_{pump}) dt \tag{2}$$

The thermal power (P_{th}) and the thermal energy is estimated by:

$$P_{th} = \Delta C_f (T_{c,out} - T_{c,int}) \tag{3}$$

where $\dot{m}$, C_f and A, $T_{c,int}$, $T_{c,out}$ are, respectively, the fluid mass flow rate, heat capacity, panel area and temperature of inlet and outlet of the PV/T panel.

The energy-saving for the DHW production (hot water demand) is calculated using Eq. (5).

$$E_{DHW,PVT} = \int_0^t \dot{m}_{DHW} C_f (T_{m,out} - T_{sup}) dt \tag{4}$$

where $\dot{m}_{DHW}$, $T_{m,out}$, T_{sup} are the mass flow rate for DHW, temperature of the water from mixing valve and temperature of the water from the mains.

The energy needs coverage factors, as function of electrical, thermal and total energies needs of the building are:

$$F_{el}(\%) = 100 \cdot \frac{E_{el,ne}}{E_{el,load}} \tag{5}$$

$$F_{th}(\%) = 100 \cdot \frac{E_{DHW,PVT}}{E_{DHW,load}} \tag{6}$$

$$F_{tot}(\%) = 100 \cdot \frac{E_{tot}}{E_{tot,load}} \tag{7}$$

3 Results and Discussion

3.1 *Main Key Temperatures Profiles*

The two PV/T solar plants are describing as a set of overlapping parts including: the glass, PV cells, Tedlar, absorber, fluid, and solar tank. Figure 4 shows the hourly PV temperature variation of the studied PV/T configurations. Indeed, a lower PV temperature is observed for the PV/T Roll-bond (the mean maximum PV temperature is about 39.25 °C). This result underlines the effective cooling process ensured by the Roll-bond design, allowing more heat extraction from the PV cells compared to the Sheet-and-tube geometry.

The fluid temperatures at the outlet of the PV/T panels for each configuration are plotted in Fig. 5. For the same inlet water temperature entering the PV/T panels [11], the outlet temperature presents a logical trend. It can be observed that the outlet temperature of the PV/T Roll-bond configuration is higher at about 43.0 °C. While in case for PV/T Sheet-and-tube one is about 39 °C. These results can be explained by the Roll-bond design transferring more heat between the PV cells and the fluid compared to the Sheet-and-tubes design.

Figure 6 shows the mean temperatures of the water in the solar tank. The Roll-bond configuration highlight a higher water temperature inside the tank compared with that of Sheet-and-tubes configuration. The maximum mean water temperature reached in the third days at around 15h:00mm is 40 °C for the PV/T Roll-bond and 37.1 °C for the other one. These results could be attributed to the increase in the temperature of the water living the PV/T panels and enter in the solar tank for the Roll-bond configuration. Further analysis to assess the performances of the two compared PV/T plants is presented thereafter.
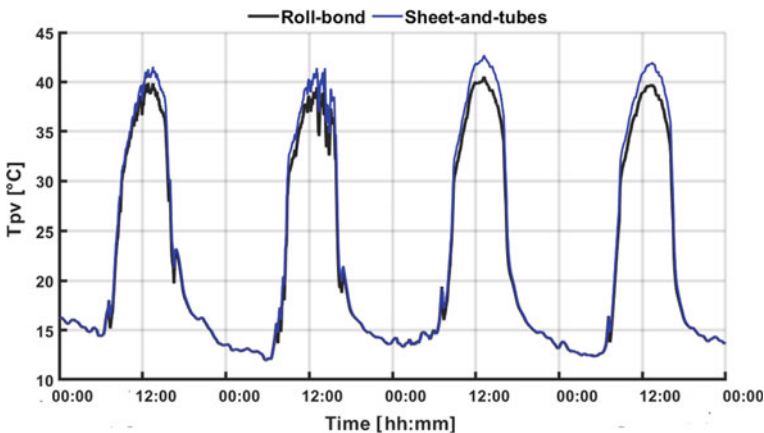
Fig. 4 PV temperature in the PV/T panels: cases of roll-bond configuration (black line) and sheet-and-tubes configuration (blue line)
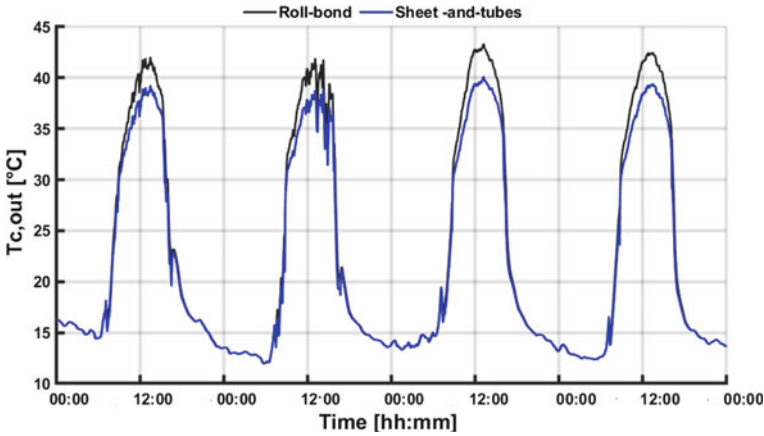
Fig. 5 Fluid temperature at the PV/T panel outlet ($T_{c,out}$): cases of roll-band configuration (black line) and sheet-and-tubes configuration (blue line)
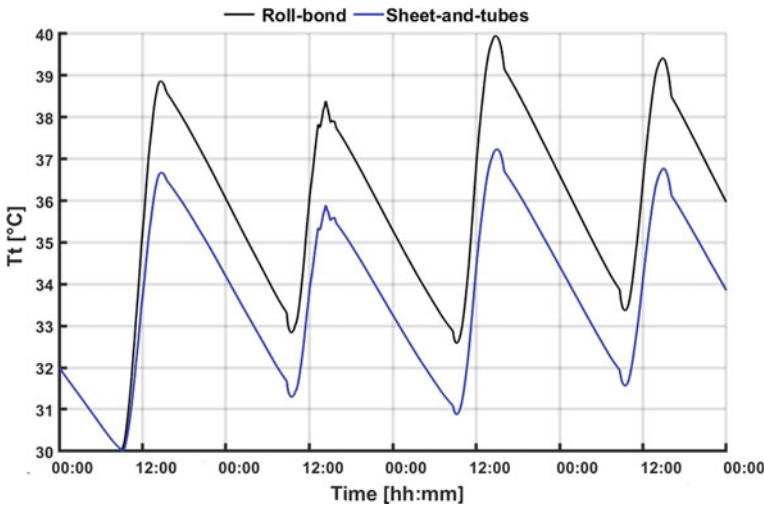
Fig. 6 Temperatures in the solar tank (T_t): cases of roll-band configuration (black line) and sheet-and-tubes configuration (blue line)

3.2 Electrical and Thermal Energy Production

The daily electrical (P_e) and thermal (P_{th}) powers profiles of the two PV/T solar plants is shown in Fig. 7. Obviously, at midday with a value of solar irradiance of 1000 W/m^2, the solar plants provide the maximum electrical and thermal powers. The Roll-bond configuration shows a better electrical and thermal production with mean maximum values of 0.9 kW and 0.45 kW, respectively, for the thermal and electrical
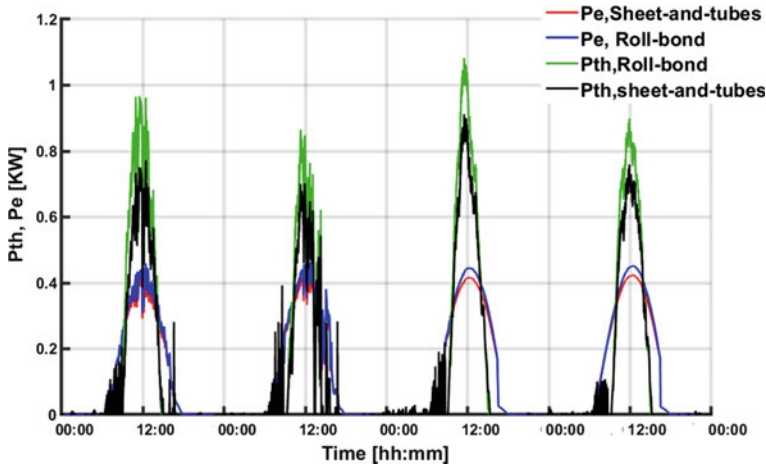
Fig. 7 Electrical and thermal power of the PV/T solar plant: cases of roll-band configuration (black line) and sheet-and-tubes configuration (blue line)

production (0.7 and 0.4 kW in case of Sheet-and-tubes design). For evaluating the daily performances of the two PV/T solar plants (on 7 March) working with a constant mass flow rate of 6 l/min, a synthetic resume of the energy performance of theses PV/T solar plants has to presented.

Table 1 shows the electrical, thermal and total energies (E_e), (E_{th}) and (E_{Tot}) produced by the PV/T solar plants. It's reports as well some key performance factors of the PV/T plants energy production. The electrical, thermal and total demand coverage factors F_{el}, F_{th}, and F_{tot} are determined as a function of the different energies demands. In this work, the electrical and thermal loads are adopted for a standard building with a floor area of 100 m^2 and net volume of 300 m^3. The daily profile of DHW demand with $T_{setpoint}$ and T_{sup} are fixed at 40 and 15 °C, according to the standard EN 15316:2007 as shown in Fig. 8 [18]. While, the daily profile of electricity demand referred to this house (lighting and appliances) is illustrated in Fig. 9 [18]. As regards the energy needs, the average electrical energy needs of an Italian house are estimated as about 3000 kWh/y [19–21] (cooling and heating needs are included). Thus, the highest electrical coverage factor (F_{el}) is referred to the PV/T Roll-bond solar plant (76.31%). Otherwise, the lowest one is achieved by PV/T Sheet-and-tubes one (73.03%). The highest thermal coverage factors (F_{th}) is achieved with the PV/T Roll-bond type (76.89%).

As results, it is worth noting that the PV/T Roll-bond solar plant provides a better energy production and it's near to fully cover this house energy needs with a total coverage factor (F_{tot}) of 76.79%.

Table 1 Comparison of the energy performance of the PV/T solar plants

	E_{el} (kWh)	E_{th} (kWh)	E_{tot} (kWh)	$E_{DHW,PVT}$ (kWh)	F_{el} (%)	F_{th} (%)	F_{tot} (%)
PV/T sheet-and-tubes	6.473	6.152	12.626	0.25198	73.03	72.39	72.4
PV/T roll-bond	6.764	8.137	14.903	0.26765	76.31	76.89	76.7
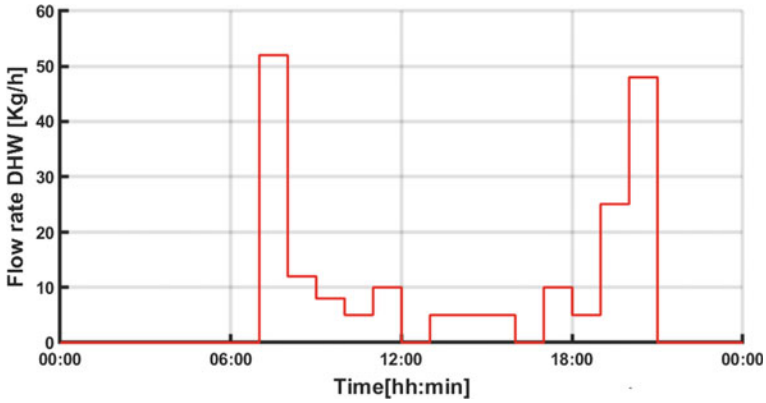
Fig. 8 Daily profile of the house DHW [18]
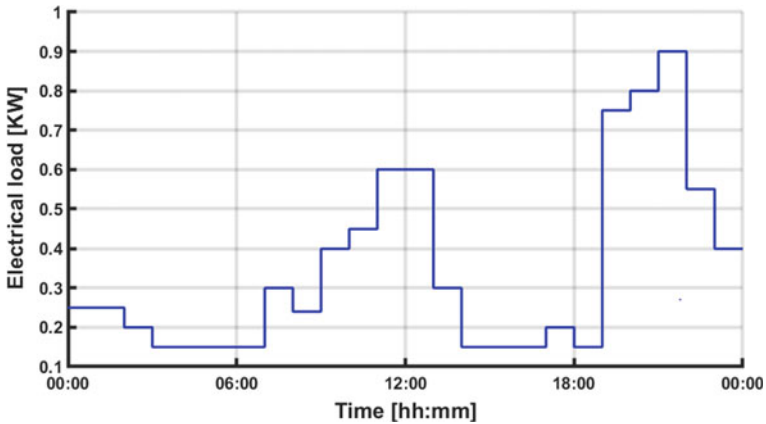
Fig. 9 Daily profile of the house electricity demand [18]

4 Conclusion

In this work, two different PV/T solar plants with Roll-bond and Sheet-and-tubes absorber designs have been numerically characterized and compared under the same

conditions. For each PV/T configuration, a numerical model has been established using the energy balance equations in dynamic regime in MATLAB. From the performances comparison of the PV/T solar plants, the PV/T Roll-bond design presents a better result than the PV/T Sheet-and-tubes with a total energy production of about 14.903 kWh per day. Thus, installing the PV/T Roll-bond solar plant in the building roof to supply the energy needs could be very promising. It is proved that this type of PV/T solar plant is near to fully cover the building energy needs, with an estimated total coverage factor (F_{tot}) of 76.79%. This result is very interesting in comparison with PV panels and solar collectors installed separately.

References

1. Michael JJ, S I, Goic R (2015) Flat plate solar photovoltaic–thermal (PV/T) systems: a reference guide. Renew Sustain Energy
2. Herrando M, Markides CN, Hellgardt K (2014) A UK-based assessment of hybrid PV and solar-thermal systems for domestic heating and power: system performance. Appl Energy 122:288–309
3. Sakellariou E, Axaopoulos P (2018) An experimentally validated, transient model for sheet and tube PVT collector. Sol Energy 174:709–718
4. Simonetti R, Molinaroli L, Manzolini G (2018) Development and validation of a comprehensive dynamic mathematical model for hybrid PV/T solar collectors. Appl Therm Eng 133:543–554
5. Guarracino I, Mellor A, Ekins-Dukes NJ, Markides CN (2016) Dynamic coupled thermal-and-electrical modelling of sheet-and-tube hybrid photovoltaic/thermal (PVT) collectors. Appl Therm Eng 101:778–795
6. Dupeyrat P, Menezo C, Rommel M, Henning H (2011) Efficient single glazed flat plate photovoltaic–thermal hybrid collector for domestic hot water system. Sol Energy 85:1457–1468
7. Calise F, Figaj RD, Vanoli L (2017) Experimental and numerical analyses of a flat plate photovoltaic/thermal solar collector. Energies
8. Aste N, Del Pero C, Leonforte F, Manfren M (2016) Performance monitoring and modeling of an uncovered photovoltaic-thermal (PVT) water collector. Sol Energy 135:551–568
9. Bombarda P, Di Marcoberardino G, Lucchini A, Leva S, Manzolini G, Molinaroli L, Pedranzini F, Simonetti R (2016) Thermal and electric performances of roll-bond flat plate applied to conventional PV modules for heat recovery. Appl Therm Eng 105:304–313
10. Del Col D, Padovan A, Bortolato M, Dai Prè M, Zambolin E (2013) Thermal performance of flat plate solar collectors with sheet-and-tube and roll-bond absorbers. Energy 58:258e269
11. El Fouas C, Hajji B, Gagliano A, Tina GM, Aneli S (2020) Numerical model and experimental validation of the electrical and thermal performances of photovoltaic/thermal plant. Energy Convers Manag 220:112939H
12. Calise F, Dentice d'Accadia M, Vanoli L (2012) Design and dynamic simulation of a novel solar trigeneration system based on hybrid photovoltaic/thermal collectors (PVT). Energy Convers Manag 60:214–225
13. Duffie JA, Beckman WA (1991) Solar engineering of thermal processes, 2nd edn. Wiley, New York
14. Khelifa A, Touafek K, Ben Moussa H (2014) Approach for the modelling of hybrid photovoltaic–thermal solar collector. IET Renew Power Gener 9(3):207–217
15. El Fouas C, El Manssouri O, Hajji B, Bouselham L, Tina GM, Gagliano A (2019) Modeling of dust deposition impact on PV/T hybrid collector performances. In: The 7th international renewable and sustainable energy conference, IRSEC 2019. https://doi.org/10.1109/IRSEC48032.2019.9078255

16. El Fouas C, Hajji M, Bouselham L, Hajji B, El Mehdi A, Bouali H. Analysis and design of an energy system based on PV/T water collector for building application. In: 2017 international renewable and sustainable energy conference (IRSEC), 2017, pp 1–5. https://doi.org/10.1109/IRSEC.2017.8477361
17. Fouas CE, Hajji M, Bouali H, Hajji B, Tina GM, Khlifi Y (2018) Absorber designs effect on the performance of PV/T water hybrid collector. In: The 1st international conference on electronic engineering and renewable energy, ICEERE. Springer. https://doi.org/10.1007/978-981-13-1405-6_78
18. Gagliano A, Tina GM, Aneli S, Chemisana D (2021) Analysis of the performances of a building-integrated PV/Thermal system. J Clean Energy Prod 320:128876
19. Gagliano A, Patania F, Nocera F, Capizzi A, Galesi A (2013) GIS-based decision support for solar photovoltaic planning in urban environment. Sustain Energy Build
20. Cicero F, Di Gaetano N, Speziale L (2010) I comportamenti di consumo elettrico delle famiglie italiane. L'Energia Elettrica 57
21. Ruggieri G. Alcune note sui consumi elettrici nel settore domestico in Italia. http://www.aspoitalia.it/attachments/220_Gianluca%20Ruggieri%20%20Consumi%20elettrici%20nel%20domestico.pdf

Simulation Numerical of a Bi-fluid Photovoltaic/Thermal Solar Panel

Oussama El Manssouri , **João Silva** , **Hajji Bekkay, José Teixeira** ,
Senhorinha Teixeira , **and Mohamed Hajji**

Abstract A photovoltaic/thermal solar collector (PV/T) is used to generate simultaneously thermal and electrical energy from received solar irradiation. In particular, bi-fluid PV/T with air and water cooling fluid is adopted for the energy supply of different applications. In this paper, a numerical study of a bi-fluid photovoltaic/thermal solar panel is performed to improve the cooling effect of two different heat exchangers: air rectangular channel with cooling fins and serpentine copper tubes. A simulation is conducted using ANSYS Fluent 21 software using the 3D steady-state analysis. Furthermore, some main key parameters, e.g., the velocity and temperature in each heat exchanger, are evaluated and analyzed. At a number of tubes and fins, respectively, of 12 and 20, the results show a good PV/T cooling effect. An improved fluid velocity and temperature for the two heat exchangers are determined to be 0.09 m/s and 46.67 °C for air and 0.27 m/s and 29.54 °C for the water.

Keywords CFD · Numerical model · Photovoltaic thermal system

O. El Manssouri (✉) · H. Bekkay · M. Hajji
Energy, Embedded System and Data Processing Laboratory, National School of Applied Sciences,
Mohamed First University, Oujda, Morocco
e-mail: elmanssouri.fso@gmail.com

J. Silva · J. Teixeira · S. Teixeira
University of Minho, Guimarães, Portugal
e-mail: js@dem.uminho.pt

J. Teixeira
e-mail: jt@dem.uminho.pt

S. Teixeira
e-mail: st@dps.uminho.pt

1 Introduction

The main advantage of a PV/T hybrid solar system is the conversion of solar radiation into electricity and heat at the same time. It could use water and/or air as cooling fluid [1, 2]. Zhou et al. [3] analyzed a three-dimensional model of a PV/T water collector with serpentine tubes which is established using the finite element method in software ANSYS. They studied the effect of many important parameters, namely the inlet velocity and spacing of the tubes, the arrangement of the tube, and the absorbing materials. Karaaslan and Menlik [4] examined the performance of the three-dimensional geometry sheet and tube PV/T-based water compared to nanofluids using software Fluent with different input velocities. The results concluded that the performance of the nanofluid gives better efficiency than water, and inlet fluid velocity plays a very important role in the thermal efficiency. Hosseinzadeh et al. [5] presented the effects of ZnO/water nanofluid and pure water on the electrical and thermal efficiencies of a photovoltaic thermal system. The 3D numerical model equations are discretized and solved using the pressure-based finite volume method by ANSYS Fluent 16.2 software. According to the results of their study, the inlet temperature of the fluid is the most effective parameter on the efficiency of the nanofluid-based PVT system. Then, Tembhare et al. [6] developed a 3D thermo-optical model thermal in ANSYS to evaluate the performance of a glazed PV/T. The effects of many important parameters, the greenhouse effect and the number of tubes, the bonding width, and the diameter of the tubes were studied. The application of the greenhouse effect increases the thermal efficiency by 12% compared to the case without this effect. Özakin and Kaya [7] studied theoretically and experimentally the air PV/T hybrid system with fins. They present the results of their work on the effect on the exergy, using a software model ANSYS for modeled fluid flow and heat transfer to show the temperature and the velocity of air. Lu et al. [8] presented and compared three types of PV/T collectors with sheet and tube-based water and air gap with different PV cell locations. Their results are based on computational fluid dynamics (CFD), and their analyses allow to choose the best performance of the three collectors.

The main objective of the present work is to analyze and compare the thermal performance of the two fluids air and water in the two different heat exchangers. The numerical modeling is performed by using the ANSYS Fluent 21 software.

2 CFD Model

First, the two geometries are presented and the system of equations governs the mixed convection flow and heat transfer in both heat exchangers. Then, the solution to this three-dimensional problem is implemented with Fluent software.

The studied system contains a wall installed on the upper wall of the air channel with fins and the serpentine copper tube for water, while the other walls are maintained

adiabatic. Figure 1a and b shows the geometry of the considered problem. The mixed convection problem in this hybrid collector, air and water system, was numerically investigated in three dimensions with a length of 1.138 m and a width of 0.52 m.

The physical systems considered are the air rectangular channel with fins for cooling the PV module as shown in Fig. 1a and the serpentine where water is flowing under the absorber as shown in Fig. 1b. The dimensions used in the numerical simulation are indicated in Table 1.
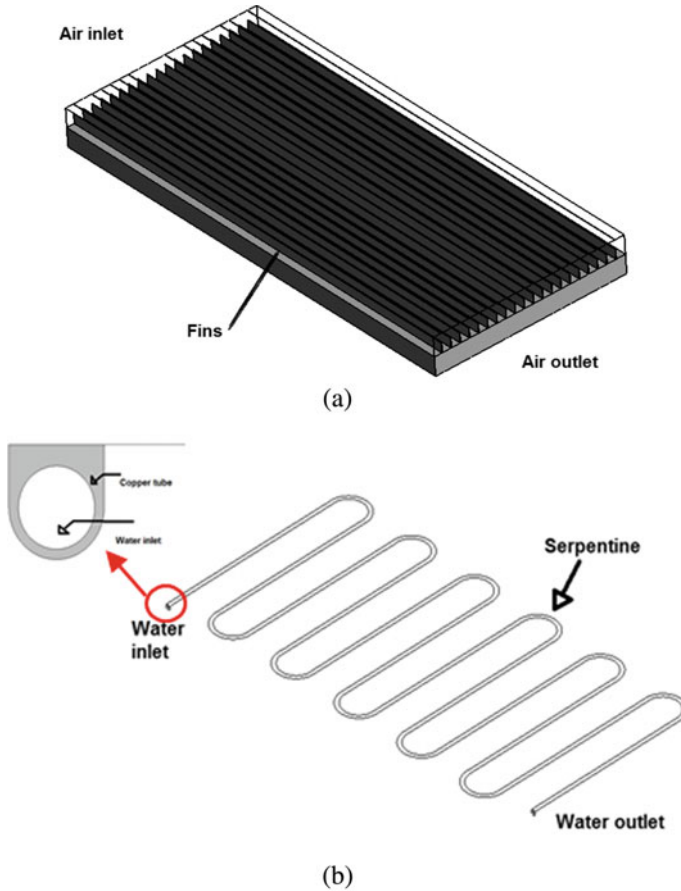
Fig. 1 Geometrical model defined for each case study: **a** the air rectangular channel with fins cooling and **b** design of the cooling tubes copper tubes serpentine

Table 1 Dimensions of two models

Parameter	Value	Parameter	Value
Number of tubes	12	Number of fins	20
Inner diameter of tube	0.008 m	Height of fins	0.01 m
Outer diameter of tube	0.01 m	Space between fins	0.01 m
Length of each tube	0.425 m	Height of air channel	0.05 m

2.1 Mathematical Model

Regarding the mathematical model, two main fields are solved, namely the fluid flow and the heat transfer. The former is simulated through the Navier–Stokes equation with the addition of a turbulence model. For the turbulence, the laminar model was chosen, being appropriated to address the flow within the air channel and in the serpentine due to their low Reynolds number. To account the heat transfer in the domain, the energy model was activated and the heat flux radiation through the wall was defined. The description of these models is presented in detail in the ANSYS Fluent User's guide [9]. Furthermore, a set of assumptions are adopted in this study, in order to simplify the mathematical modeling of the problem. These assumptions are based on the physical properties of the air flow and water flow with heat transfer in the regime of mixed convection.

- The flow is three-dimensional;
- The flow is laminar;
- Solar irradiation fixed at 1000 W/m^2.
- The thermo-physical properties of the different geometries were considered constant in both models.

Regarding the calculations developed by the software, the simulation was carried out using the steady-state formulation, and the convergence criterion was chosen as 10^{-5} and 10^{-7} for continuity and energy and momentum equations, respectively. The Semi-Implicit Method for Pressure-Linked Equations algorithm was used to solve the pressure–velocity coupling since it provides more efficient and robust single-phase flows. The contribution of the gravity acceleration was implemented in a downward direction.

2.2 Computational Domain, Boundary Conditions, and Mesh

The mathematical model must be solved under certain well-defined boundary conditions, which correspond to the characteristics of the mixed convection problem in

Table 2 Boundary conditions

Boundary condition	Value	Boundary condition	Value
Inlet air	306 K	Inlet water	300 K
Inlet velocity air	0.05 m/s	Wall adiabatic	300 K
Mass flow rate of water	0.035 kg/s	Inlet velocity water	0.14 m/s
Mass flow rate of air	0.007 k/s		
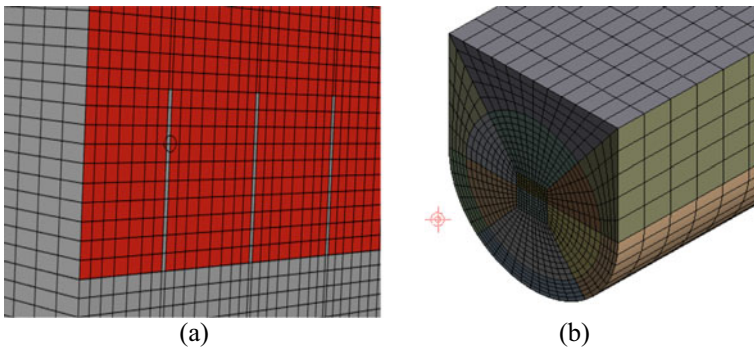
Fig. 2 Meshing the air channel with fins (**a**) and mesh in serpentine (**b**)

the two geometries. The types of conditions imposed in this study are given in Table 2 for both cases.

A structured three-dimensional mesh with hexahedron elements was carried out. Figure 2a and b shows the meshing details of the two exchangers.

3 Numerical Results and Discussion

For this purpose, the base case conditions mentioned in assumptions are considered and the intensity of solar irradiation used is 1000 w/m². For an inlet temperature of air is 33.70 °C, it can be observed that the temperature of the air fluid in the outlet reached at 46.67 °C is shown in Fig. 3a and it presents a good evolution of the temperature. This excellent cooling performed in the water cooling mode which allows for better heat extraction.

The velocity evolution outlet of air channel with fins is illustrated in Fig. 3b. From the mass flow rate and the surface flow rate and the density of the air, we calculated and we obtained the inlet velocity at 0.04 m/s.
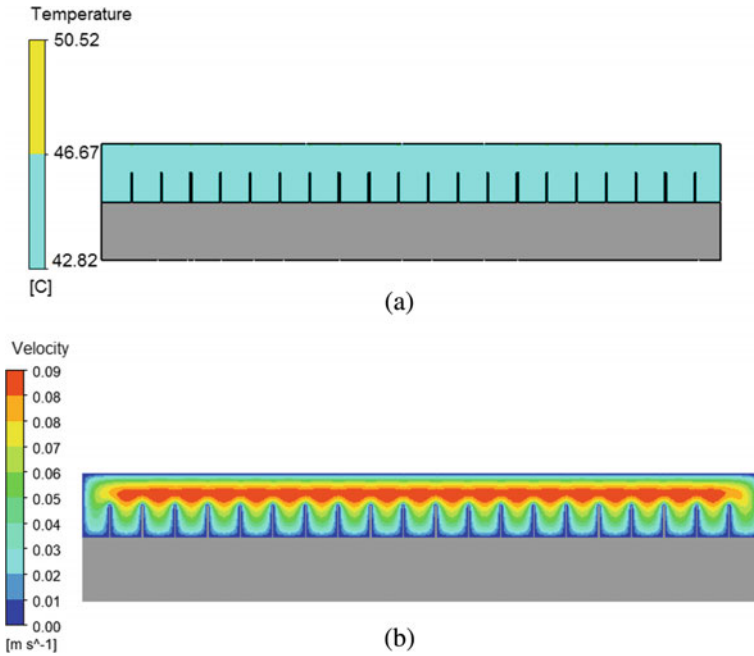
Fig. 3 Contours of **a** temperature and **b** velocity at the outlet of the serpentine

For the outlet, it can be observed that velocity arrived at center of channel at 0.09 m/s and between fins and upper and down at 0.06 m/s. Then, it is observed that the air velocity is very close to zero in the regions near the fins. Because the viscous forces and the friction effect cause the no-slip condition, the air velocity is zero at all points where it touched to the fins. In opposition to the areas near the fins, the velocity of air is faster and more laminar in the regions at the inlet and outlet and in the regions between the fins and upper of fins.

The temperature distributions in all the tubes from the first to the end are illustrated in Fig. 4. As can be seen from Fig. 5, the upper surface of tubes is uniformly heated with the inlet temperature is 26 °C, the water reached equal to 29.12 °C at 1000 W/m².

The velocity evolution outlet of water in serpentine is illustrated in Fig. 5b. From the mass flow rate and the surface flow rate and the density of the water, the inlet velocity of 0.14 m/s can be obtained. It is observed that the air velocity is very near to zero in the regions near the boundary of tube. For the outlet, velocity of water arrived at 0.27 m/s in the center of the tube.
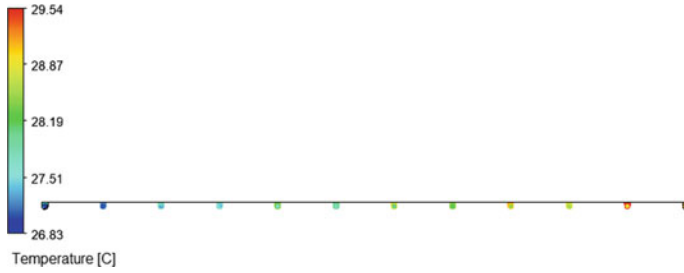
Fig. 4 Contours of temperature distributions of all the tubes
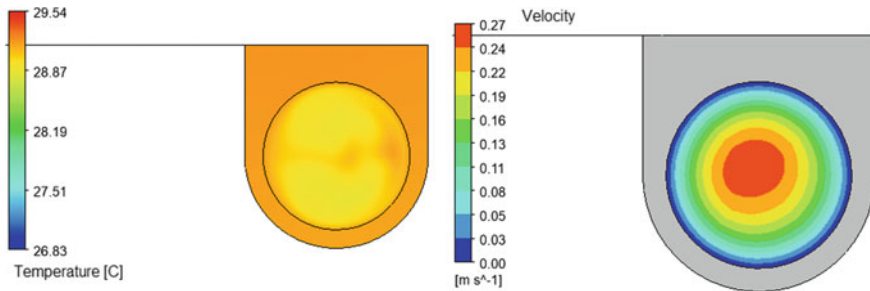
Fig. 5 Contours of **a** temperature and **b** velocity at the outlet of the serpentine

4 Conclusions

The present work concerns a numerical modeling of PV/T bi-fluid system with two different heat exchangers. The first one is with air channel and fins design, and the second is with a serpentine copper tube. A 3D steady-state numerical model is developed in ANSYS Fluent software in order to predict the thermal behavior of the cooling fluid. The simulation results show that using this heat exchange design allows an efficient cooling of the PV cells. At a fixed number of tube and fins about 12 and 20, respectively, an improved fluid velocity and temperature for the two heat exchangers are noted to be 0.09 m/s and 46.67 °C for air and 0.27 m/s and 29.54 °C for the water.

References

1. El Manssouri O, Hajji B, Tina GM, Gagliano A, Aneli, S (2021) Electrical and thermal performances of bi-fluid PV/thermal collectors. Energies 14. https://doi.org/10.3390/en1406 1633
2. El Manssouri O, El Fouas C, Hajji B, Rabhi A, Tina GM, Gagliano A (2020) Modeling and performances assessments of PV/T bifluid hybrid collector: three cooling modes operation case, pp 1–6. https://doi.org/10.1109/iceit48248.2020.9113233
3. Zhou J, Ke H, Deng X (2018) Experimental and CFD investigation on temperature distribution of a serpentine tube type photovoltaic/thermal collector. Sol Energy 174:735–742. https://doi.org/10.1016/j.solener.2018.09.063
4. Karaaslan I, Menlik T (2021) Numerical study of a photovoltaic thermal (PV/T) system using mono and hybrid nanofluid. Sol Energy 224:1260–1270. https://doi.org/10.1016/j.solener.2021.06.072
5. Hosseinzadeh M, Salari A, Sardarabadi M, Passandideh-Fard M (2018) Optimization and parametric analysis of a nanofluid based photovoltaic thermal system: 3D numerical model with experimental validation. Energy Convers Manag 160:93–108. https://doi.org/10.1016/j.enconman.2018.01.006
6. Tembhare SP, Barai DP, Bhanvase BA (2022) Performance evaluation of nanofluids in solar thermal and solar photovoltaic systems: a comprehensive review. Renew Sustain Energy Rev 153:111738. https://doi.org/10.1016/j.rser.2021.111738
7. Özakin AN, Kaya F (2019) Effect on the exergy of the PVT system of fins added to an air-cooled channel: a study on temperature and air velocity with ANSYS Fluent. Sol Energy 184:561–569. https://doi.org/10.1016/j.solener.2019.03.100
8. Lu L, Wang X, Wang S, Liu X (2017) Analysis of three different sheet-and-tube water-based flat-plate PVT collectors. J Energy Eng 143:04017022. https://doi.org/10.1061/(asce)ey.1943-7897.0000456
9. ANSYS (2013) ANSYS Fluent theory guide

Numerical and Parametric Analysis of Nanofluid-Based PV/T System for Hydrogen Production

Safae Margoum, Chaimae El Fouas, Mohamed Hajji, Hajji Bekkay, and Abdelhamid Rabhi

Abstract Hydrogen production from the PV/T hybrid system becomes an attractive option in recent years. As a good alternative to the fuel, hydrogen can be produced by electrolysis of water using the electrical output of the PV/T system. In this work, a Cu/water nanofluid PV/T system is designed for producing hydrogen through the electrolysis of water in a proton exchange membrane electrolysis cell (PEMEC). Assessment of electrical power required by the PEMEC for separation of the oxygen and hydrogen molecules is performed, based on a dynamic numerical model of PEMEC-PV/T developed in MATLAB software. To further enhance the amount of hydrogen produced, a parametric analysis had been proceeded to investigate the impact of the Cu/water nanofluid mass flow rate and inlet temperature. In addition, the effect of glazing the PV/T panel on the hydrogen generation from the whole system is studied. From the analysis of the obtained results, the PEMEC-PV/T is seemed very promising for generating hydrogen. An important amount of hydrogen produced is found to be 1.556 mol/h in the case of an unglazed PV/T system, where effective energy provided by the Cu/water nanofluid PV/T system is ensured adopting an optimum mass flow rate of 216 kg/h and inlet temperature of 303.15 K.

S. Margoum (✉) · C. El Fouas · M. Hajji · H. Bekkay
Renewable Energy, Embedded System and Information Processing Laboratory, National School of Applied Sciences, Oujda, Mohammed First University, 60000 Oujda, Morocco
e-mail: safae.margoum@ump.ac.ma

C. El Fouas
e-mail: c.elfouas@ump.ac.ma

M. Hajji
e-mail: hajjiyam@gmail.com

H. Bekkay
e-mail: hajji.bekkay@gmail.com

A. Rabhi
Modelization, Information and System Laboratory, Picardie Jules Verne University, Amiens, France
e-mail: abdelhamid.rabhi@u-picardie.fr

1 Introduction

Hydrogen can be generated by electrolysis of water by adopting the solar electrical production from PV panels, without using fossil sources and emission of carbon dioxide. However, because of temperature issue presented by the PV panel, coupling the PV/T system with electrolysis device for producing hydrogen could be an effective technique. Among various types of electrolysis device, the alkaline and polymer exchange membrane (PEMC) electrolysers are the most commercially available [1]. In particular, the PEMEC electrolysis system is the most used to produce pure hydrogen [2]. The work of [3] aims to assess the monthly performance of an integrated PV/T for cooling and hydrogen production from PEMC system. As a results, the amount of hydrogen produced is found to be maximum about 9.7 kg in August. The effectiveness of PV/T system for providing the required electrical power of the PEMEC and preheat the feed water is proved trough the work of [4]. In other study, a numerical model to investigate the hydrogen generating from proton exchange membrane (PEM) electrolyzer powered by a photovoltaic thermal system (PV/T) is developed by [5]. It is observed that 0.018 kg of hydrogen can be produced by using the electrical energy of the PVT system. The performance of the PEMEC, powered by the (PVT) system, is also examined by [6]. Accordingly, it is found that the PVT-TEG-PEMEC system outperforms other systems in hydrogen production. A novel solar hydrogen generation PV/T system is proposed, and a numerical model for evaluating and optimizing the PV/T thermodynamic performance is established in [7]. They investigated the effect of various parameters, including solar radiation, the fluid, and PV emissivity on the hydrogen production. The possibility of hydrogen production from PV/T solar plant formed by two PV/T system in series is examined as well by [8], where a considerable amount of hydrogen produced is depicted about 60 ml/min. From previous literature works, many studies have been conducted on the application of conventional PV/T systems for the hydrogen generation. In recent years, a few research works have been suggested the nanofluids-based PV/T system for producing a pure hydrogen. The possibility of applying nanofluid-based PVT system for hydrogen generation is experimentally investigated by [9]. According to this research, using nanofluids improves the PVT system's electrical and thermal performance as well as its hydrogen generation. It is estimated that 51 kWh of energy is able to produce one kilogram of hydrogen. Another PV/T model using MWCNT and Fe_2O_3 nanofluids as passive cooling agents is built for improving its electrical production [10]. This PV/T system is designed to generate hydrogen from water electrolysis. The results show that Fe_2O_3 nanofluid was the promising source to enhance the working ability of the PV/T system, a peak production of hydrogen of 17.3 mL/min was observed for the 0.01 kg/s of Fe_2O_3.

This paper is focused on the coupling of proton exchange membrane electrolysis cell with a Cu/water nanofluids-based PV/T system for producing hydrogen. Furthermore, the evaluation of the impact of some main PV/T parameters on its electrical output and the hydrogen production is proceed.

2 PEMEC-PV/T System Description

2.1 PEMEC-PVT System Modeling

In this work, hydrogen is produced by the electrolysis of water in a proton exchange membrane electrolysis cell (PEMEC) using the electrical output of the PV/T panel. In general, a PEMEC consists of an anode and a cathode electrode, and a thin membrane as described in Fig. 1. By applying a voltage in the PEMEC device, water molecules are dissociated into hydrogen and oxygen gas. On the anode side, an atom of oxygen, two hydrogen protons, and two electrons are born. While hydrogen gas is generated when hydrogen protons and electrons are transferred through the membrane to the cathode side. As follows, the main reaction describing the steps of the water electrolysis process in the PEMEC device is presented [6]:

$$\text{Anode reaction } H_2O \rightarrow 2H^+ + 1/2O_2 + 2e^- \tag{1}$$

$$\text{Cathode reaction } 2H^+ + 2e \rightarrow H_2 \tag{2}$$

$$\text{Complete reaction } H_2O \rightarrow H_2 + 1/2O_2 \tag{3}$$
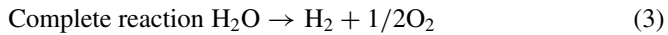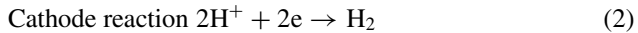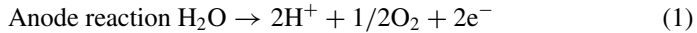
In order to assess the PV/T panel electrical power required for the energy supply of the PEMEC device, a dynamic thermo-electrical PV/T model is developed based on a heat balance equations for each PV/T layer, using previous literature works [11, 12]. This electrical power is a function of the PV cell temperature T_{cell}, reference electrical

Fig. 1 Schematic diagram of hydrogen production from the PEMEC-PV/T system
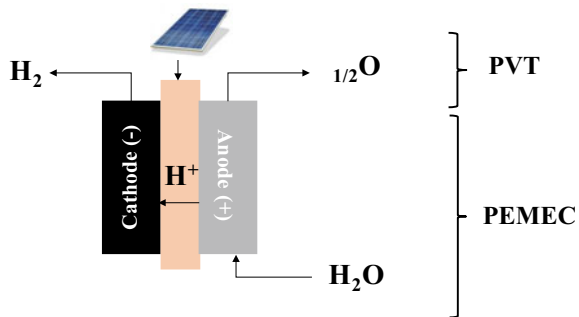
Table 1 Expressions of the PEMEC parameters

Parameters	Equations	
The amount of hydrogen production (mol/h) [12]	$\dot{m}_{\text{Hydrogen}} = \frac{E_e}{2\vartheta F}$	(5)
PEME voltage electrolysis (V) [12]	$\vartheta = \vartheta_{\text{rev}} + \vartheta_{\text{anode}} + \vartheta_{\text{cathode}} + \vartheta_{\text{ohm}}$	(6)
The reversible cell voltage (V) [6]	$\vartheta_{\text{rev}} = 1.229 - 8.5e^{-4}(T_{\text{cell}} - 298.15)$	(7)
The anode and anode polarization voltage (V) [6]	$\vartheta_{\text{anode}} = \frac{RT_{\text{cell}}}{\alpha_{\text{anode}}} \ln\left(\frac{j}{2j_{\text{anode}}} + \sqrt{1 + \left(\frac{j}{2j_{\text{anode}}}\right)^2} \right)$	(8)
The anode and cathode polarization voltage (V) [6]	$\vartheta_{\text{anode}} = \frac{RT_{\text{cell}}}{\alpha_{\text{cathode}}} \ln\left(\frac{j}{2j_{\text{cathode}}} + \sqrt{1 + \left(\frac{j}{2j_{\text{cathode}}}\right)^2} \right)$	(9)
The ohmic voltage (V) [6]	$\vartheta_{\text{ohm}} = \frac{j\delta_{\text{membrane}}}{k}$	(10)
The membrane conductivity (Ω/m) [6, 14]	$k_{\text{membrane}} =$ $(0.5139\lambda - 0.00326)\exp\left(1268\left(\frac{1}{303} - \frac{1}{T_{\text{cell}}}\right)\right)$	(11)

efficiency η_{ref} and a solar irradiance G [13]. For calculating the hydrogen produced in the PEMEC unit, Table 1. Expressions of the PEMEC parameters resumes all the governing equations and expressions of PEMEC parameters included in the simulation.

$$E_e = \eta_{\text{ref}}[1 - \beta(T_{\text{cell}} - 25)]GPA\tau_{\text{glass}} \tag{4}$$

η_{ref} reference PV cell efficiency, β reference temperature coefficient, P packing factor, A collector surface, τ_{glass} glass transmittivity, their values are 12%, 0.0045 K^{-1}, 1, 1.15 m^2, 0.95 respectively.

3 Results and Discussion

3.1 PEMEC-Cu/Water Nanofluid PV/T Hydrogen Production

In this section, the hydrogen production from water electrolysis ensured by PEMEC-PVT system-based Cu/water nanofluid is evaluated, where the electrical output of the PV/T panel is used for the energy supply of the electrolysis system; from Fig. 2, it is obvious that the hydrogen produced is directly proportional to the electrical power.

During the day, the electrical production presents the same trend with the solar irradiance. In Fig. 3, the solar irradiance and ambient temperature reach a maximum value at 13 h of 1000 W/m^2 and 309 °K which affects the electrolysis process. As
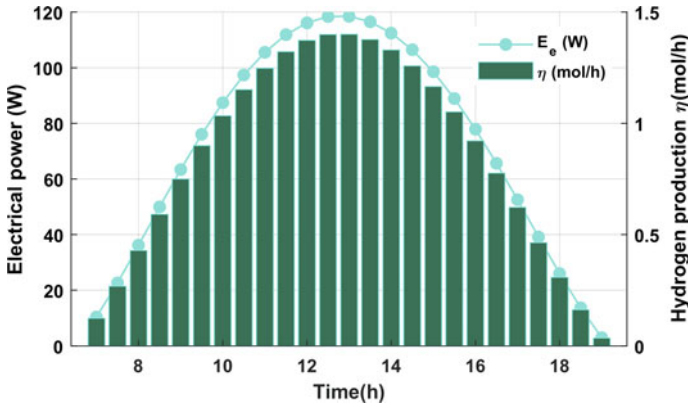
Fig. 2 Cu/water-based PVT/T hydrogen production
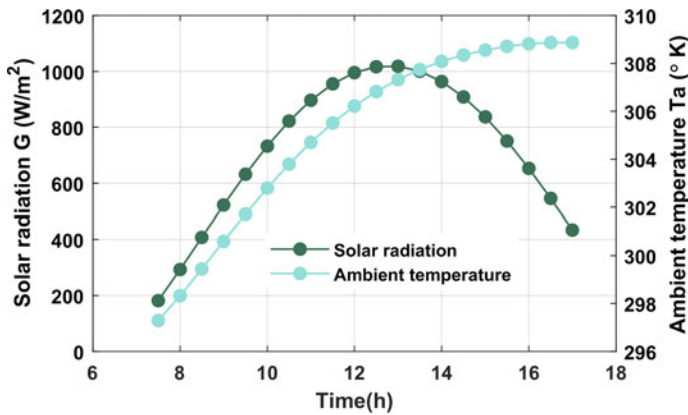
Fig. 3 Environmental data: solar irradiance and ambient temperature

observed, a maximum value of electrical power and hydrogen production is reached at 13h00 about 118.5 W and 1.401 mol/h respectively.

3.2 Effect of Cu/Water Nanofluid Mass Flow Rate and Inlet Temperature on the PEMEC-PV/T Hydrogen Production

In order to improve the amount of hydrogen generated by the PEMEC-PV/T system, the effect of some PV/T main key parameters, e.g., the Cu/water nanofluid mass flow rate and inlet temperature on the effective cooling of the PV/T panel, is investigated in this section.
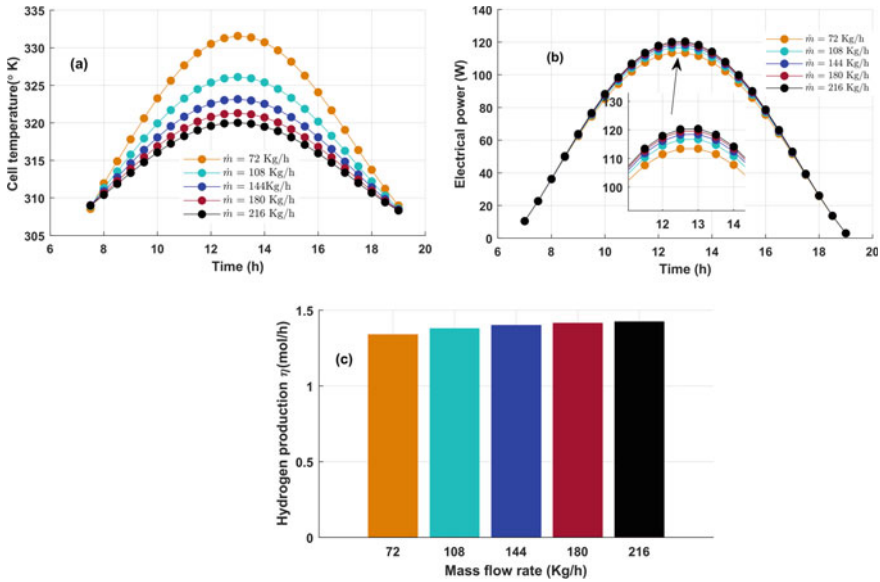
Fig. 4 Evolution of cell temperature (**a**), electrical power (**b**) and hydrogen production (**c**) with the Cu/water nanofluid mass flow rate

3.2.1 Effect Cu/Water Nanofluid Mass Flow Rate

The impact of Cu/nanofluid mass flow rate on the hydrogen produced is determined considering five mass flow rate values of 72, 108, 144, 180, and 216 kg/h. Figure 4 shows the evolution of the cell temperature (a), electrical power (b), the hydrogen production (c) with the Cu/water nanofluid mass flow rate increases.

It is observed from Fig. 4a that increasing the mass flow rate from 72 to 216 kg/h reduces the cell temperature from a maximum value of 331.6 to 320 °K. As a result, the maximum electrical power is increased from 113.4 to 120.4 W (Fig. 4b). This is related to the effective extraction of the heat at the PV cell back. Thus, increasing the mass flow rate enhances the operating voltage of the PEMEC and improve the hydrogen production by 6.27% as illustrated in Fig. 4c.

3.2.2 Effect of Cu/Water Nanofluid Inlet Temperature

The effects of increasing the Cu/water nanofluid inlet temperature on the cell temperature, electrical power, and hydrogen production of the PEMEC-PVT system are presented in Fig. 5a, b, c, respectively. The cell temperature and the electrical power have been significantly affected by an increase in the inlet temperature of Cu/water. As illustrated in Fig. 5a and b by rising the inlet temperature from 303.15 to 318.15 °K the PV cell temperature is increased by 1.34%. While a reduction in electrical power of about 6.68% and hydrogen production of about 7.26% are noted Fig. 5c. This is
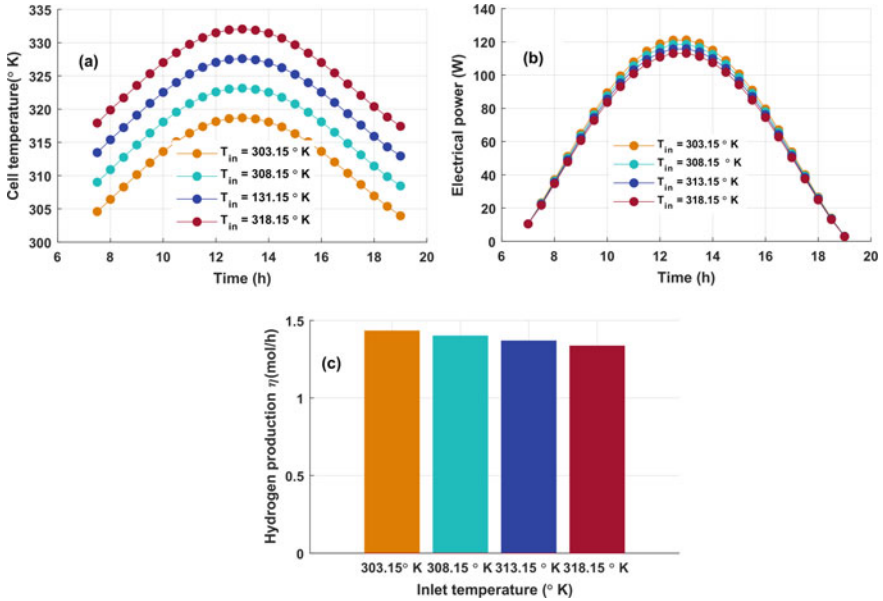
Fig. 5 Evolution of cell temperature (**a**), electrical power (**b**) and hydrogen production (**c**) with the Cu/water nanofluid inlet temperature

due to the lower amount of heat extracted from the PV cells, using a higher inlet temperature Cu/water nanofluid inlet temperature. Then, important production of hydrogen by the PEMEC-PV/T system can be performed while ensuring an effective cooling of the PV cells.

3.3 Effect of Glazing the PV/T Panel on the PEMEC-PV/T Hydrogen Production

The glazing and unglazing of the PV/T panel could affect the effective cooling as well as the energy production of the panel. The variation of the PV cells temperature, electrical power, and hydrogen production of PEMEC-PVT system for glazed and unglazed cases is shown in Fig. 6a, b, c, respectively. According to these figures, the glazed PV/T presents a higher PV temperature of 323.2 °K at 13h00 compared with the unglazed one (312.5 °K). This can be explained by the fact that adding the glass cover contribute to more heating of the PV cells, which affects its electrical production. Consequently, electrical power is enhanced by 16 W while removing the glass cover as depicted in Fig. 6b. In terms of the hydrogen produced by the PEMEC-PV/T system, maximum production is obtained for the PV/T unglazed 1.556 mol/h as shown in Fig. 6c (1.401 mol/h for the glazed PV/T).
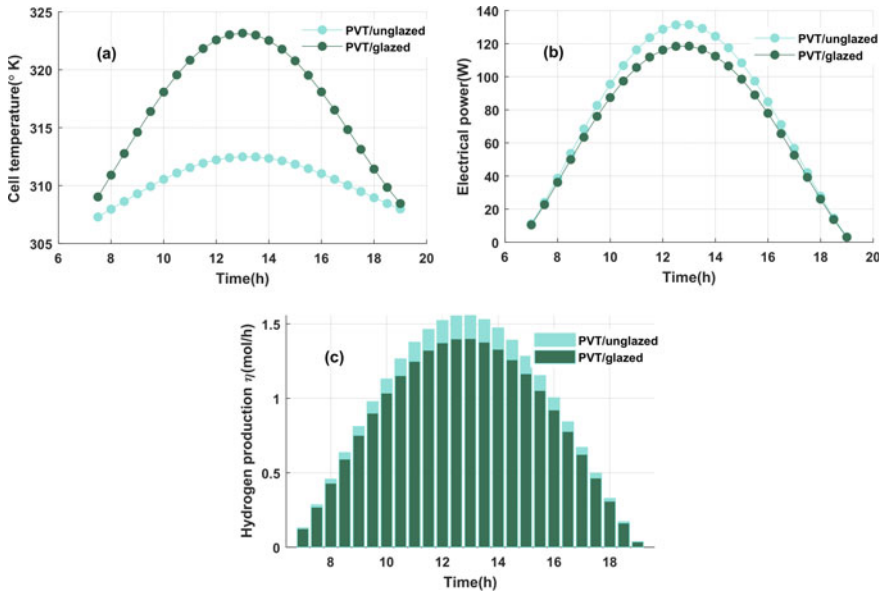
Fig. 6 Variation of PV cell temperature (**a**), electrical power (**b**), hydrogen production (**c**) for glazed and unglazed Cu/water nanofluid PV/T

4   Conclusion

In this study, a parametric analysis of Cu/water nanofluid PV/T system coupled with a proton exchange membrane electrolysis cell (PEMEC) for hydrogen production is performed. The objective is to evaluate the effect of some main key parameters of the Cu/water nanofluid PV/T system, e.g., nanofluid mass flow rate and inlet temperature on the PV/T energy production and hydrogen produced. Next, the impact of glazing the PV/T panel on the PEMEC-PV/T production is investigated. The obtained results driven from the developed numerical PEMEC-PV/T model allow dressing the following conclusions:

- Increasing the Cu/water nanofluid mass flow rate enhances the PV/T system electrical output and hydrogen generated by PEMEC. An optimum value of the mass flow rate is fixed to be 216 kg/h.
- While the Cu/water nanofluid inlet temperature is lower with a given value of about 303.15 K, important production of hydrogen is allowed.
- Unglazing of the PV/T panel is recommended to further improve the production of the PEMEC-PV/T system. In this case, a significant amount of hydrogen produced is found around 1.556 mol/h.

References

1. Toghyani S, Afshari E, Baniasadi E, Atyabi SA, Naterer GF (2018) Thermal and electrochemical performance assessment of a high temperature PEM electrolyzer. Energy 152:237–246. https://doi.org/10.1016/j.energy.2018.03.140
2. Sellami MH, Loudiyi K (2017) Electrolytes behavior during hydrogen production by solar energy. Renew Sustain Energy Rev 70:1331–1335. https://doi.org/10.1016/j.rser.2016.12.034
3. Dincer I, Acar C (2015) Review and evaluation of hydrogen production methods for better sustainability. Int J Hydrogen Energy 40(34):11094–11111. https://doi.org/10.1016/j.ijhydene.2014.12.035
4. Görgün H (2006) Dynamic modelling of a proton exchange membrane (PEM) electrolyzer. Int J Hydrogen Energy 31(1):29–38. https://doi.org/10.1016/j.ijhydene.2005.04.001
5. Cilogulları M, Erden M, Karakilcik M, Dincer I (2017) Investigation of hydrogen production performance of a Photovoltaic and Thermal System. Int J Hydrogen Energy 42(4):2547–2552. https://doi.org/10.1016/j.ijhydene.2016.10.118
6. Hydrogen production performance of a photovoltaic thermal system coupled with a proton exchange membrane electrolysis cell—ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0360319921044888?via%3Dihub. Accessed 04 Mar 2022
7. Wang H, Li W, Liu T, Liu X, Hu X (2019) Thermodynamic analysis and optimization of photovoltaic/thermal hybrid hydrogen generation system based on complementary combination of photovoltaic cells and proton exchange membrane electrolyzer. Energy Convers Manag 183:97–108. https://doi.org/10.1016/j.enconman.2018.12.106
8. Study and modeling of energy performance of PV/T solar plant for hydrogen production. Springer. https://doi.org/10.1007/978-981-15-6259-4_91. Accessed 05 Mar 2022
9. Sangeetha M (2020) Progress of MWCNT, Al_2O_3, and CuO with water in enhancing the photovoltaic thermal system. Int J Energy Res. https://doi.org/10.1002/er.4905. Accessed 15 June 2021
10. Anderson A et al (2022) Effects of nanofluids on the photovoltaic thermal system for hydrogen production via electrolysis process. Int J Hydrogen Energy. https://doi.org/10.1016/j.ijhydene.2021.12.218
11. Salari A, Kazemian A, Ma T, Hakkaki-Fard A, Peng J (2020) Nanofluid based photovoltaic thermal systems integrated with phase change materials: numerical simulation and thermodynamic analysis. Energy Convers Manag 205:112384. https://doi.org/10.1016/j.enconman.2019.112384
12. Han B, Steen SM, Mo J, Zhang F-Y (2015) Electrochemical performance modeling of a proton exchange membrane electrolyzer cell for hydrogen energy. Int J Hydrogen Energy 40(22):7006–7016. https://doi.org/10.1016/j.ijhydene.2015.03.164
13. Maadi SR, Kolahan A, Passandideh-Fard M, Sardarabadi M, Moloudi R (2017) Characterization of PVT systems equipped with nanofluids-based collector from entropy generation. Energy Convers Manag 150:515–531. https://doi.org/10.1016/j.enconman.2017.08.039
14. Olivier P, Bourasseau C, Bouamama PrB (2017) Low-temperature electrolysis system modelling: a review. Renew Sustain Energy Rev 78:280–300. https://doi.org/10.1016/j.rser.2017.03.099

A Hybrid HGWO-PSO Approach for Combined Economic Emission Dispatch Problem Optimization

Naima Agouzoul, Faissal Elmariami, Aziz Oukennou, Ali Tarraq, and Rabiaa Gadal

Abstract Currently, thermal power station pollution requires special attention. In fact, the majority of the energy demand is produced by thermal power industry. Therefore, the objective of combined economic emission dispatch (CEED) is to optimize the total cost of energy production, by minimizing the CO_2 emissions. The challenge of CEED is to find the trade-off between these two contradictory objectives simultaneously while considering the valve point effect. The said valve point that acts nonlinearly and influences the total energy price. In this paper, we are combining two algorithms to develop a new hybrid algorithm named "HGWO-PSO." This algorithm goal is to contribute toward the optimization of the CEED problem using the particle swarm optimization (PSO) and the gray wolf optimization (GWO) algorithms. The hybrid HGWO-PSO technique has been shown to be an effective technique. It finds the global optimal convergence solution with the fewest verified iterations using the standard IEEE-30 BUS network system. An observable improvement in system behavior has been seen while comparing HGWO-PSO with conventional PSO and GWO methods.

Keywords Cost · CEED · CO_2 emission · GWO · HGWO-PSO · Optimization · PSO

1 Introduction

The amount of attention paid to climate change caused by air pollution is growing day by day. This rise is due to an increase in electrical energy demand where thermal power units provide majority of this energy. For example, in China more than 70% of energy is delivered by thermal power station [1]. It is undeniable that renewable

N. Agouzoul (✉) · F. Elmariami · A. Tarraq · R. Gadal
Laboratory of Energy and Electrical Systems, National Higher School of Electricity and Mechanics (ENSEM), El Jadida Road, km 7, Casablanca, Morocco
e-mail: naima.agouzoul@ensem.ac.ma

A. Oukennou
SARS Team, National School of Applied Sciences (ENSA), CA University, Safi, Morocco

energy sources have contributed to a significant reduction of CO_2 emissions, nevertheless their full integration into the global power grid still under development and requires time. Hence, the aim of CEED is to operate the energy with a minimum cost and emissions, while satisfying all the system constraints. Several hybrid PSO-GWO techniques have recently been used to solve a number of problems, demonstrating their efficacy. In [2], hybrid PSO-GWO was applied to optimize microgrid scheduling problem. In [3], PSO-GWO is compared to PSO, GWO, ABC, and SSA as well as three different hybrid methods of PSO-GWO and tested on five benchmark functions and three different real problems. In [2, 3], GWO was only used in part with a few iterations and population in PSO execution. In [4], the pertinence of GWO-PSO is revealed when minimizing power loss and voltage deviation in ORPD problem. In [5], the success of PSO-GWO is verified by the significant reduction of congestion. Here GWO intervenes during the execution of PSO with small probabilities to replace some random generated particles with improved ones. In [6], the hybrid PSO-GWO approach has ensured a high reduction of total harmonic distortion of the source current. In [7], PSO was executed first, followed by GWO in each iteration. Although this technique is simple to implement, it has a long execution time. In [8], the dynamic finite automata learning method, based on a hybrid PSO-GWO, was able to efficiently harness sensor energy, extend the network lifetime and transmit data in an optimal path. Here the process was initialized by running GWO followed by PSO using modified equations. Until now, many approaches combining PSO and GWO algorithms have been proposed and applied to solve real-world optimization problems as mentioned before. Nevertheless, in our knowledge we have not found papers applying this hybrid approach to solve the CEED problem. The proposed technique is validated by comparing PSO to GWO algorithm in an IEEE-30-bus test system. The following is the paper's outline: Sect. 2 describes the CEED model. Section 3 gives a brief overview of PSO and GWO methods and explains the operation of the proposed HGWO-PSO algorithm. While Sect. 4 describes the experimental results with discussions, and the paper is concluded in Sect. 5.

2 CEED Problem Formulation

2.1 Objective Function

To reduce the energy production unit's fuel costs and emissions, the price penalty factor (PPF) was used to combine two objective functions into a single one.

The model used is described above using the weighted sum technique [9]:

$$\text{minOF} = w_{f1}C_{\text{tot}} + w_{f2}h_T E_{\text{tot}} \qquad (1)$$

where, minOF is the objective function, C_{tot} is the operating cost, E_{tot} represents the emissions function, w_{f1} and w_{f2} are weight' factors, h_T denotes the total PPF

in (\$/kg). According to [10], the min–max approach (PPF) was chosen because it provides relevant results. We use the equation below to calculate PPF of every generator as h_i.

$$h_i = C_i\left(P_i^{\min}\right)/E_i\left(P_i^{\max}\right) \tag{2}$$

where C_i and E_i are the operating cost and the emissions of power generating unit i, respectively, $P_i^{\min}$ and $P_i^{\max}$ are, respectively, the minimum and maximum output limit of the i power unit (MW).

The total PPF is given as the summation of PPF power units:

$$h_T = \sum_{i=1}^{N_g} h_i \tag{3}$$

N_g: the number of power units.

The operating cost and emission functions

The operating cost function is given as:

$$C_{\text{tot}} = \sum_{i=1}^{N_g} C_i = \sum_{i=1}^{N_g} a_i P_i^2 + b_i P_i + c_i + \left|e_i \sin\left(f_i\left(P_i^{\min} - P_i\right)\right)\right| \tag{4}$$

where a_i, b_i and c_i are the cost coefficients of power unit i. e_i, f_i are the cost coefficients of generator i representing the valve point effect, $P_i^{\min}$ and P_i are the minimum output power limit and the output power of the i unit, respectively (MW).

The thermal unit's emissions are expressed mathematically as shown below:

$$E_{\text{tot}} = \sum_{i=1}^{N_g} E_i = \sum_{i=1}^{N_g} \gamma_i + \beta_i P_i + \alpha_i P_i^2 \tag{5}$$

where E_i and α_i, β_i, γ_i are emissions and emission coefficients of unit i, respectively.

2.2 Equality and Inequality Constraints

To ensure the system power balance, the total power generated from thermal units must be equal to the summation of the total load demands and network losses.

$$\sum_{i=1}^{Ng} P_i = P_D + P_L \tag{6}$$

where P_D presents the total load demand, P_L is the network losses calculated using B matrix method [7].

$$P_L = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_i B_{ij} P_j + \sum_{i=1}^{N_g} P_i B_{i0} + B_{00} \tag{7}$$

where B_{ij}, B_{i0}, and B_{00} are coefficient losses.

The power generators must be bounded between its lower and upper values.

$$P_i^{\min} \leq P_i \leq P_i^{\max} \tag{8}$$

3 Proposed Methodology

3.1 Briefs on Basics of PSO and GWO Algorithms

In PSO, particles move through a space to search an optimum solution [11]. The movement of each particle is updated based on the best position on its previous history ($P_{i \cdot \text{Pbest}}^n$) and best global position of the group ($P_{i \cdot \text{Gbest}}^n$) using this function:

$$v_i^{n+1} = w \cdot v_i^n + c_1 \cdot r_1 \left(P_{i \cdot \text{Pbest}}^n - x_i^n \right) + c_2 \cdot r_2 \left(P_{i \cdot \text{Gbest}}^n - x_i^n \right) \tag{9}$$

where w is the inertia weight. The velocity and the position of the ith particle are v_i^n and x_i^n at the nth iteration. While c_1 and c_2 denote the coefficients acceleration, r_1, r_2 are random numbers within [0, 1].

The updating position of each particle during iteration i is expressed as:

$$x_i^{n+1} = x_i^n + v_i^{n+1} \tag{10}$$

The GWO [12] is a recent meta-heuristic technique inspired by gray wolves hunting prey behavior. Based on a hierarchy composed of four levels of wolves: α, β, δ, and γ.

The GWO operations are modeled mathematically as:

$$D = \left| C \times X_p(t) - X(t) \right| \tag{11}$$

$$X(t + 1) = X_p(t) - (A \times D) \tag{12}$$

t corresponds to the running iteration. X and X_p are the position of wolf and the prey, respectively. A and C are the coefficient vectors that are formulated as follows:

$$A = a \times (2 \times r_1 - 1) \tag{13}$$

$$C = 2 \times r_2 \tag{14}$$

r_1 and r_2 are random numbers in [0, 1]. In addition, during each iteration, "a" is decreasing linearly from 2 to 0.

The α, β and δ wolves update their locations according to the following equations:

$$\overrightarrow{D_\alpha} = \left| \overrightarrow{C_1}\overrightarrow{X_\alpha} - \vec{X} \right|; \quad \overrightarrow{D_\beta} = \left| \overrightarrow{C_2}\overrightarrow{X_\beta} - \vec{X} \right|; \quad \overrightarrow{D_\delta} = \left| \overrightarrow{C_3}\overrightarrow{X_\delta} - \vec{X} \right| \tag{15}$$

$$\overrightarrow{X_1} = \left| \overrightarrow{X_\alpha} - \overrightarrow{A_1}\overrightarrow{D_\alpha} \right|; \quad \overrightarrow{X_2} = \left| \overrightarrow{X_\beta} - \overrightarrow{A_2}\overrightarrow{D_\beta} \right|; \quad \overrightarrow{X_3} = \left| \overrightarrow{X_\delta} - \overrightarrow{A_3}\overrightarrow{D_\delta} \right| \tag{16}$$

Thus, the optimal solution is the average value of X_1, X_2, and X_3 calculated as follows:

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3} \tag{17}$$

3.2 HGWO-PSO Algorithm Proposed Mechanism

According to [13], PSO is known for its good exploitation. However, its exploration capacity remains limited [14]. Our proposed algorithm combines the advantages of both techniques. The novelty of our approach is reflected in the proposed structure of the PSO and GWO combination, where we have avoided the successive execution of both algorithms in all iterations. The arbitrary execution of PSO is conditioned by a randomly generated number defined between 0 and 1 and less than a number predefined in the control parameters list in Table 1. In this case, GWO is executed for 300 iterations while PSO is executed for 30 iterations instead of 300. The mechanism of this hybrid algorithm is shown in Fig. 1.

Table 1 Parameter setting of HGWO-PSO for solving CEED problem	Parameters	Values
	Maximum iterations of GWO	300
	Population size of GWO	50
	Maximum iterations of PSO	30
	Population size of PSO	20
	Inertia weight	$w = 0.9$
	Acceleration factors	$c_1 = 2$ and $c_2 = 2$
	The generated number	0.5
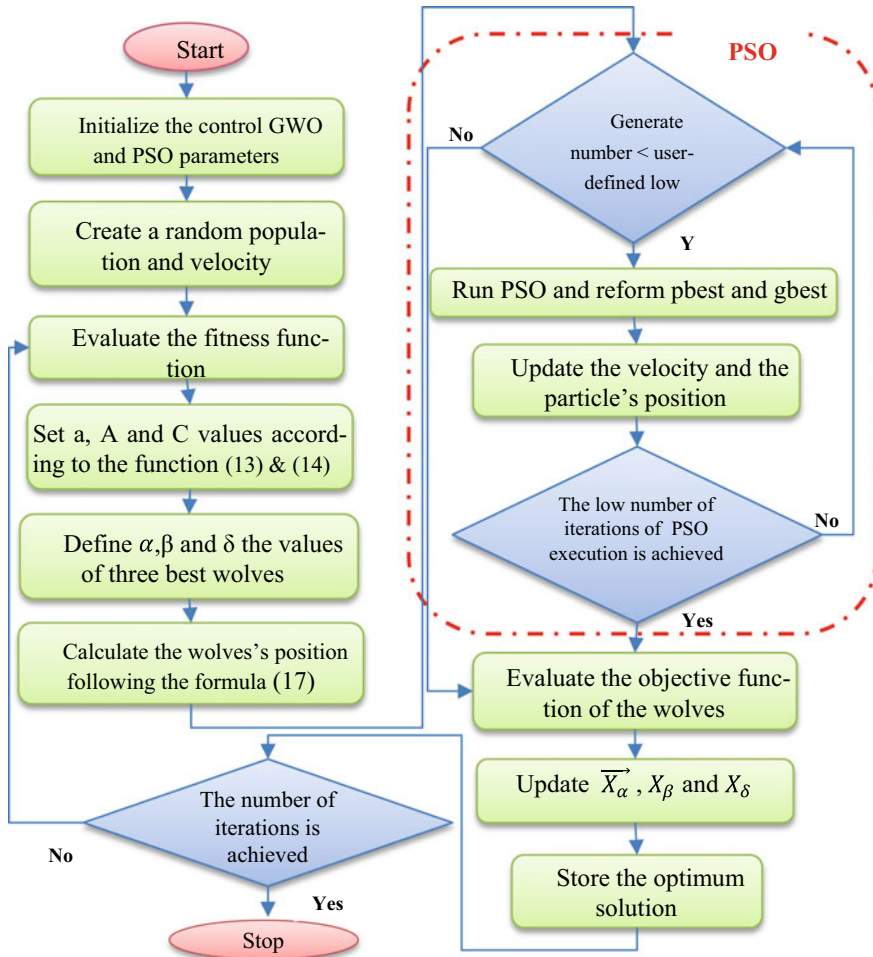
Fig. 1 Flowchart of the proposed hybrid HGWO-PSO algorithm

4 Simulation Process and Results

The HGWO-PSO has been tested on the IEEE-30 bus test system. The toolbox of MATPOWER contains the total data of this system [15]. The reliability of HGWO-PSO algorithm is proven using MATLAB by comparing it with PSO and GWO methods.

Figure 2 shows the algorithm convergence in the case where w_f1 = 1 and w_f2 = 0. The HGWO-PSO method achieves the lowest operation value cost of 822.6185$/h with a fast convergence in a record time at iteration 105 comparing to separate GWO and PSO which reached a cost of 823.445$/h for 242 iterations and 832.4755$/h after 136 iterations, respectively.
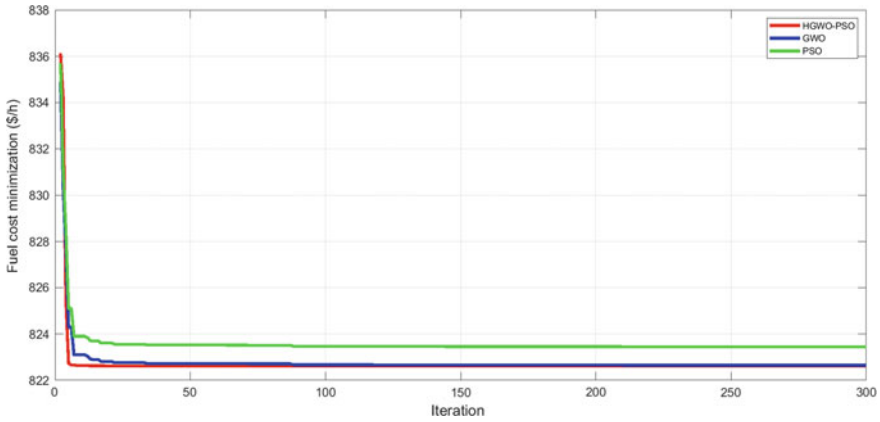
Fig. 2 Convergence characteristics for fuel cost

Indeed, Fig. 3 shows the convergence characteristics of CO_2 emissions. In the case where w_f1 = 0 and w_f2 = 1, the rate of emissions found by HGWO-PSO is equal to the one found by PSO (363.6757 kg/h). Whereas HGWO-PSO is more efficient because it obtains better results with minimum number of iterations according to Table 2. Table 2's bolded numbers show each algorithm's least power losses, minimum number of convergence iterations, and optimal values of the objective function for each algorithm.

Figure 4 shows the results of combining production costs with emissions, i.e., in the case, where w_f1 = 0.5 and w_f2 = 0.5, the total minimum cost found by GWO and PSO using the (PPF) is equal to 931.024$/h and 930.9825$/h, respectively. By applying the HGWO, the total cost was reduced to 930.0035$/h.
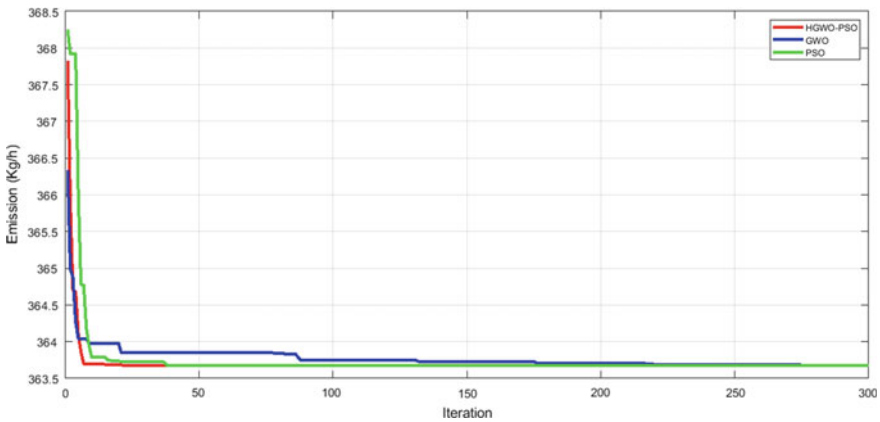
Fig. 3 Convergence characteristics for CO_2 emissions

Table 2 Simulation results of the HGWO-PSO, GWO, and PSO algorithms

		Fuel cost ($/h)	CO_2 emissions (kg/h)	Losses (MW)	Convergence in iteration	The generators power outputs (MW)					
						Gu1	Gu2	Gu3	Gu4	Gu5	Gu6
HGWO-PSO	$w_{f1}=1$ and $w_{f2}=0$	822.6185	518.1196	13.331	**105**	192.32	57.40	15.00	10.00	10.00	12.00
	$w_{f1}=0$ and $w_{f2}=1$	**909.4612**	**363.6757**	**5.918**	**48**	106.91	48.67	36.34	32.80	30.00	34.60
	$w_{f1}=0.5$ and $w_{f2}=0.5$	**930.0035**	**383.821**	**6.679**	244	117.99	94.99	30.86	33.30	28.74	29.18
GWO	$w_{f1}=1$ and $w_{f2}=0$	823.445	517.1729	13.451	242	194.91	51.90	15.00	11.00	11.55	12.54
	$w_{f1}=0$ and $w_{f2}=1$	909.5758	363.6758	5.919	293	106.93	48.66	36.33	32.83	30.00	34.57
	$w_{f1}=0.5$ and $w_{f2}=0.5$	931.024	430.8291	6.892	281	118.99	57.40	35.93	25.87	22.64	29.46
PSO	$w_{f1}=1$ and $w_{f2}=0$	832.4755	**494.859**	**9.426**	136	149.73	50.37	15.00	25.71	22.57	29.45
	$w_{f1}=0$ and $w_{f2}=1$	918.4333	363.6757	6.021	51	106.96	55.94	36.33	32.18	26.24	30.83
	$w_{f1}=0.5$ and $w_{f2}=0.5$	930.9825	383.8386	9.908	**63**	119.24	57.40	35.94	25.71	22.57	29.45
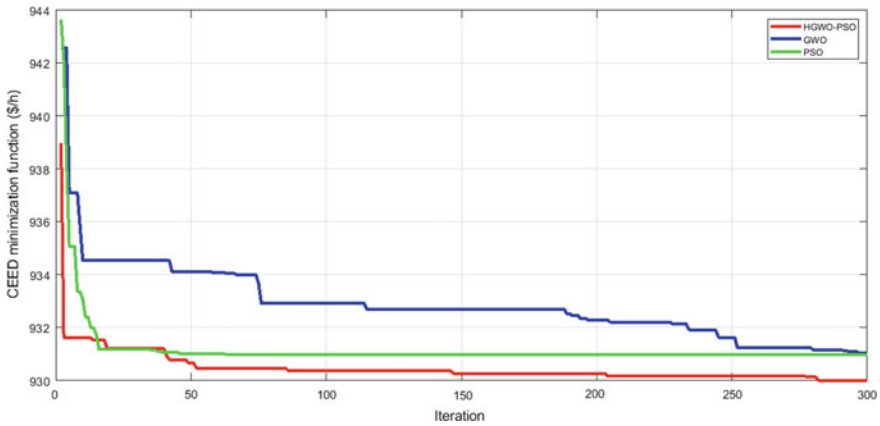
Fig. 4 Convergence characteristics for CEED optimization using HGWO-PSO, GWO, and PSO

The last six columns of Table 2 contain the power distribution of the six generators. Furthermore, for the 300 tests performed, the generators power production limits are respected in the three test cases. The analysis of the simulation results that the integration of HGWO-PSO proves to be a great saving with a minimum number of iteration and a minimum number of transmission losses. In fact, HGWO-PSO outperforms PSO and GWO in terms of obtaining the best total production cost combined with CO_2 emissions. This shows the effectiveness and efficiency of the proposed closure.

5 Conclusion

In this paper, a new application of the hybridization of two powerful algorithms, GWO and PSO, was presented for the optimal resolution of CEED problem. The experiment was conducted on IEEE 30 transmission framework and allowed to reduce the total cost of energy to an ideal total cost of 930.0035$/h. According to the simulation results, the introduced technique has the ability to obtain reasonable solution to address this problem. Many features can be considered a continuation of this current work. It may be more realistic to apply the same algorithm to large power systems with many control variables and intermittent sources. The goal is to exploit the advantages of the hybrid method to operate the system in an optimal way and in the best possible conditions.

References

1. An Y, Zhai X (2020) SVR-DEA model of carbon tax pricing for China's thermal power industry. Sci Total Environ 734:139438. https://doi.org/10.1016/j.scitotenv.2020.139438
2. Suman GK, Guerrero JM, Roy OP (2021) Optimisation of solar/wind/bio-generator/diesel/battery based microgrids for rural areas: a PSO-GWO approach. Sustain Cities Soc 67:102723. https://doi.org/10.1016/j.scs.2021.102723
3. Şenel FA, Gökçe F, Yüksel AS, Yiğit T (2019) A novel hybrid PSO–GWO algorithm for optimization problems. Eng Comput 35(4):1359–1373. https://doi.org/10.1007/s00366-018-0668-5
4. Shaheen MAM, Hasanien HM, Alkuhayli A (2021) A novel hybrid GWO-PSO optimization technique for optimal reactive power dispatch problem solution. Ain Shams Eng J 12(1):621–630. https://doi.org/10.1016/j.asej.2020.07.011
5. Gautam A, Sharma P, Kumar Y (2021) Mitigating congestion by optimal rescheduling of generators applying hybrid PSO–GWO in deregulated environment. SN Appl Sci 3(1):69. https://doi.org/10.1007/s42452-020-04084-0
6. Mishra AK, Das SR, Ray PK, Mallick RK, Mohanty A Mishra DK (2020) PSO-GWO optimized fractional order PID based hybrid shunt active power filter for power quality improvements. IEEE Access 8:74497–74512. https://doi.org/10.1109/ACCESS.2020.2988611
7. Qiao B, Liu J (2020) Multi-objective dynamic economic emission dispatch based on electric vehicles and wind power integrated system using differential evolution algorithm. Renew Energy 154:316–336. https://doi.org/10.1016/j.renene.2020.03.012
8. Prithi S, Sumathi S (2021) Automata based hybrid PSO–GWO algorithm for secured energy efficient optimal routing in wireless sensor network. Wirel Pers Commun 117(2):545–559. https://doi.org/10.1007/s11277-020-07882-2
9. Khaloie H et al (2020) Co-optimized bidding strategy of an integrated wind-thermal-photovoltaic system in deregulated electricity market under uncertainties. J Clean Prod 242:118434. https://doi.org/10.1016/j.jclepro.2019.118434
10. Dey B, Bhattacharyya B, Srivastava A, Shivam K (2020) Solving energy management of renewable integrated microgrid systems using crow search algorithm. Soft Comput 24(14):10433–10454. https://doi.org/10.1007/s00500-019-04553-8
11. Hussain S, Al-Hitmi M, Khaliq S, Hussain A, Asghar Saqib M (2019) Implementation and comparison of particle swarm optimization and genetic algorithm techniques in combined economic emission dispatch of an independent power plant. Energies 12(11):2037. https://doi.org/10.3390/en12112037
12. Wang X, Song W, Wu H, Liang H, Saboor A (2022) Microgrid operation relying on economic problems considering renewable sources, storage system, and demand-side management using developed gray wolf optimization algorithm. Energy 248:123472. https://doi.org/10.1016/j.energy.2022.123472
13. Liu W, Wang Z, Zeng N, Alsaadi FE, Liu X (2021) A PSO-based deep learning approach to classifying patients from emergency departments. Int J Mach Learn Cybern 12(7):1939–1948. https://doi.org/10.1007/s13042-021-01285-w
14. Shami TM, El-Saleh AA, Alswaitti M, Al-Tashi Q, Summakieh MA, Mirjalili S (2022) Particle swarm optimization: a comprehensive survey. IEEE Access 10:10031–10061. https://doi.org/10.1109/ACCESS.2022.3142859
15. Kyomugisha R, Muriithi CM, Edimu M (2021) Multiobjective optimal power flow for static voltage stability margin improvement. Heliyon 7(12):e08631. https://doi.org/10.1016/j.heliyon.2021.e08631

VSAS Models for Energy Processes and a Greenhouse Simulation and Control

Nacer K. M'Sirdi and Fabrice Aubépart

Abstract A Variable Structure Automatic Systems (VSAS) modeling approach is proposed for cyber-physical processes using renewable energies. It is developed for a greenhouse dynamic analysis. It is also applicable for control of vehicles dynamics and cooperating robots. The VSAS models have energy exchanges with environment and are subject to commutations and switching.

1 Introduction

Several Multi-Input Multi-Output (MIMO) plants have intermittent interactions with their environment and/or other neighbor processes. This class of systems is in the main interest of this work.

They are in general nonstationary, time varying or/and nonlinear and operate in changing environment conditions.

The climate and environment dependent processes are affected by switched phenomena with intermittent effects. These are driven by some events (wind blowing, rain, rising sun, shadows, etc.). Their effects on the plant operation mode, are switched on and off regard to the associated events.

We can try to stabilize the climate, inside a greenhouse, by controlling of those effects based on multiple feedback loops and Lyapunov functions. One can also,

N. K. M'Sirdi (✉) · F. Aubépart
Aix Marseille Univ, Université de Toulon, CNRS, LiS UMR CNRS 7020, Avenue Escadrille Normandie Niemen , 13397 Marseille Cedex 20, France

HyRES Lab, Marseille Cedex 20, France
e-mail: nacer.msirdi@lis-lab.fr
URL: http://nkms.free.fr/MGEF/HyRESLab.htm

develop several observers, predictions and then be able for diagnosis and test to monitor the plant.

Our project deals with diagnosis and control of the behavior of a VSAS process, after state observation and estimation of its non-measured variables. The considered process is a Greenhouse equipped with IoT and wireless sensors.

We propose in this paper a modeling approach based on Variable Structure Automatic Systems (VSAS) to cope very well to plant the dynamics. We will apply this kind of model to a greenhouse and show that it allows estimations and identification of parameters for the separate phenomena (Temperature, humidity, condensation, etc.).

The piece-wise continuous equations (linear or note) are well suited to get Ordinary Differential Equations (ODEs)-based models for Energy sources, greenhouses [1–3] and some robotic applications.

Some greenhouse modeling methods use a classic linear model with context-dependent Markov-switching [4]. Other ones are based on hierarchical fuzzy logic commutation and supervision. Nonlinear extension based on PCA are preferred in [1]. Markov chains [4] and Kalman filters [5] can be used for prediction and then to drive the commutations [6].

Neural Networks and Fuzzy Logic or neuro-fuzzy techniques for learning to get black box local representations have also been used [3].

The local models do not have physical interpretations and depend on the operating points.

When using a linear model, the parameters are fast time varying during operation in a changing environment. The parameters identification will be difficult [3, 7, 8]. The control will then be difficult to adjust in time for each operating region.

In [1] authors combine known part of the involved phenomena (gray box) in a hybrid representation.

Knowledge on the events is needed to track the structure variations and to be able to observe or identify the model parameters [9].

For the system state observation and the parameters identification, adequate models are required [8]. Event governed models are used in [9]. The events occurrences are assumed known.

For the Complex Nonlinear systems in interaction with their environment, we propose the Variable Structure Automatic Switched Systems (VSAS) modeling approach. This model class easily describe, simulate and control the behavior of a Greenhouse. This will simplify the process stabilization, the diagnosis and monitoring.

This approach is modular and well adapted to use of IoTs for measurements by wireless sensors, acquisition and local control loops. We can eventually use different sampling periods. The control can then be composed by modular feedback loops and adapted trough IoTs.

We present in Sect. 2 the formulation and the stability tools useful for this approach. Section 3 illustrates the effectiveness of this approach for a greenhouse. Section 4 concludes by discussion on Greenhouse results.
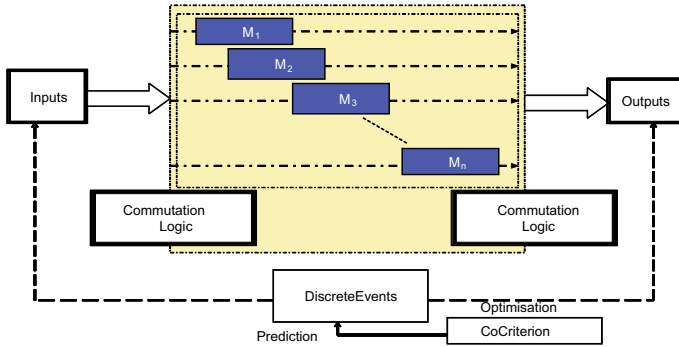
Fig. 1 VSAS nonlinear model structure for switched system

2 Problem Formulation

2.1 VSAS Systems Model

Let us use a similar approach than the one we have developed previously in [1, 8], which is described by Fig. 1.

The main features of this switched multi-model approach, are the modular structure, the combination of switched models (M_p) (representing partial dynamics of the plant) and compatibility with use of IoTs and local sensors.

The sub-models describe the process dynamics (due to the involved phenomena) in different state space regions.

$D_p \subset \mathbb{R}^n$ and time periods (for $p = 1, 2, \ldots, P$) $(\dot{z}_p = f_{m_p}(z_p, u, t), \ t \in \mathbb{R}$.

The subspace $D_p \subset \mathbb{R}^n$ correspond to activity of a phenomenon and then the (action) validity of the corresponding sub-model.

The global system gathers the sub-models (addition of the phenomena effects) activated by some logic conditions of Switching ($\xi_p(t_k)$ depending on operating zone) indicating when they are active.

As consequence we get the proposed VSAS model:

$$\begin{cases} \dot{z}_p(t) = f_{m_p}(z_p, x_e, u, t) & \text{for } p = 1, \ldots, P \\ \xi_p(t_k) = F[t_{k-1}, u, x, z_p, x_e, t] & \text{for } p = 1, \ldots, P \\ \dot{x}(t) = g_m(x, x_e, u, z_p, f_{m_p}, \xi_p, t) = \sum_{p=1}^{P} \xi_p(t_k) \cdot f_{m_p}(x, x_e, u, z_p, f_{m_p}, t) \end{cases} \quad (1)$$

where

- u define the inputs used to control the heating, the brumisation, the ventilation, the crop watering, windows opening, etc.),
- x_e are the external inputs external temperature, humidity, radiation, wind (T_e, H_e),
- $x(t)$ is continuous state vector for the greenhouse,

- M_p is the sub system describing one phenomenon (heating, ventilation, sun radiation effect, vaporization, misting, etc.),
- z_p is the (partial) state vector affected by one phenomenon M_p,
- ξ_p is the weight of the contribution of the model M_p in the dynamics of the global system,
- F is the discrete event function defining the switching ξ_p (heating on/off, ventilation dark-night,sunny day, sun radiation effect, vaporization),
- t_k time instant when the weights $(\xi_p(t_k))$ changes.

For the P sub-models we have the variables $\xi_p(t_k)$ defined by Eq. (1) of F, from the system state, the sub-model state, the inputs, the external inputs and the control inputs by the following equation:

$$\xi_p(t_k) = F[t_{k-1}, u, x, z_p, x_e, t] \quad \text{for } p = 1, \ldots, P \quad (2)$$

The variables $\xi_p(t_k)$ are used for weighting the effects of the different phenomena in the models combination.

The composition of the function F depends on the **Time periods**. For example the day can be split in four time periods, at least [10]: • night (or dark-night), • day (sunny day) or • daybreak and • night-break. During the day solar radiation is on and contribute to heating. During the night there is no radiation and the outside temperature is lower so the condensation start on the canopy and the roof. It will depend on temperature difference, canopy and roof surfaces. During night-break and daybreak the radiation, the external temperature and humidity will be different. The time is needed for selection and weighting of corresponding the corresponding sub-model equations, like night, night-break, day or daybreak, when solar radiation is null, or weak and increasing, or full or weak and decreasing (respectively). Then condensation can appear or not. The discrete event function F depends on the inputs, the wind, the ventilation, the misting and the heating can be activate or not.

The system sub-models combination is governed by a **logic function** of the plant behavior, the state variables and external inputs (e.g., temperature is too low (respectively high) then start heating (respectively ventilation).

If, during the dark-night the outside temperature in high (like in summer) the condensation will not occur and the heat still flows from outside to the inside of the greenhouse.

The event function F depends on the state variables or feature like condensation, on snow on the roof or on the canopy among others. It depends also on the **sub-models Interactions**.

Equation (1) defines what is called the VSAS class of models which is driven by functions of Discrete Events (2).

This class is more close to the system for physical interpretation and more efficient for description than the literature proposed ones. Compared to the hybrid systems description of [1–4], the VSAS models are more easy to use.

The obtained model is modular and more simple. It is more compatible to use IoTs, wireless sensors and remote decentralized control. We can easily observe and estimate the systems states $\hat{\xi}(t)$. The sub-models M_p are physically interpretable.

2.2 Stability Tools for VSAS System

As in the VSAS class there are several sub-models Mp, we can use multiple Lyapunov's functions for stability analysis [1].

The plant switching nature requires use of several Lyapunov functions corresponding to the sub-models commutations and switches $\xi_p(t_k)$ [11–13].

Let us recall, for stability analysis, the Lyapunov theorem.

Theorem 1 (Lyapunov's Stability) *Let z_{ep} be an equilibrium point in domain D_p, for the system M_p (in Eq. 1 and a chosen real valued Lyapunov function $V_p(z)$ in the domain D_p. We get the local stability if $V_p(z)$ satisfies the following conditions:*

(C1): $V_p(z) > 0$ for $z \neq 0$ and $V_p(0) = 0$
(C2): $\dot{V}_p(z) < 0$: for any z, $z \in D_p$: $\dot{V}_p(z) = \frac{\partial V_p(z)}{\partial z} f_p(x, u, t) \leq 0$.

Theorem 2 (Composed Lyapunov function) *For an equilibrium point to $x_e = 0$ of the VSAS plant, assume that, for each vector field $f_p(z_p, u, t)$ of a P-switched plant,we define a Lyapunov function $V_p(z_p)$, which verify theorem 1. As $\xi_p(t_k)$ define the event sequence of switching the M_p sub-models, then we have in addition the condition:*

(C3): $V_p(z_p(t_{k+1})) \leq V_p(z_p(t_k))$ for all the switching times t_k.

The set V gathers the P Lyapunov candidates in $V = \{V_1, \ldots, V_p\}$, we get a description of the energy flows in the switched system.

The energy evolution depends on the discrete events $\xi_p(t_k)$, the sub-models $f_p(x, u, t)$.

To ensure stability (an always decreasing energy), over the models M_p switching, the Lyapunov function must decrease (after switching) or be controlled with a limited growth.

3   Greenhouse VSAS Model

One use a greenhouse to create optimal conditions for plant growth. It is used to soften the weather conditions. It tracks (measure and detect) the external disturbances, like hard solar radiation, temperature excess, wind blow, etc. It must avoid the disadvantages, reduce the wind and rain effects by the control of temperatures, the internal hydrometry and the carbon dioxide.

Agricultural greenhouses are in general made of the following sub-systems , at least:

- A room with canopy and roof,
- A heating, by hot air flow or a hot water circulation buried tubes; solar energy is used for water heating.
- A Misting: water vaporization by a ramp running through the greenhouse.
- Air Ventilation, by manual sashes and the roof openings.

3.1 Greenhouse Description

It is a simple and small greenhouse with glass canopy and roof which can be placed in a house garden. It interests more and more families, after the period of confinement. The greenhouse inside Surface is $18.70\,\mathrm{m}^2$ (6.04 m long and 3.09 m large). The height varies from 1.8 to 2.47 m when the roof is open. The total Air Volume inside is $40\,\mathrm{m}^3$.

A greenhouse has at least, 8 inputs and 2 outputs to be controlled:

- 4 actuators as inputs for misting, openings, shading and heating
- 4 disturbances inputs or measurements which are solar radiation R_g (W/m2), external hydrometry H_e (%), external temperature T_e (°C), and wind speed v_w (km/h),
- 2 outputs to control: internal temperature T_i (°C) and hydrometry H_i (%).

The greenhouse heating is done by hot water circulation in buried tube under the ground. Water is heated by a PVT panel harnessing solar energy. The energy harnessing and storage is not considered in this paper.

For Humidity control Water is vaporized.

Air flow control is done by Ventilation and (walls and the roof) windows openings.

Wireless sensors, connected to IoT systems are used for data acquisition and the remote control.

We start this project with a small set of sensors as we are interest first by modeling study.

The required sensors for full observability and controllability of greenhouse operation and plant will be studied in the next work.

Combined sensors give measurement of the internal air humidity and temperature. The greenhouse sensors used are:

- Temperature sensors of the rooting substrate,
- External sensors of the greenhouse,
- A combined humidity-temperature sensor,
- A global radiation sensor (insulation),
- A wind speed sensor,
- Soil temperature is measured near the crop rooting substrate,
- Outside Sensors measure the external temperature and air humidity,
- Sensor to measure solar radiation and wind speed.

The following states Variables are also considered:

- the state of the substrate watering,
- the crop deficit in Water,
- the concentration of CO_2 in the greenhouse,
- the humidity saturation of Air,
- We consider also several sensors to get measurements in several points (inside and outside) in the greenhouse environment.

Measurement and Control System: Several connected nodes are used to collect data from the measured quantities in the greenhouse (Temperature, Humidity, radiation, wind speed, etc.). They constitute a wireless network. Each nodes associates a LORA SX1278 chip from Semtech and a module integrating a 32-bits micro-controller (ESP32). The node is associated with an XBee communication chip in order to allow the possibility of 3 wireless communication protocols modes: LORA, Zigbee and Wifi, see Fig. 1.

The WiFi and Bluetooth are integrated by use of a micro-controller (ESP32) from Espressif Systems (Table 1).

3.2 The VSAS Physical Model of the Greenhouse

Several physical phenomena are combined in the Greenhouse behavior. They occur as water and vapor fluxes transfers (Q_i) and heat transfers (Φ_i) (radiations). They are gathered as $Q_S = \sum_i Q_i$ and $\Phi_S = \sum_i \phi_i$.

In this paper, we consider only the temperature and humidity.

The greenhouse thermal and humidity behavior can be described by the following four equations involving the states of internal temperature, internal humidity, soil temperature and the heating fluid temperature (t_i, H_i, T_s, T_f).

$$
\begin{aligned}
\frac{dT_i}{dt} &= \frac{1}{C_p} \cdot \Phi_S(t) \\
\frac{dH_i}{dt} &= Q_S \\
\frac{dT_s}{dt} &= \phi_{\text{heating}} - \phi_{\text{loss}} \\
\frac{dT_f}{dt} &= -\frac{h_{pf} \cdot A_f}{m_f \cdot C_f} \cdot (T_f - T_s) + \frac{\dot{m}_e \cdot c}{m_f \cdot C_f} \cdot (T_f - T_{\text{fin}})
\end{aligned}
\tag{3}
$$

where Q_S is the water mass balance of the air humidity and Φ_S is the sum of the heat flows weighted by the corresponding $\xi_p(t)$ (in Eqs. (1) and (2)).

The last two equations describe the heating system. The ground heating receives the heat flow $\phi_{\text{heating}} = \frac{h_{sf} \cdot A_f}{m_s \cdot C_s} \cdot (T_{\text{fout}} - T_s)$ given by the circulation of the heating fluid which has as temperature T_f between T_{fin} at the inlet and T_{fout} at the outlet when leaving the greenhouse.

The quantity of heat loss from the ground is transmitted to the air inside the greenhouse $\phi_{\text{loss}} = \frac{h_{si} \cdot A_s}{m_s \cdot C_s} \cdot (T_s - T_i)$.

Table 1 Inputs and state variables and parameters

Variable	State variables	Units
T_i	Temperature inside air	°C
H_i	Humidity rate inside	kg/m^3
T_s	Soil temperature	°C
T_f	Heating fluid temperature	°C
T_{fout}	Fluid outlet temperature	°C
Variable	Environment variables	Units
H_e	Humidity of air outside	kg/m^3
R_g	Solar radiation	
T_e	Outside temperature	°C
T_{fin}	Inlet fluid temperature	°C
T_b	Temperature of the misting water	°C
v_w	Velocity of the wind	m/s
q_a	Air flow ventilation	g/s
H^*	The water content % at saturation	
Variable	Input variables	Units
$\dot{m}_f$	Water flow in the heating circuit	l/s
R_v	Air flow rate control	–
F_b	Misting flow control	–
R	Ventilation air flow Rate $R = \frac{1}{V_g} q_a$	l/s
Variable	System parameter	Units
S_e	Surface of Canopy exposed to sun	m^2
A_s	Greenhouse soil surface	m^2
A_f	Heating circuit contact surface	m^2
A_p	Canopy surface	m^2
V_g	Greenhouse air volume	40 m^3
τ	Absorption coefficient of the soil substrate	
K_c	Heat transmission coefficient of the roof	
a and b	The crop evaporation parameters	–

The heating system injects the fluid mass quantity $\dot{m}_c$ inside the greenhouse buried tube.

3.3 VSAS Flows Commutations

The heat flow balance Φ_S inside the greenhouse volume at time t is the sum of input flows Φ_{in} and the produced flow Φ_{prod} minus the losses Φ_{loss} and the flow Φ_{Cons} consumed by the crop (assumed null in this paper).

$$\Phi_S(t) = \Phi_{\text{in}} + \Phi_{\text{prod}} - \Phi_{\text{Cons}} - \Phi_{\text{loss}}$$
$$\Phi_{\text{in}} = \Phi_{\text{Ray}} + \Phi_{\text{Vent}} + \Phi_{\text{Brum}} + \Phi_{\text{Heat}} \qquad (4)$$
$$\Phi_{\text{prod}} = \Phi_{\text{Conv-Sol}} + \Phi_{\text{Plants}} + \Phi_{\text{Conv-Pers}}$$

The input heat flow Φ_{in} gathers heat from solar radiation, ventilation, misting and heating. The soil convection, the crop, and people in the greenhouse lead a produced flow Φ_{prod}.

Thermal flows analysis led us to

$$\Phi_{\text{loss}} = C_p \cdot h_p \cdot \frac{A_p}{V_g} \cdot (T_i - T_e)$$
$$\Phi_{\text{Conv-Sol}} = h_s \cdot \frac{A_s}{V_g}(T_i - T_s)$$
$$\Phi_{\text{Vent}} = -C_p \cdot \frac{dT_i}{dt} = -R \cdot C_p(T_i - T_e) \qquad (5)$$
$$\Phi_{\text{Brum}} = L_v \cdot \frac{F_b \cdot (T_i - T_e)}{V_g}$$
$$\Phi_{\text{Chauf}} = m_s \cdot C_s A_f \cdot (T_s - T_f)$$
$$\Phi_{\text{cons}} = 0$$

The air humidity balance (vapor mass balance Q_S) gathers the soil evaporation and condensation on canopy and roof, the misting and the ventilation effect.

Study of humidity flows combination gives us:

$$Q_1 = R_b/1.3 \cdot Vg$$
$$Q_2 = (a + b \cdot R_g) \cdot D_i$$
$$Q_3 = \frac{1}{V_g}\frac{dv}{dt}(H_i - H_e) = (H_i - H_e) \cdot R_v$$
$$Q_4 = (a_s + b_s \cdot R_g) \cdot (H^* - H_i) \qquad (6)$$
$$Q_5 = C_{\text{paroi}} \cdot C_{m,h}(c_{\text{res}} \cdot H^* - H_i)$$
$$Q_6 = C_{\text{roof}} \cdot C_{m,h}(H_{\text{rouf}}(T_e) - H_i)$$
$$Q_S = Q_1 + Q_2 + Q_3 + Q_4 + Q_5 + Q_6$$

where $R = \frac{1}{V_g}\frac{dV_g}{dt} = \frac{1}{V_g}q_a$ is the volume rate of air exchange produced by the ventilation and $dV_g/dt = q_a$ the air flow.

In a previous work an experimental study proved that we can split a day in four periods: night, night-break, day and daybreak [1].

During the night there is no solar radiation but heating may be necessary and ventilation may be stopped. This allow to mind about definition of the weighting commutation functions for each sub-model.

Each line of In Eqs. (5) and (6) we will weight each Q_i and Φ_i by the corresponding $\xi_p(t_k)$.

The $\xi_p(t_k)$ can be equal to 1 when active and to 0 when the function is inactive. If needed, smooth transitions from 0 to 1 and reciprocally can be used. As an example the the sun radiation can be weighted by 0 during dark-night, 1 in the day and growing from 0 to 1 during the night-break (which is from 5:00 to 7:30 mornings during the Summer-days) and decreasing from 1 to 0 during the daybreak (from 19:30 to 22:00).

This finally gives us the VSAS model of the greenhouse:

$$\begin{aligned}
\frac{dT_i}{dt} &= \frac{1}{m_p \cdot C_p} \cdot \Phi_S(t) \\
\frac{dHi}{dt} &= Q_1 + Q_2 + Q_3 + Q_4 + Q_5 + Q_6 \\
\frac{dT_f}{dt} &= -\frac{h_{pf} \cdot A_f}{m_f \cdot C_f} \cdot (T_f - T_s) + \frac{\dot{m}_e \cdot c}{m_f \cdot C_f} \cdot (T_f - T_{\text{fin}}) \\
\frac{dT_s}{dt} &= -\frac{h_{si} \cdot A_s}{m_s \cdot C_s} \cdot (T_s - T_i) - \frac{h_{sf} \cdot A_f}{m_s \cdot C_s} \cdot (T_s - T_{\text{fout}})
\end{aligned} \tag{7}$$

The VSAS model class is then modular and adjustable by commutations. We can also adjust the weighting functions depending on the world region and the period in the year.

4 Conclusion

The VSAS class gives models easy to handle for physical systems description. It can give incremental models. It is also modular and efficient to use with IoTs-based hardware.

The VSAS modeling approach copes very well to model the dynamics of a greenhouse, buildings, renewable energy sources and plants composed by multiple subsystems. It uses the material and energy balances, and then it is adequate for physical models development, for the stability analysis, for observation and diagnosis and also for control. It leads to systems which can easily be verified and managed.

The VSAS model developed in this paper is modular and copes well for equipment of the measurements and control by wireless sensors and IOTs. Control and measurement can also be done with different and adapted sampling period for acquisition and control. Trough IoTs we can apply decentralized measurement and local feedback loops for control. Simulations can be done by incremental models end then are well adapted to process prototyping.

The next works will focused be on the diagnosis based on observations and predictions. The next step will then be to design robust control approaches which may efficiently implemented using wireless technologies, IoTs and distributed electronic computing.

References

1. M'Sirdi NK, Rajaoarisoa A, Naamane LH, Duplaix J, Balmat J (2007) Switched complex system analysis for modeling, control and diagnosis. Int J Sci Tech Autom Control. hal-01966643
2. Monteils C, Amouroux M (1993) Analyse thermique de systèmes de chauffage localiséá basse température dans les serres agricoles: paillage radiant et tubes entérréés. J Phys III France 3(1993):1981–2031
3. Salgado P, Boaventura Cunha J (2005) Greenhouse climate hierarchical fuzzy modelling. Control Eng Pract 13:613–628

4. Murray-Smith R, Johansen T, Shorten R (1997) Multiple model approaches to modeling and control. Taylor and Francis, London

5. Theilliol D, Sauter D, Ponsart J (2003) A multiple model based approach for fault tolerant control in non-linear systems. IFAC Proc Vol 36(5):149–154. 5th IFAC symposium on fault detection, supervision and safety of technical processes 2003, Washington, DC, 9–11 June 1997 (online). Available at: https://www.sciencedirect.com/science/article/pii/S1474667017364856

6. Narendra K, Balakrishnan J, Ciliz M (1995) Adaptation and learning using multiple models, switching, and tuning. IEEE Control Syst Mag 15(3):37–51

7. Monteils C, Issanchou G, Amouroux M (1991) Modél energétique de la serre agricole. J Phys III1:429–454

8. Rajaoarisoa L, M'Sirdi K, Boukas EK, Balmat JF, Duplaix J (2008) Greenhouse climate modeling : observability and identification for supervision. In: Proceedings of the 5th international conference on informatics in control, automation and robotics (ICINCO 08), Portugal, 2008 (online). Available at: https://hal.archives-ouvertes.fr/hal-01778919

9. Belhani A, M'Sirdi NK (2011) An observer based backstepping control for a greenhouse. In: 2011 international conference on communications. CCCA, computing and control applications

10. M'Sirdi NK, Aubepart F, Belhani A (2021) Smart autonomous greenhouse design modeling and simulation as a variable structure automatic system. In: 2021 9th international renewable and sustainable energy conference (IRSEC), Nov 2021, pp 1–7

11. Branicky M (1998) Multiple Lyapunov functions and other analysis tools for switched and hybrid systems. IEEE Trans Autom Control 43(4)

12. Donde V (2001) Development of multiple Lyapunov functions for a hybrid power system with a tap changer. In: ECE 497DL. ECE 497DL, Spring 2001

13. Decarlo RA, Branicky MS, Pettersson S, Lennartson B (2000) Perspectives and results on the stability and stabilizability of hybrid systems. Proc IEEE 88:1069–1082

Impact of Wind Power Integration on the Moroccan Electrical Grid Reliability

M. El Fahssi, T. Ouchbel, S. Zouggar, M. Larbi Elhafyani, M. Oukili, M. Maaouane, and M. Chennaif

Abstract Wind energy integration into Morocco's electricity grid is increasing. Due to the unpredictable nature of wind, this increase can have a significant impact on the electrical system's ability to handle the load. As a result, the goal of this research is to employ a non-sequential Monte Carlo simulation to assess the impact of wind power and load evolution on the reliability of the Moroccan electrical network at the hierarchical level HLI (HLI: load covering ability under the assumption of infinite node). To accomplish so, a set of empirical equations has been used to calculate Weibull parameters based on wind medium speed and to quantify the electrical power generated by wind farms in Morocco. Following that, the modeled wind power was integrated into the developed Monte Carlo simulation program to evaluate the influence of increased wind power penetration on Moroccan power system reliability.

Keywords Reliability · Wind energy · Monte Carlo simulation · Weibull parameters

1 Introduction

In order to decrease its strong energy dependence, Morocco adopted its National Energy Strategy (NES) with corresponding targets for 2020 in 2009 and renewed it in Paris end of 2015 with targets until 2030. It challenged the three principal challenges of modern energy policy, security of supply, affordability, and sustainability.

M. El Fahssi (✉) · T. Ouchbel · S. Zouggar · M. Maaouane · M. Chennaif
Laboratory of Electrical Engineering and Maintenance (LEEM), School of Technology Oujda, University Mohammed 1st, BP, 473, 60000 Oujda, Morocco
e-mail: mohammed.elfahssi@ump.ac.ma

M. Larbi Elhafyani
National School of Applied Sciences, University Mohammed 1st, Oujda, Morocco

M. Oukili
Faculty of Sciences and Technology, University Cadi Ayyad, Marrakech, Morocco

Among the goals of this NES is securing energy supply, especially by reducing the dependence on imported energy carriers through the development of domestic RE sources (from 96% in 2015 to 82% by 2030) [1] and the increased exploration of conventional energy sources.

Based on the NES's goals and the related National Priority Action Plan (PNAP, 2009/2015) set target to increase the total installed capacity of renewable energy (RE) in the electricity sector to 42% by 2020 and to 52% by 2030 [2].

Morocco has huge potential and favorable conditions especially for wind and solar energy. Due to its long coastline with high wind speeds (up to 11 m/s), the total wind power potential is estimated at around 5000 TWh/year [2].

A major plank in Morocco's NES is the progressive rollout of RE technologies through the Moroccan Integrated Wind Program, the Moroccan Solar Plan as well as a continuation of the country's hydro-electric plans.

The Moroccan Integrated Wind Program aims to achieve 2000 MW installed wind power capacity by 2020 and up to 5000 MW by 2030 (additional 4200 MW from 2016 to 2030) [1]. In light of this plan, Morocco has opted to build many wind farms to generate electricity. But, this haphazard energy source has a significant impact on the system's ability to cover the load.

The Weibull parameters (k, c) are calculated by the empirical equations of each wind farm for the twelve considered regions. Then, the fluctuating electrical wind power is then generated. Finally, a non-sequential Monte Carlo simulation [3–10] by well-being indices is used under the MATLAB environment in order to quantify the impact of increased wind power penetration on the Moroccan power grid.

The paper is organized as follows. Section 2 presents electrical wind power in Morocco. Next, Sect. 3 introduces the impact of wind power on the reliability of the Moroccan electrical grid. Then, simulation results and interpretations are discussed in Sect. 4. Finally, Sect. 5 presents conclusions.

2 Electrical Wind Power in Morocco

2.1 Weibull Parameters of Moroccan Wind Farms

Based on wind medium speed taken from Moroccan wind map Fig. 1 and the empirical Eqs. (1) and (2), we calculated the Weibull parameters (k, c) [11] of each wind farm for the twelve considered regions, Table 1.

$$k = 0.9 + 0.2V_m \tag{1}$$

$$c = \frac{V_m}{\Gamma[1 + (1/k)]} \tag{2}$$
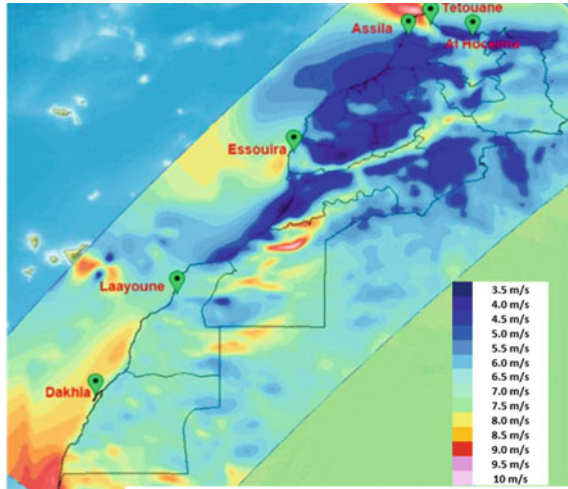
Fig. 1 Moroccan wind map

Table 1 Moroccan wind farms at 2019

Wind farm	Installed power (MW)	Region (province)
Akhefnir 1	101.5	Tarfaya
Amogdoul	60	Essaouira
Foum al Oued	50.6	Laayoune
Tanger I	140	Tanger
Tarfaya	301	Tarfaya
Haouma	50.6	Tanger
A.Torres (Koudia El Baida)	53.5	Tétouan
Afftissat	201.6	Boujdour
Jbel Khalladi	120	Ksar Sghir
Akhefnir 2	100.23	Tarfaya
CIMAR	5	Laayoune
LAFARGE	32	Tétouan

V_m: the wind medium speed, c: the scale parameter (information on the average wind speed), and k: the shape parameter (wind repartition around the average value).

The Weibull parameters for each of the twelve wind farms are presented in Table 1. and are validated by the Moroccan wind map (Fig. 1; Table 2).

Table 2 Calculated Weibull parameters for each wind farm

Wind farm	k	c (m/s)	V_m (m/s)
Akhefnir 1/Akhefnir 2	2.1	6.77	6
Amogdoul/Tarfaya	2.3	7.9	7
Foum al Oued/Haouma/A.Torres (Koudia El Baida)	2.4	8.46	7.5
Afftissat/Jbel Khalladi/LAFARGE	2.5	9.02	8
Tanger I/CIMAR	2.6	9.57	8.5

2.2 Wind Speeds in the Different Regions

To determine the hourly fluctuations of the power generated by wind farms, we need to sample the hourly wind speeds. In that way, we use the method based on the inversion of the cumulative distribution function of Weibull associated with each of the investigated regions. Practically, the procedure is divided into two steps [6, 12]:

- Initially, a uniformly distributed on the interval [0, 1] random number is assigned to P.
- Then, the wind speed for each value of P is determined by Eq. (3):

$$V = c(-\ln(1 - P))^{1/k} \qquad (3)$$

Figure 2 shows the hourly wind speeds generated for Foum El Oued park using the proposed sampling method. It should be noted here that each time evolution is independent since, at each hour, a different random number (between [0, 1]) is sampled for each of the twelve regions.

2.3 Global Generated Wind Power by the Considered Wind Farms

In this study, we use the Moroccan wind farms operated in 2019. The technical description of each of these twelve parks is given in Table 3 [7, 10].

The previously generated wind speeds (Sect. 2.2) allow calculating the hourly electrical power produced by the twelve considered wind farms. Indeed, based on the wind turbines characteristics of the investigated wind parks, it is possible to convert each hourly wind speed into generated power (Fig. 3).

The power curve associated with all kinds of wind turbines is illustrated in Fig. 3.

When converting wind speeds into power, it is supposed that all wind is subject to the same hourly wind speed. Therefore, every hour, one different wind speed is sampled for each wind farm and converted into power using the adequate power curve (Fig. 3). The hourly wind power produced in Morocco is then achieved by adding the hourly power generated for each of the twelve wind farms (4).
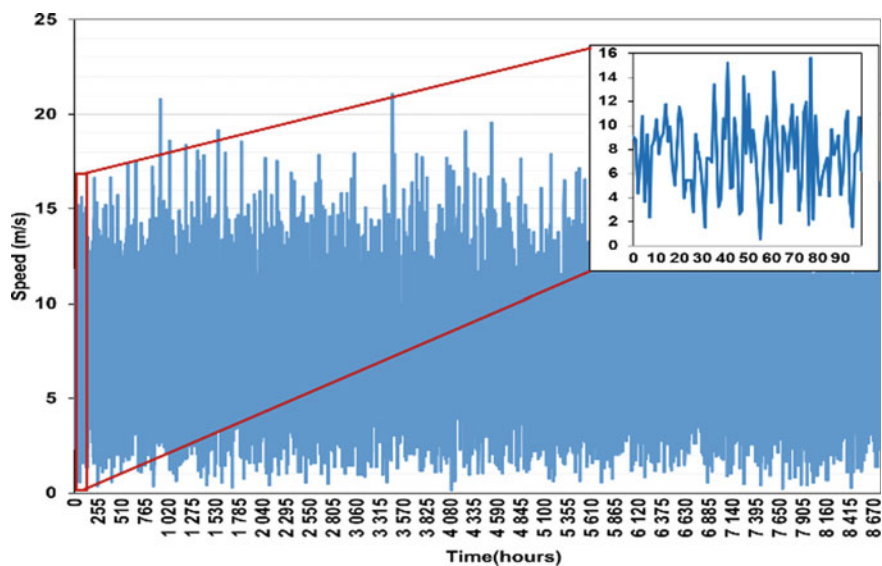
Fig. 2 Hourly wind speeds generated for Foum El Oued park

Table 3 Technical descriptions of all considered wind farms

| Wind farm	Turbine	Number of turbines
Akhefnir 1	Alstom Eco74	61
Amogdoul	Gamesa G52/850	71
Foum al Oued	Siemens SWT-2.3-101	22
Tanger I	Gamesa G52/850	165
Tarfaya	Siemens SWT-2.3-101	131
Haouma	Siemens SWT-2.3-93	22
A.Torres (Koudia El Baida)	Vestas V44/600 Enercon E40/500	83 7
Afftissat	Siemens 3.6-130	56
Jbel Khalladi	Vestas V90/3000	40
Akhefnir 2	GE 1.7-100	56
CIMAR	Gamesa G52/850	6
LAFARGE	Gamesa G52/850 Gamesa G80/2000	12 11
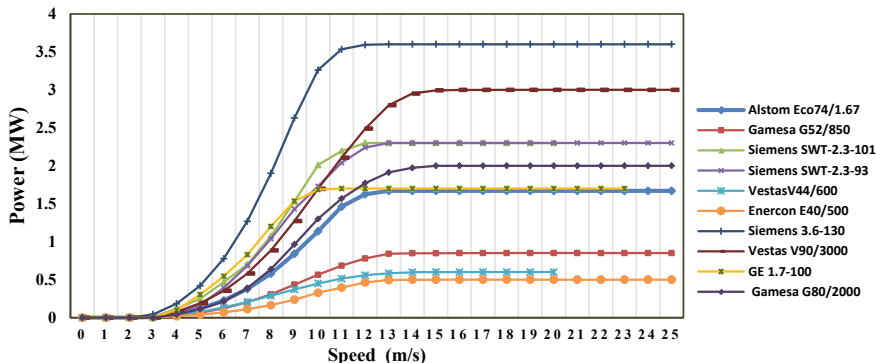
Fig. 3 Power curve for considered wind turbines

$$P_G = 61 P_{A1} + 56 P_{A2} + 71 P_M + 22 P_F + 165 P_{TN} + 131 P_A$$
$$+ 22 P_H + 83 P_{T1} + 7 P_{T2} + 56 P_{AF} + 40 P_J + 6 P_C$$
$$+ 12 P_{L1} + 11 P_{L2} \tag{4}$$

With

- P_G the global electrical power produced by all parks;
- P_{A1} the power produced by the Akhfenir 1 wind park;
- P_{A2} the power produced by the Akhfenir 2 wind park;
- P_M the power produced by the Amogdoul wind park;
- P_F the power produced by the Foum al Oued wind park;
- P_{TN} the power produced by the Tanger I wind park;
- P_{TA} the power produced by the Tarfaya wind park;
- P_H the power produced by the Haouma wind park;
- P_{T1} the power produced by the Koudia El Baida wind park (VestasV44/600);
- P_{T2} the power produced by the Koudia El Baida wind park (Enercon E40/500);
- P_{AF} the power produced by the Afftissat wind park;
- P_J the power produced by the Jbel Khalladi wind park;
- P_C the power produced by the CIMAR wind park;
- P_{L1} the power produced by the LAFARGE wind park (Gamesa G52/850);
- P_{L2} the power produced by the LAFARGE wind park (Gamesa G80/2000).

Figure 4 shows the global power produced by the twelve parks for 8760 h. Again, the fluctuating nature of the latter is clearly illustrated.
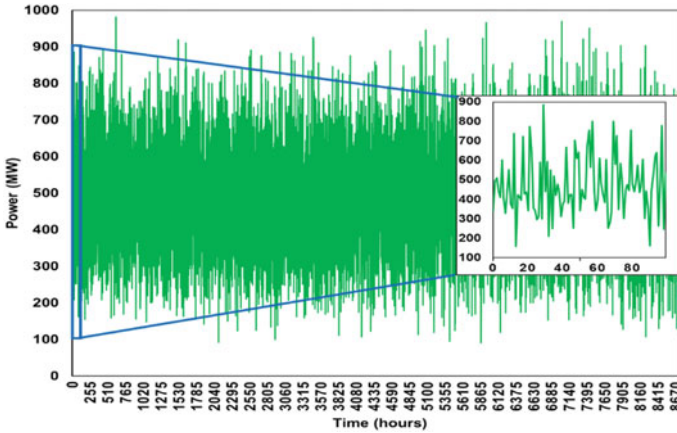
Fig. 4 Global power produced by the twelve wind farms

3 Impact of Wind Power on the Reliability of the Moroccan Electrical Grid

3.1 Non-sequential Monte Carlo Simulation

A non-sequential Monte Carlo simulation was used to evaluate the reliability indices [3, 13–19] concerning the Moroccan electrical system. This simulation depicts the "grid's life" as a series of events that alter the system's status. In our simulation, each created system state is taken into account hourly. Furthermore, depending on the studied hierarchical level, the algorithm recognizes load changes, possible failures, or maintenance of production units (hierarchical level HL1), as well as any overloads or line unavailability (weather, falling branches, etc.) in the case of hierarchical level 2 (HL2: Production + Transmission). It should be noted that the reliability evaluation in this research will be limited to the hierarchical level HLI.

Each traditional production unit (thermal, gas turbine, hydraulic) can be classified into two states in the context of our research: completely available and unavailable. Using the following approach, a uniformly distributed (in the interval [0, 1]) number is sampled for each production unit during the Monte Carlo simulation to determine its operating state [3, 6, 9, 18, 20]:

- If $\mu \leq$ FOR (Forced Outage Rate: rate of unplanned downtime associated with a production unit), the production unit is assumed to be unavailable.
- If $\mu >$ FOR, the production unit is fully available.

In practice, the global power is equal to the sum of the power produced by each conventional unit (considering the latter's availability state) plus the sampling wind power for the twelve wind farms examined during each created system state.

3.2 Reliability Indices: Well-Being Analysis of the Electrical System

The electrical system reliability indices are calculated by comparing the global generated electrical power with the load consumption during each generated state. The indices used in our research are known as electrical system "well-being" indices. Roy Billinton [6, 8, 18] established these indicators in 1999, and they are presented in terms of healthy, marginal, and risky states using the following definitions:

- **Healthy state**: The available global power, even if the most significant production unit is stopped, can cover the required load within adequate operating conditions of the electrical grid.
- **Marginal state**: The available global power is greater than the corresponding hourly load and permits to cover the latter within adequate operating conditions.
  However, if the most significant production unit is stopped, this power becomes lower than the needed load.
- **Risky state**: The available global power is directly below the needed consumption and does not permit adequately covering the latter.

During the electrical system "well-being" analysis, reliability indices are defined as follows:

$$\text{Healthy state probability} = P(H) = n(H)/N \cdot 8760 \tag{5}$$

$$\text{Marginal state probability} = P(M) = n(M)/N \cdot 8760 \tag{6}$$

$$\text{Risky state probability} = P(R) = n(R)/N \cdot 8760 \tag{7}$$

where $n(H)$, $n(M)$, and $n(R)$ are the numbers of healthy, marginal, and risky states, respectively and are simulated during the Monte Carlo algorithm, and N is the total number of simulated years.

3.3 The Moroccan Generating Facilities and Evolution of the Load

In addition to the wind parks presented in Table 1, the classical production units installed in Morocco in late 2019 are also considered. The total installed capacity was then reaching 8746 MW (without considering the installed renewable power intermittent) dispatched in 1269.1 MW of hydraulic type, 5488 MW of thermal type, 1489.4 MW of gas turbine type and 464 MW of STEP standard [21]. This study does not consider the photovoltaic and CSP power installed.

Table 4 Classical generating units considered in the Monte Carlo simulation

Total hydraulic park	1269.1 MW
Total thermal park	5488 MW
Total Gaz turbine park	1489.4 MW
STEP park	464 MW

Table 5 FOR parameter for each type of conventional production [6, 20]

Type of production unit	FOR (%)
Hydraulic	1.5
STEP	1.5
Gaz turbine	1.2
Thermal	2.5

Table 4 shows the various units that constitute the Moroccan conventional production park, while Table 5 provides the FOR associated with each type of conventional production [6, 18, 20].

Finally, Fig. 5 illustrates the temporal evolution of the simulated load. This last one is modeled by the use of a uniformly distributed law (linear approximation of the load distribution of probability) between the annual peak load and the base consumption [4–6]. Table 6 gives the peak load and base consumption in Morocco for 2019.

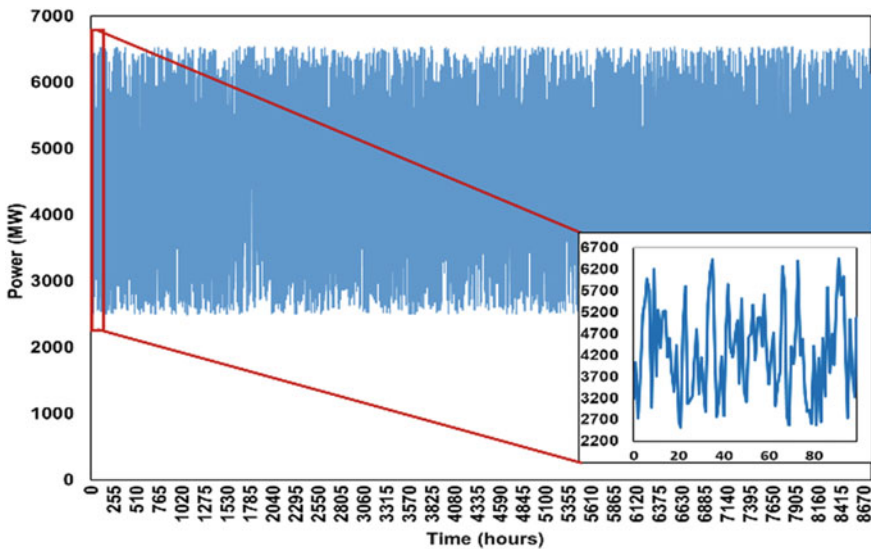$$CH = u(C_{\text{pic}} - C_{\text{base}}) + C_{\text{base}} \tag{8}$$

Fig. 5 Evolution of the simulated load

Year	Base load (MW)	Peak load (MW)
2019	2500	6540

where CH, C_{pic}, C_{base} are power load, peak load, and base load, respectively.

4 Simulation Results and Interpretations

The first part of this research is to determine the impact of wind generation on the Moroccan power grid's reliability.

A Monte Carlo simulation has been launched to assess this impact, integrating wind power (11%) to the classical production park described in Table 4 and replacing this wind power with conventional generation (keeping the same installed power).

The simulation results are provided in Table 7.

Firstly, we see from the results of Table 7. that when we replace wind power (11%) in the electrical system with conventional power (without changing the load), the probability of finding the system in a risky state decreases considerably ($P(R) = 15.8 \times 10^{-5}$ with wind energy $P(R) = 2.4 \times 10^{-5}$ with conventional energy). The results confirm that the intermittency of wind energy impacts the electrical system's reliability.

Secondly, we also note from the results of Table 7. The energy produced by the system constituted by conventional units greatly exceeds 11% of the wind power (Fig. 6).

The second part of this study evaluates the effect of adding the installed wind power to see at which global installed power we will reach the same reliability indices of 100% conventional energy (Table 8).

Based on the results of Table 8, we can first note that reliability indices decrease when the wind energy progressively increases. Furthermore, these reliability indices do not reach the conventional 100% reliability indices despite when 100% in wind

Table 7 MCS results with replacement wind power by conventional plants

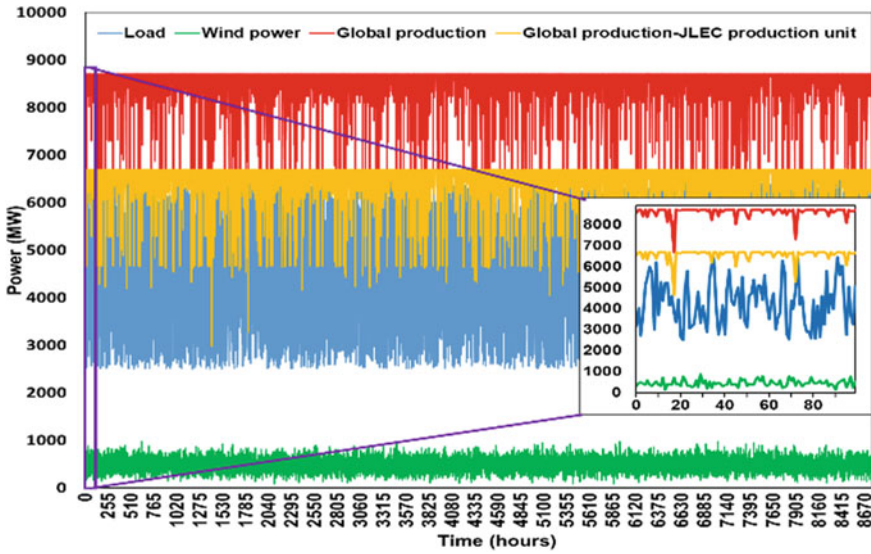|	With wind power 11% (2019)	Replacement wind power by conventional energy (2019)
Installed power (MW)	10,677	
$P(H)$	0.98455	0.99404
$P(M)$	0.015292	0.0059349
$P(R)$	15.8×10^{-5}	2.4×10^{-5}
LOLE (h/year)	1.38	0.21
Yearly electricity generation (GWh)	78,970	85,100

Fig. 6 Simulation of global production, wind power, and load in Morocco on 2019

Table 8 MCS results with adding wind power to installed power

% wind power added	+ 100%	+ 60%	+ 40%	+ 20%
Installed power	21,354	17,083	14,947.8	12,812.4
$P(H)$	0.99624	0.99541	0.99455	0.99245
$P(M)$	0.0037206	0.0045377	0.0053965	0.0074871
$P(R)$	3.6244×10^{-5}	4.7374×10^{-5}	5.061×10^{-5}	6.697×10^{-5}
LOLE (h/year)	0.3175	0.415	0.44333	0.58667
Yearly electricity generation (GWh)	115,936	101,142	93,758	86,350

power of the initially installed power is added, which confirms the effect of the intermittency on the electrical reliability system.

Secondly, in terms of energy, we see from Table 8 that the yearly electricity generation, including added wind power, increases gradually and exceeds the energy produced by 100% conventional from 20% of the wind power added hence the importance of having means of energy storage and interconnections with the other countries.

The third part of this research evaluates the impact of progressively replacing the conventional power installed on 2019 by wind energy to verify which percentage of wind power the Loss Of Load Expectation (LOLE) [13–17] exceeds the threshold set CREG and NERC as shown in Table 10.

Table 9 MCS results of replacing conventional power installed by wind power

% wind power	17%		15%		11% (2019)
Wind installed power	1815		1601.5		1220
$P(H)$	0.96611		0.97675		0.98455
$P(M)$	0.033314		0.022932		0.015292
$P(R)$	0.00057277		0.00032249		15.8×10^{-5}
LOLE (h/year)	5.0175		2.825		1.3833

Table 10 Threshold of LOLE fixed by CREG and NERC

	Threshold fixed by CREG[a] (h/year)	Threshold fixed by NERC[b] (h/year)
LOLE	2.72	2.4

[a] CREG: Commission de Régulation de l'Électricité et du Gaz
[b] NERC: North American Electric Reliability Corporation

From the result of Table 9, we can see that at 11% of wind power, LOLE is lower than the threshold fixed by CREG and NERC; however, at 15%, LOLE exceeds this threshold, which confirms that 11% of installed power is an adequate percentage of wind power on 2019.

5 Conclusion

In this paper, we firstly evaluated the Moroccan power grid's reliability in the presence of wind power for the hierarchical level HLI (HLI) using a non-sequential Monte Carlo simulation.

The simulation results have evaluated the effect of wind power generation on the reliability of the Moroccan power grid. When we replace wind power (11%) with conventional power (without changing the load), the probability of finding the system in a risky state decreases considerably, which confirms the intermittency's impact of wind energy on the electrical system's reliability.

Secondly, we note that reliability indices decrease when the wind energy is progressively increased, without reaching the conventional 100% reliability indices despite when 100% in wind power of the initially installed power is added, which confirms the effect of the intermittency on the electrical system reliability.

In terms of energy, the yearly electricity generation, including wind power, increases gradually, hence the importance of having means of energy storage and interconnections with other countries.

Finally, when we progressively replace the conventional power installed on 2019 by wind energy, we note that 11% of wind power LOLE is lower than the threshold fixed by CREG and NERC. However, at 15%, LOLE exceeds this threshold, which

confirms that 11% of installed power is an adequate percentage of wind power in 2019.

References

1. Schinke B, Klawitter J (2016) Background Paper: Country Fact Sheet Morocco Energy and Development at a glance 2016 Project: (MENA-SELECT), Germanwatch
2. International Energy Agency, Morocco 2019: Energy policies beyond IEA countries, 2019, Paris
3. Vallée F, Versèle C, Lobry J, Moiny F (2013) Non-sequential Monte Carlo simulation tool in order to minimize gaseous pollutants emissions in presence of fluctuating wind power. Renew Energy 50:317–324
4. Oukili M, Zouggar S, Vallée F, Seddik M, Ouchbel T (2013) Comparative study of the moroccan power grid reliability in presence of photovoltaic and wind generation. Smart Grid Renew Energy
5. Oukili M, Zouggar S, Vallée F, Seddik M, Ouchbel T (2010) Impact of wind power integration and load evolution on the Moroccan electrical grid reliability. IJEET 6(1)
6. Vallée F, Lobry J, Deblecker O (2008) System reliability assessment method for wind power integration. IEEE Trans Power Syst
7. Abdulwhab A, Billinton R (2004) Generating system well-being index evaluation. Electr Power Energy Syst 26:221–229
8. Billinton R, Karki R (1999) Application of Monte Carlo simulation to generating system well-being analysis. IEEE Trans Power Syst 14(3)
9. Billinton R, Sankarakrishnan A (1995) A comparison of monte carlo simulation techniques for composite power system reliability assessment. In: IEEE WESCANEX 95, communications, power, and computing. Conference proceedings
10. Billinton R, Sankarakrishnan A (1994) Adequacy assessment of composite power systems with HVDC links using Monte Carlo simulation. IEEE Trans Power Syst 9(3):1626–1633
11. Daoudi M, Ait Sidi Mou A, El Khomri M, Elkhouzai E (2019) Statistical analysis of wind speed & wind direction in Tantan Province, Morocco. Int J Recent Technol Eng (IJRTE)
12. Vallée F, Lobry J, Deblecker O (2008) Application and comparison of wind speed sampling methods for wind generation in reliability studies using non-sequential Monte Carlo simulations. Eur Trans Electr Power 18:1–14
13. Billinton R (1994) Evaluation of reliability worth in an electric power system. Reliab Eng Syst Saf 46:15–23
14. Allan RN (1994) Power system reliability assessment—a conceptual and historical review. Reliab Eng Syst Saf 46:3–13
15. Billinton R, Allan RN (1990) Basic power system reliability concepts. Reliab Eng Syst Saf 27:365–384
16. Billinton R, Allan RN (1988) Concepts of power system reliability evaluation. Electr Power Energy Syst 10(3)
17. Billinton R, Allan RN (1984) Power system reliability in perspective. IEEJ Electron Power 30:231–236
18. Vallée F, Deblecker O, Lobry J (2009) Clustering methodologies for wind power sampling in large scale network generating system. Reliab Stud J Electr Syst
19. Batut J (1987) Fiabilité du réseau électrique, vol 3, no 9. Persée, Cahier / Groupe Réseaux, pp 33–42
20. Vallée F, Lobry J, Deblecker O (2007) Impact of the wind geographical correlation level for reliability studies. IEEE Trans Power Syst 22(4)
21. Office National d'Electricité (2008) Quand l'énergie… vient du vent. Documentation ONE

Development of an Energy Management Approach in a Residential Building Integrating Renewable Energies

Abdelouahid El Youssefi, Noudi Nsangou Mama, and Abdallah Saad

Abstract Morocco has implemented an ambitious energy policy in terms of energy efficiency and integration of renewable energies into the electricity network. These distributed generators should represent more than 50% of the total installed capacity in 2030. The current paper concerns this policy and aims to integrate a maximum of renewable energies in buildings to reduce the part provided by the electricity distributor. This contribution is limited to the integration of photovoltaic (PV) energy into a residential building. The exact energy needs were evaluated by modeling different appliances load profiles. These were adjusted to the real ones obtained by experimental measurements. We then sought to cover the needs with locally produced and optimized PV. Without using storage solutions nor injecting surplus PV production into the network, we adopted an optimization approach based on load scheduling without affecting comfort. Therefore, the energy supplied by the electricity distributor was substantially reduced.

Keywords Load scheduling · Modeling load profiles · Optimization · PV energy · Residential building

1 Introduction

World energy policy has highlighted sustainable development and the limitation of greenhouse gases to overcome the depletion of fossil fuels and global warming, hence the development of research into alternative renewable energies and the energetic efficiency.

Morocco, fully involved in these environmental issues, has put in place a very ambitious energy policy which has made it a forerunner at the regional and even global level. Through the policy implemented, it aspires to meet more than 50% of its needs by 2030 through renewable energies (solar and wind). Through the strategy

A. El Youssefi (✉) · N. Nsangou Mama · A. Saad
Department of Electrical Engineering, ENSEM-HASSAN II University, Casablanca, Morocco
e-mail: abd.elyoussefi1971@gmail.com

adopted in terms of energy efficiency, it aims to achieve an energy saving of around 20% by 2030.

Our work fits into this context and aims at the energy efficiency of a building and its energy autonomy via PV energy sources. Indeed, residential and tertiary buildings in Morocco represent a significant part of the energy consumed in Morocco, more than 50% in 2015 according to the International Energy Agency.

The potential for energy savings in the building sector is estimated at around 40%, and this largely through economically viable measures according to the "450" scenario of the said agency in 2009.

This potential can be achieved by aggregating the effect of several individual measures, according to a study carried out by Plan Bleu in 2009.

Our work concerns residential buildings where we are looking for:

- On the one hand to use equipment with high energy efficiency, providing the required comfort but in an optimal way.
- On the other hand, to set up a solar PV production such that the load curve of the building where the building element considered matches the consumption profile, the unused part therefore injectable in the network having to be minimized. The storage and return of the latter are not mentioned in this article.

To do this, we first modeled the equipment of a home and studied the load profiles. The established models as well as the modeled profiles are validated by experimental measurements. The loads are then managed dynamically, in particular the times and durations of use, to arrive at a profile best covered by the PV production. Once the global PV consumption/production management model was validated, we implemented it to reduce the energy required from the network without affecting user comfort.

The reminder of this article is structured as follows. Section 2 describes and details models and data specifications. Section 3 addresses the scheduling algorithm. Section 4 investigates the case study of our work and summarizes the results obtained, and finally, Sect. 5 presents conclusions and future works.

2 Data Specifications and Models

For the purposes of our study, we considered a 100 m^2 apartment in a four-story building located in Casablanca. It is composed of a living room, a kitchen, three bedrooms and two bathrooms.

The list of equipment in the apartment is specified in Table 1. This table mentions the powers and the operating time slots.

1. **Load profile**

The modeling of buildings and their equipment has been the subject of numerous studies [1-3]. We use the models developed to analyze the energy needs of housing.

Devices	Power (W)	User time preference	Observations (priority)
Fridge-freezer	140	–	
Washing machine	2400	8:00	Schedulable (1)
Electric cooker	1000	7:00–7:15 13:00–14:00 19:00–20:30	
Dishwasher	2000	17:00	Schedulable (2)
Heating/air conditioning	1200	7:00–8:30 17:00–23:00	
Flat iron	1200	15:30–17:00	Schedulable (3)
Water kettle	1500	7:00–7:15 13:30–14:00 20:30–21:00	
Microwave	800	7:00–7:15 13:30–14:00 19:30–20:00	
Coffee maker	1000	7:00–7:15 13:45–14:00	
Water heater	1600	6:00–8:15 17:00–21:00	
Lighting	300	6:30–8:00 18:00–23:00	
Other goods	500	19:30–21:30	

Table 1 Data specifications of the appliances

Figure 2 represents the simulation (Obtained by MATLAB/Simulink) of the variation of the powers absorbed during 24 h of operation. The coefficients assigned to the models have been adjusted to have practically the curves obtained from the experimental measurements.

2. **Experimental Measurement**

The house also includes an experimental device for measuring the consumption data of the various equipment as shown in Fig. 1. This is mainly the Home Energy Manager (HEM). Its main function is to manage consumption and control of loads as well as the interface with the network manager (including weather and consumption forecasts).

A load curve of a dishwasher observed experimentally is shown in Fig. 2 and the global load of residence is shown in Fig. 3.

The proposed solution is based on an Arduino board. It consists of the following materials:

- Processing: Arduino Mega2560 board
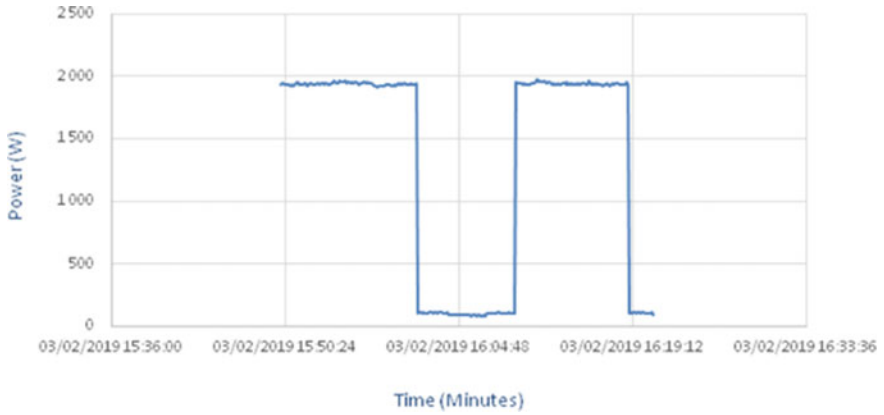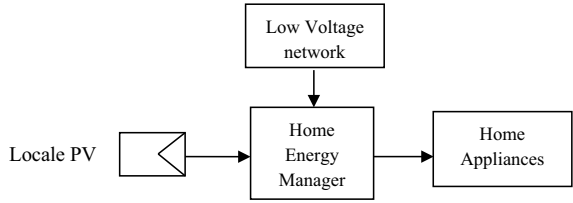
Fig. 1 Home installation
diagram

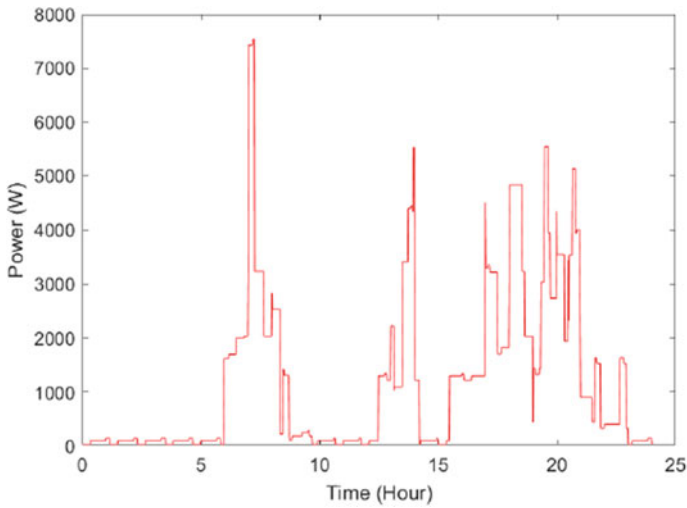**Fig. 2** Load curve of a dishwasher observed experimentally

Fig. 3 Global load profile of residence

- Display: LCD I2C
- Measurement sensors
- Conditioning card.

3. **PV profile**

For this study, the residential PV load profile was generated. Figure 5 indicates the profiles for both summer and winter periods in Casablanca city (Fig. 4).

4. **Adaptation of the load profile to the experimental results**
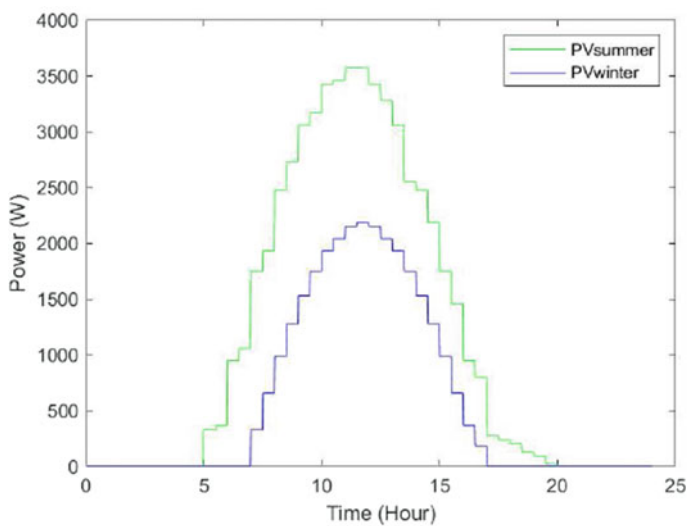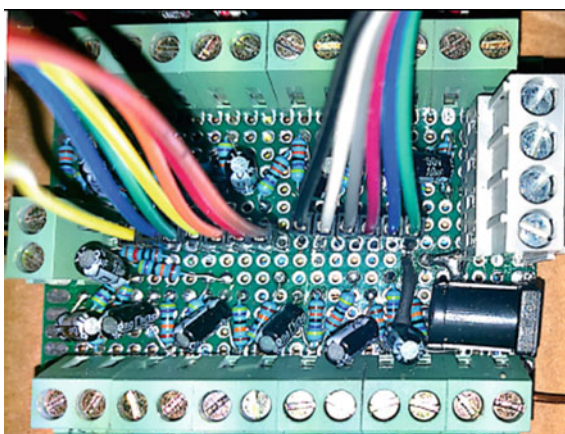

Fig. 4 Conditioning card

Fig. 5 Power of the residential PV installation

For each equipment, the theoretical consumption profile obtained by simulation is compared with the measured profile. We have determined by the method of least squares the parameters of the model which makes the two profiles coincide. According to this method, if we have N measurements (y_i) with $i = 1, \ldots, N$ the "optimal" parameters (y_i) are those which minimize the sum:

$$S(\theta) = \sum_{i=1}^{N} (y_i - f(x_i; \theta))^2 = \sum_{i=1}^{N} r_i^2(\theta) \tag{1}$$

where the $r_i(\theta)$ are the residuals of the model: $r_i(\theta)$ is the difference between the measurement y_i and the prediction $f(x_i; \theta)$ given by the model.

$S(\theta)$ can be considered as a measure of the distance between the experimental data and the theoretical model which predicts these data. The least squares prescription dictates that this distance be minimal.

3   Scheduling Algorithm

Solar energy is an intermittent source. When it is available, it should be consumed first. The objective of this paragraph is to present an optimal management method for loads to maximize the use of this renewable source.

In the literature, there are several methods for optimizing energy consumption in an electrical installation. As presented in [2], they are:

- Linear (LP) and nonlinear (NLP) optimizations
- Applied dynamic programming (DP)
- Optimal linear quadratic (LQ) control
- Genetic method, …

In general, the methods presented above have a sampling period of the order of 1 h. This is not suitable for domestic appliances which, for some, can only work for a few minutes. In addition, only the average load curve is modeled. Their application to the management of building loads is not easy. Therefore, the authors in [2] have developed an approach that combines two methods: nonlinear programming and the so-called "branch and bound" separation and evaluation method. A study of the exact method of managing the loads of a building is presented in [4].

We have developed a nonlinear optimization method based on the algorithm below.

4 Case Study

Numerical studies have been done to evaluate the performance of the algorithm presented in Fig. 6. By using MATLAB, the experiments are performed on the HP computer with an Intel Core i3 CPU, 2.30 GHz speed and 4 GB of memory.

The energy consumed by all the n equipment:

$$E_L = \sum_{i=1}^{n} p_i t_i \tag{2}$$

The energy produced by E_{PV} photovoltaic (PV) panels is a function of time. For durations of the order of a minute, it can be estimated by:

$$E_{PV} = \sum_{i=1}^{n} P_{PV} \cdot t_i \tag{3}$$

P_{PV}: Power produced by the PV and which varies according to time.

In our case study, E_{PV} is known from meteorological data. It is then necessary to optimize the load curve of the next day to consume the maximum possible energy produced by the PV: Therefore, minimizing the energy produced by the PV and not consumed.

As described in [5], the objective function is:

Fig. 6 Organizational chart of scheduling algorithm
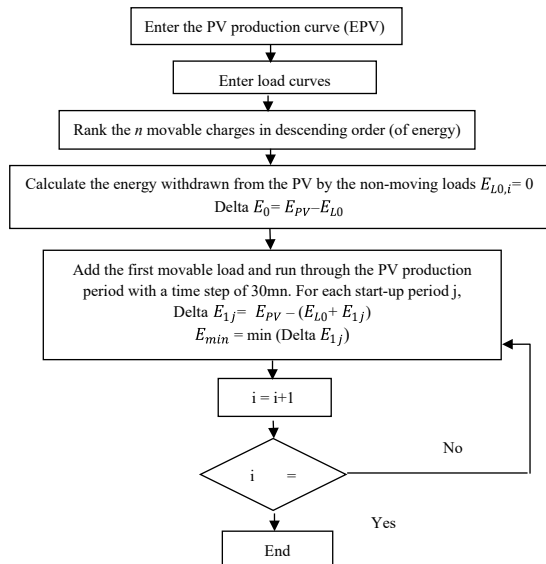
$$\text{Min } f(c) = E_{\text{PV}} - E_L(c) = \sum_{i=1}^{n} \text{Ppv} \cdot t_i - \sum_{i=1}^{n} cp_i t_i \qquad (4)$$

$$\text{Constraints}: \quad 0 \le P_{\text{PV}}.$$

Without optimization, the energies produced by the PV and not consumed are given in the table below:

The energy paid to the E_{Dist} distributor is:

$$E_{\text{Dist}} = E_L - E_{\text{PVC}} \qquad (5)$$

$$E_{\text{PVC}} = E_{\text{PV}} - E_{\text{PVNC}} \qquad (6)$$

With

E_{PVC}: PV energy produced and consumed.

E_{PVNC}: PV energy produced and not consumed.

The energy paid to the distributor is reduced by 9.64 kWh in summer and 4.80 kWh in winter. That is a decrease of 34% and 17% of the energy paid to the distributor in summer and winter, respectively, and this, by the sole fact of using a PV source and without optimizing the load curve.

Using the proposed optimization algorithm, we get (Tables 2, 3 and 4):

Table 2 PV energy without optimization

Period	Average energy consumed per day E_L (kWh)	PV energy produced E_{PV} (kWh)	PV energy produced and not consumed E_{PVNC} (kWh)
Summer	28.25	27.35	17.71
Winter	28.25	13.84	9.04

Table 3 PV energy with optimization

Period	Average energy consumed per day, E_L (kWh)	PV energy produced, E_{PV} (kWh)	PV energy produced and not consumed, E_{PVNC} (kWh)
Summer	28.25	27.35	15.56
Winter	28.25	13.84	6.11

Table 4 Optimal operating periods of movable loads

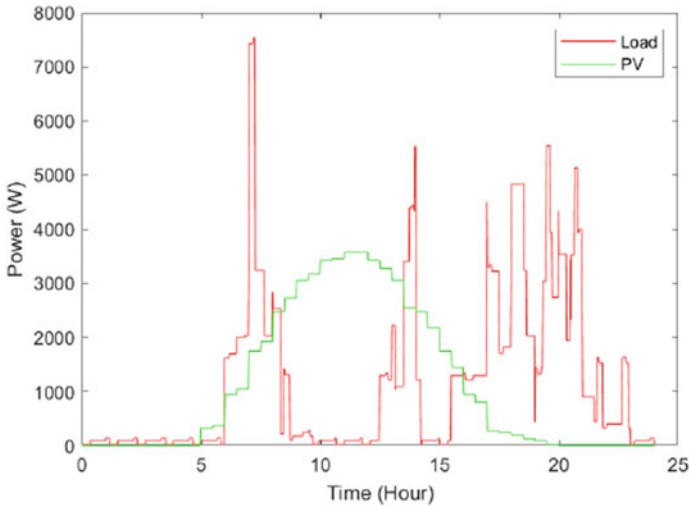| Period	Optimal time of use		
	Iron	Dishwasher	Washing machine
Summer	11:30	10:30	9:30
Winter	10:00	10:30	12:00

Fig. 7 Load and PV curves during summer period without optimization

With the optimization, the PV energy not consumed in summer is reduced by 2.15 kWh. That is a decrease of 12%. This value increases to 2.93 kWh (32%) in winter. The energy paid to the distributor is reduced by 11.79 kWh (42%) in summer and 7.73 kWh (27%) in winter. This contributes significantly to the fight against pollution caused by fossil energy sources.

The algorithm used for the optimization is an exact method whose implementation time is about 17 min.

Figures 7 and 8 represent the different load and production profiles in summer and winter without and with optimization, respectively.

5   Conclusion

In this article, we have evaluated the impact of load management on the electrical energy consumption of an apartment in the city of Casablanca. After the experimental consumption measurements, we studied the load models of the different electrical equipment. These load profiles have been adjusted to actual values. This made it possible to have the overall load curve of the apartment. Then, we developed a method for managing movable loads. This gave very interesting results since we were able to reduce the consumption paid to the distributor by 42% in summer and 27% in winter. Similarly, the share of PV energy produced and not consumed decreased by 12% and 32% respectively in summer and winter.

Moreover, the proposed algorithm is classified among the exact methods. The implementation time is 17 min for 12 pieces of equipment, 3 of which are movable.
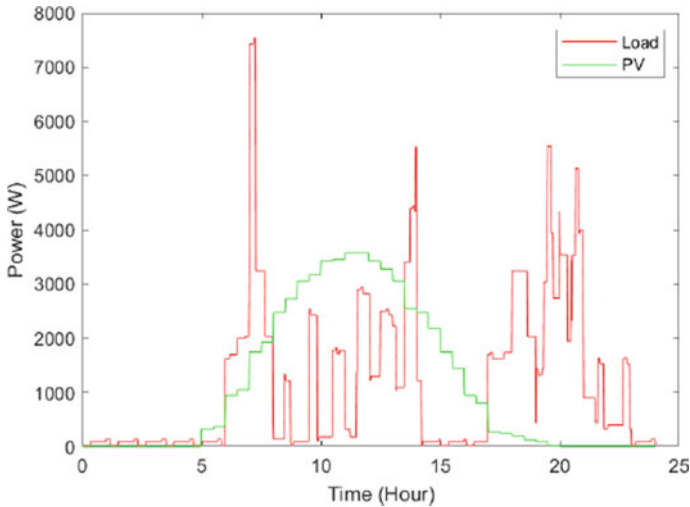
Fig. 8 Load and PV curves during summer period with optimization

This time can become very long if the number of controllable loads increases. As indicated by the authors in [4], this time can be significantly reduced by using a heuristic method: hence, the need to make a comparative study with a heuristic method. Similarly, it is possible to consider integrating local storage of PV energy to reduce the share of PV energy produced and not consumed.

References

1. Gueddouch T (2017) Study to optimize and improve the energy performance of residential buildings. Ph.D. dissertation, ENSEM, Univ. Hassan II, Casablanca
2. Le K (2008) Optimal management of energy consumption in the buildings. Ph.D. dissertation, Univ. Grenoble and Univ. Danang
3. Abarkan M (2014) Modelization and analysis of the behavior of a building integrating multi energy systems. Ph. D. dissertation, Univ. Aix Marseille and Univ. Sidi Mohamed Ben Abdellah, Fes
4. Ogwumike C, Short M, Denai M (2015) Near-optimal scheduling of residential smart home appliances using a heuristic approach. In: Proceedings of the 2015 IEEE conference on industrial technology, Seville, Spain, 17–19 March 2015, pp 3128–3133
5. Paris B, Blanco E, Hazyuk I (2010) Programmation linéaire pour la gestion de l'énergie électrique d'un bâtiment. In: Conférence IBPSA France, 2010

Numerical Analysis of Heat Transfer in a Solar Collector Submitted the Flow of Nanofluid

Fatima-Zahra Barhdadi, Ikrame Jamal, Kamal Amghar, and Salah Daoudi

Abstract This work is based on a two-dimensional numerical simulation, to study the effects of various nanofluids (Al_2O_3, C_uO, S_iO_2, and Z_nO) on heat transfer performance along the rectangular channel mounted with two different ribbed: rectangular and triangular. The effects of this ribbed on heat transfer performance are evaluated with Reynolds number in the range from 5000 to 20,000 with a uniform heat flux of 10,000 W/m^2 at upper and lower walls. The governing equation is based predicated on the continuity equation, momentum, and energy equations, which are solved by using the finite volume method. Numerical results are analyzed, such as the profiles of velocity, the factor coefficient, and average Nusselt number. In addition, the model studied is tested by comparing the literature and the results obtained that are in favorable agreement with those reported in literature. Moreover, the results show that the thermal performance is highly significant for the case of ribbed rectangular also with nanofluid S_iO_2–water in which shows the greatest improvement in heat transfer compared to other nanofluids. Finally, through ribbed channels, the average Nusselt number improved with increasing Reynolds number.

Keywords Forced convection · Turbulent flow · Ribbed channel · Nanofluid · Heat transfer · Numerical simulation

F.-Z. Barhdadi (✉) · I. Jamal · K. Amghar · S. Daoudi
Laboratory of Mechanics and Energy (LME), Faculty of Sciences, Mohammed First University, 60000 Oujda, Morocco
e-mail: Fatima-zahra.barhdadi@ump.ac.ma

K. Amghar
e-mail: k.amghar@uae.ac.ma

K. Amghar
Department LSIA, National School of Applied Sciences, Abdelmalek Essaâdi University, Al Hoceima, Morocco

1   Introduction

Recently, the improvement of forced convection heat transfer in a rib–groove channel has attracted the attention of many engineering researches, as it contributes to the increase of the efficiency of heat exchangers. The most geometric shapes of the rib–groove channel used in all domains of engineering are the following: rectangular, triangular, square, and circular [1]. The use of nanofluids on the heat transfer fluids in rib–groove channels is one of the most widely used methods to improve heat transfer [2–5]. These are solutions containing nanoparticles with very high thermal conductivity suspended in a base fluid to increase the effective thermal conductivity of the mixture, thus improving its thermal performance. Arroub et al. [6] conducted a numerical study which is performed to investigate the influence of Al_2O_3–water nanofluid on the mixed convection heat transfer in a rectangular channel. The obtained results indicate that the presence of nanoparticles in the base fluid acts to increase the heat transfer rate and the average temperature inside the channel. Monireh et al. [7] studied numerically the heat transfer by two-dimensional incompressible nanofluids around four sinusoidal lateral obstacles in a horizontal channel. The results obtained show that the average Nusselt number increases with the increase in the concentration of nanoscale particles and the decrease in the distance between the obstacles, which improves thermal transmission. Parsazadeh et al. [8] presented a numerical investigation of forced convection turbulent of nanofluid in a two-dimensional channel, and the results indicate that S_iO_2 nanofluid improves the heat transfer in a channel with detached rib arrays due to its higher thermal conductivity and lower mass density. It is shown that heat transfer increases with increasing nanoparticle concentration and decreasing nanoparticle diameter. Vanaki and Mohammed [9] studied a numerical work to investigate the thermal and hydraulic characteristics of turbulent flow of nanofluid by forced convection in channels using transverse ribs of different shapes. According to the results, the triangular ribs have the highest Nu number. It was found that the S_iO_2 nanofluid with a nanoparticle diameter of 30 nm and a concentration of 4% provides the highest values of Nusselt number and friction factor. Alsabery et al. [10] numerically investigated the effect of the Reynolds number and wall corrugation number on the flow and heat transfer of nanofluids in a two-dimensional channel and wavy. The results indicate that when the Reynolds number increases, the average particle velocity increases, on the other hand, the temperature is lower than the average temperature. In addition, as the number of oscillations increases, the fluid leaves the channel at a higher average temperature.

2   Physical Model

According to Fig. 1, the configuration represents a rectangular channel mounted with the ribbed rectangular and triangular which are installed on the walls of the channel. The study is analyzed at three configurations (a), (b), and (c). Consequently, the
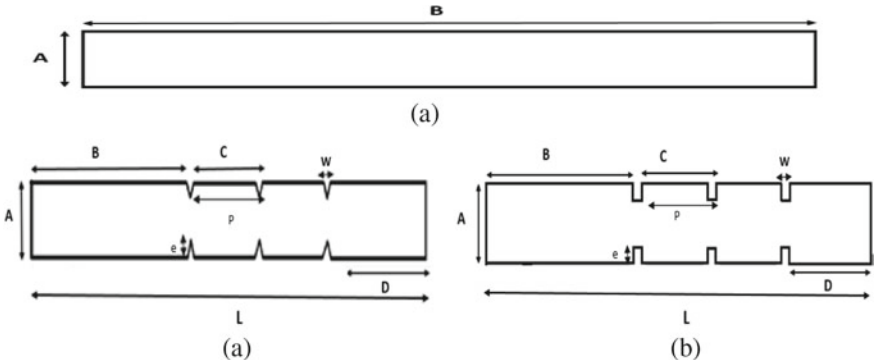
Fig. 1 Studied configuration, **a** smooth channel, **b** rectangular ribbed channel, **c** triangular ribbed channel

Table 1 Dimensions of the configuration

L	A	B	e	W	P	D
104 mm	10 mm	40 mm	2 mm	2 mm	20 mm	22 mm

numerical simulation is carried out to study the effect of various nanofluids and the effects of configuration on heat transfer performance. The dimensions of geometries are illustrated in Table 1.

The flow of nanofluid is assumed to be turbulent with intensity of 1% and at room temperature of $T_{in} = 300$ K. The top and bottom walls are maintained at a uniform heat flux of 10,000 W/m². There is no slip condition imposed on the channel walls and rib surface in all simulation.

3 Mathematical Formulation

3.1 Governing Equations

The physical problem is examined by two-dimensional form of the Navier–Stokes equations. The system of equation is solved numerically by finite volume method. The standard k–ε model is chosen for modeling the turbulence. For a Cartesian coordinate system, these equations can be written as follows:

Continuity equation:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{1}$$

Momentum equation:

$$\frac{\partial(\rho u_i)}{\partial t} + \frac{\partial(\rho u_i u_j)}{\partial x_j} = -\frac{\partial P}{\partial x_i} + \left[\mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)\right] + F_j \tag{2}$$

Energy equation:

$$u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = \frac{K_{nf}}{(\rho c_\rho)_{nf}}\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right) \tag{3}$$

The standard k–ε model is a semiempirical model in which two transport equations are used, one for the turbulence kinetic energy k and rate of kinetic energy dissipation. The equations demonstrating the turbulent flow can be formed as follows:

Equation of the kinetic energy K:

$$\frac{\partial}{\partial x_j}(\rho u_j K) = \frac{\partial}{\partial x_i}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial K}{\partial x_j}\right] + G_k - \rho\varepsilon \tag{4}$$

Turbulent kinetic energy dissipation:

$$\frac{\partial}{\partial x_i}(\rho u_i \varepsilon) = \frac{\partial}{\partial x_i}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial\varepsilon}{\partial x_j}\right] + C_{1\varepsilon}\left(\frac{\varepsilon}{k}\right)G_k + C_{2\varepsilon}\rho\left(\frac{\varepsilon^2}{k}\right) \tag{5}$$

3.2 Governing Parameters

$$\text{Reynolds number (Re)} : \ R_e = \frac{\rho u D_h}{\mu} \tag{6}$$

$$\text{Local Nusselt number (Nu)} : \ N_u = \frac{h D_h}{K_{nf}} \tag{7}$$

$$\text{Friction factor } (f) : \ f = \frac{2\Delta P D_h}{L\rho U^2} \tag{8}$$

3.3 Thermophysical Properties of Nanofluids

The following equations are used to calculate the thermophysical properties of nanofluids, which are the density and heat capacity [11], according to these expressions:

$$\rho_{nf} = (1 - \varphi)\rho_f + \varphi\rho_p \tag{9}$$

$$(\rho C_P)_{nf} = (1 - \varphi)(\rho C_P)_f + \varphi(\rho C_p)_P \tag{10}$$

In addition, the parameter of thermal conductivity is calculated by using Brownian motion of nanoparticles in rib–groove channel, and the effective thermal conductivity can be obtained by using the following mean empirical correlation [12]:

$$k_{\text{eff}} = k_{\text{static}} + k_{\text{Brownian}} \tag{11}$$

$$K_{\text{static}} = \frac{K_{np} + 2K_f + 2(K_{np} - K_f)\varphi}{K_{np} + 2K_f + (K_{np} + K_f)\varphi} K_f \tag{12}$$

$$K_{\text{Browniam}} = 5 \times 10^4 \beta\varphi\rho_f C_{\rho,f} \sqrt{\frac{KT}{2\rho_{np}R_{np}}} f(T, \varphi) \tag{13}$$

where $k = 1.3807 \times 10^{-23}$ J/K.

$$f(T, \varphi) = \left(2.8217 \times 10^{-2}\varphi + 3.917 * 10^{-3}\right)\left(\frac{T}{T_0}\right)$$
$$+ \left(-3.0669 \times 10^{-2} - 3.3.91123 \times 10^{-3}\right)$$

On the other hand, the dynamic viscosity [13] is calculated as follows:

$$\mu_{\text{eff}} = \mu_f \frac{1}{\left(1 - 34.87\left(\frac{d_p}{d_f}\right)^{-0.3} * \varphi^{1.03}\right)} \tag{14}$$

where $d_f = \left[\frac{6M}{N\pi\rho_{bf}}\right]^{\frac{1}{3}}$.

4　Numerical Procedures and Code Validation

The code validation is based to study the independence of the mesh. The study is applied for different meshes with that based on different numbers of nodes such us 13,626; 46,308; 61,218; 86,751; 173,451; and 231,267 nodes. Figure 2 shows the variation of the average Nusselt number as a function of the numbers of nodes for Al_2O_3–water nanofluid. In all simulations, we have chosen the mesh of 173,451 nodes, and the latter gave an ease of convergence and an acceptable result with minimal computation time. In addition, the validation of the code was done based on the work used from Vanaki et al. [9], as well as the Dittus–Boelter equation and

correlations of Blasius [14]. According to Fig. 2, the results show a good agreement
(Fig. 3).

$$\text{Correlations of Dittus}-\text{Boelter}: \quad N_u = 0.023 R_e^{0.8} P_r^{0.4} \tag{15}$$

$$\text{Correlations of Blasius}: \quad f = 0.316 R_e^{-0.25} \quad \text{for } 3000 \leq \text{Re} \leq 20{,}000 \tag{16}$$
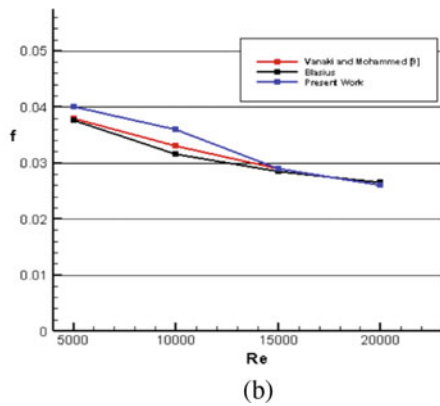
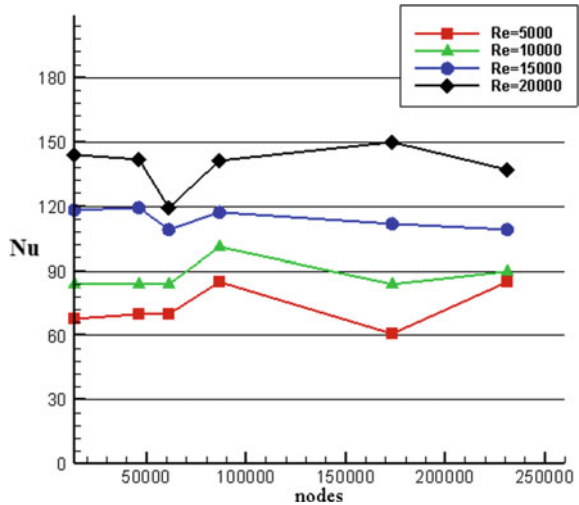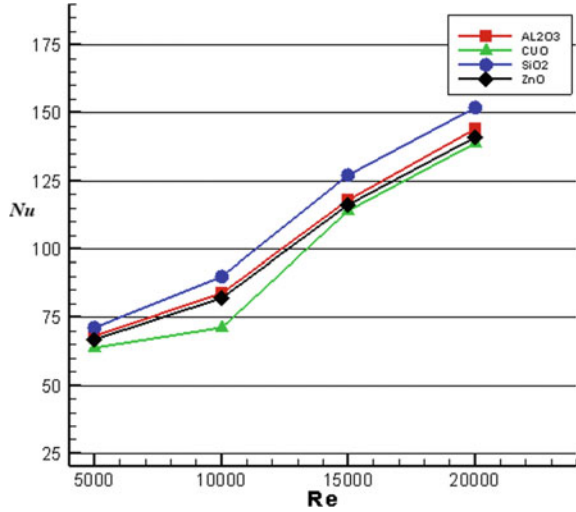Fig. 2 Mesh independence
study

(a)

(b)
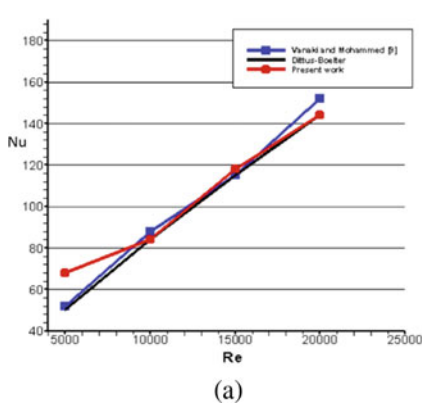
Fig. 3 Comparison of the results obtained with those numerical methods, **a** average Nusselt number
and **b** friction factor

Fig. 4 Effect of nanofluids on the average Nusselt number

5 Results and Discussion

5.1 The Effect of Various Nanofluids

In Fig. 4, we examined the effect of different types of nanofluids on the heat transfer performance along the smooth channel, and the results found that SiO_2 nanofluid gives high values of average Nusselt number compared with other nanofluids. This is explained by the S_iO_2–water nanofluid which has a particle diameter of 30 nm and a concentration of 4%. So, we will use the S_iO_2–water nanofluid to study in all simulation, which next goals is to study the effect of two different ribs of types rectangular and triangular on heat transfer performance and conservation of mass.

5.2 Velocity Profiles

Figure 5 shows the results of the velocity field distribution along the channel with triangular ribs using S_iO_2–water. We note that the velocity profiles are uniform at the entrance of the channel, while they become zero on all the walls because of the non-slip condition applied to the walls, and the value of the velocity increases while moving toward the center of the channel precisely for the ribbed part. We can explain these results by the friction between the fluid particles and the channel walls.
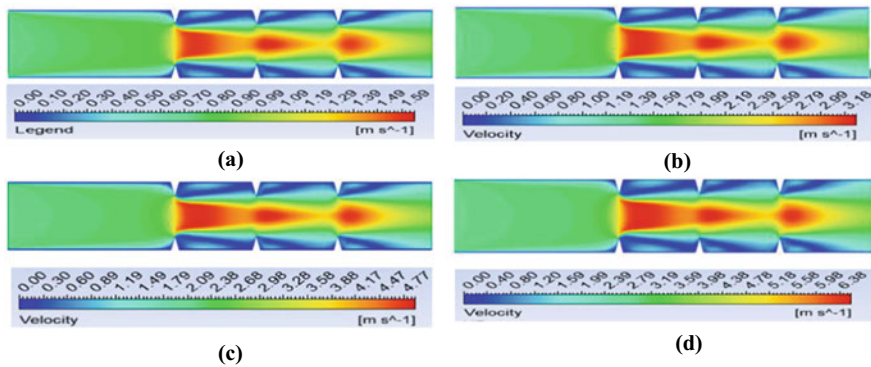
(a)

(b)

(c)

(d)

Fig. 5 Velocity distribution along the channel with ribbed triangular. **a** Re = 5000, **b** Re = 10,000, **c** Re = 15,000, and **d** Re = 20,000

5.3 *Average Nusselt Number and Friction Factor*

According to Fig. 6, the average Nusselt number increases as a function of the Reynolds number. This augmentation is attributed to the higher velocity, which can lead to the disturbance of the flow and thus improve the heat transfer. It was discovered that the average Nusselt number of a ribbed channel is greater than that of a smooth channel. This is due to the induction of a high recirculation flow and a thin boundary layer in the ribbed channels, leading to higher temperature gradients. The triangular ribbed channel provides the highest average Nusselt number. In addition, Fig. 7 shows the variation of the friction factor along the channel as a function of the Reynolds number for the smooth channel and the two cases. From this figure, it is observed that the highest values of the friction factor are related to the ribbed channel. Moreover, the Reynolds number increases to improve the friction factor. We conclude that the friction factor values for the triangular rib are the highest compared to the other configurations.

6 Conclusion

In this work, a numerical study was carried out to investigate the thermal and hydraulic characteristics of nanofluid flow by forced convection turbulent in a two-dimensional channel using rectangular and triangular ribs under a constant heat flow and Reynolds number varied from 5000 to 20,000. According to the results, it was observed that the increased value of Re causes an increase in heat transfer along the channel. Moreover, the SiO_2–water nanofluid considered perfect, which has a particle diameter of 30 nm and a concentration of 4%, provides the highest values of Nu number and friction factor among other types of nanofluids due to its higher thermal conductivity and lower mass density. On the other hand, the channel mounted by triangular ribs has the

Fig. 6 The effect of ribs on the average Nusselt number as a function of Re
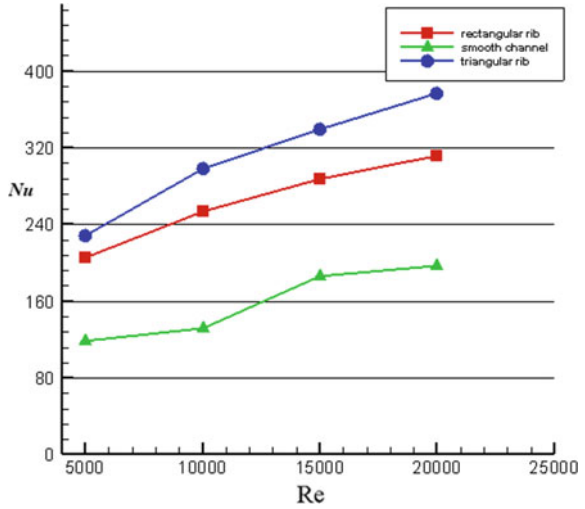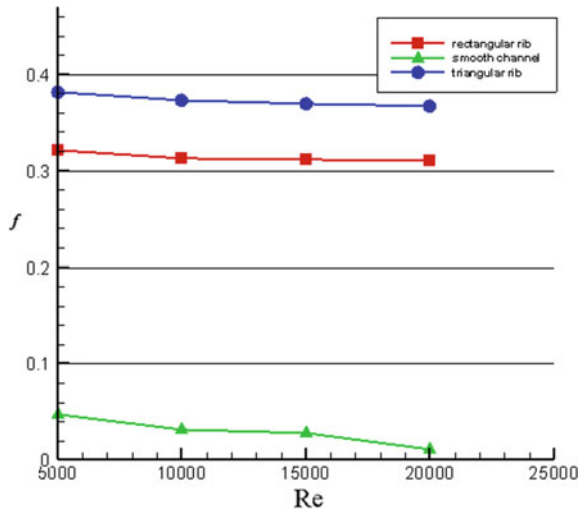

Fig. 7 The effect of ribs on the friction factor as a function of Re

important average Nusselt number. Finally, it is concluded that the type of nanofluid and the shape of ribs are considered important to improving heat transfer.

References

1. Manca O, Nardini S, Ricci D (2012) A numerical study of nanofluid forced convection in ribbed channels. Appl Therm Eng 37:280–292; Author F, Author S (2016) Title of a proceedings paper. In: Editor F, Editor S (eds) CONFERENCE 2016. LNCS, vol 9999. Springer, Heidelberg, pp

1–13

2. Salman B, Mohammed H, Munisamy K, Kherbeet A (2013) Characteristics of heat transfer and fluid flow in micro-tube and micro-channel using conventional fluids and nanofluids. Renew Sustain Energy Rev 28:848–880; Author F (2010) Contribution title. In: 9th International proceedings on proceedings. Publisher, Location, pp 1–2

3. Shanbedi M, Heris S, Amiri A, Baniadam M (2014) Improvement in heat transfer of a two-phased closed thermosyphon using silver-decorated MWCNT/water. J Dispers Sci Technol 35(8):1086–1096

4. Rasul M, Rashidi M, Mohsen I, Nor Azwadi C, Hong W (2018) Forced convection of nanofluids in an extended surfaces channel using lattice Boltzmann method. Int J Heat Mass Transfer 117:1291–1303

5. Mohammed H, Bhaskaran G, Shuaib N, Saidur R (2011) Heat transfer and fluid flow characteristics in microchannels heat exchanger using nanofluids: a review. Renew Sustain Energy Rev 15:1502–1512

6. Arroub I, Bahlaoui A, Raji A, Hasnaoui M, Naïmi M (2018) Cooling enhancement by nanofluid mixed convection inside a horizontal vented cavity submitted to sinusoidal heating. Eng Comput 35(4):1747–1773

7. Monireh A, Omid S, Rasul M, Mikhail A (2019) Enhancement of heat transfer of nanofluids in the presence of sinusoidal side obstacles between two parallel plates through the lattice Boltzmann method. Int J Mech Sci 156:159–169

8. Parsazadeh M, Mohammed H, Fathinia F (2017) Influence of nanofluid on turbulent forced convective flow in a channel with detached rib-arrays. Int Commun Heat Mass Transfer 46:97–105

9. Vanaki Sh, Mohammed H (2015) Numerical study of nanofluid forced convection flow in channels using different shaped transverse ribs. Int Commun Heat Mass Transfer

10. Alsabery A, Hajjar A, Sheremet M, Ghalambaz M, Hashim I (2021) Impact of particles tracking model of nanofluid on forced convection heat transfer within a wavy horizontal channel. Int Commun Heat Mass Transfer 122:105176

11. Williams W, Hu L, Buongiorno J (2008) Experimental investigation of turbulent convective heat transfer and pressure loss of alumina/water and zirconia/water nanoparticle colloids (nanofluids) in horizontal tubes. J Heat Transfer 130(4)

12. Vajjha R, Das D (2009) Experimental determination of thermal conductivity of three nanofluids and development of new correlations. Int J Heat Mass Transfer 52(21):4675–4682

13. Corcione M (2011) Empirical correlating equations for predicting the effective thermal conductivity and dynamic viscosity of nanofluids. Energy Convers Manag 52(1):789–793

14. Incropera F, DeWitt D (1990) Introduction to heat transfer. Wiley

Numerical Investigation on Ejector Optimization and Performance Using the Refrigerant R134a

Ikrame Jamal, Fatima Zahra Barhdadi, Kamal Amghar, and Salah Daoudi

Abstract In this paper, the computational fluid dynamics (CFD) technique is used to study the effects of condensing pressure and geometrical parameters on the performance of an ejector. The working fluid used in this study is R134a gas, which was employed in a numerical simulation of a supersonic refrigerant ejector. The numerical calculation is performed using Ansys Fluent software. The Soave–Redlich–Kwong (SRK) real gas model is used for the equation of state. The Navier–Stokes equations and the turbulence model based on the realizable $k - \varepsilon$ model are solved by the finite volume method using the SIMPLE algorithm for the pressure–velocity coupling scheme. Close agreement in terms of entrainment ratio is obtained between the numerical and experimental results, with a relative error that does not exceed 5.12% compared to the experimental data. Furthermore, the numerical results showed that the ejector performance is strongly influenced by the condensing pressure and geometrical parameters, in addition to the existence of an ideal H_{NE} and NXP value for the ejector to work at its best.

Keywords Numerical simulation · Refrigeration · Entrainment ratio · Supersonic ejector · R134a

1 Introduction

Most traditional air conditioning and refrigeration systems run on electricity. Researchers have considered replacing these traditional systems with systems that use renewable or low-grade heat sources. Thermally driven cooling systems, such as

I. Jamal (✉) · F. Z. Barhdadi · S. Daoudi
Laboratory of Mechanics and Energy (LME), Faculty of Sciences, Mohammed First University, 60000 Oujda, Morocco
e-mail: ikrame.jamal@ump.ac.ma

K. Amghar
Department LSIA, National School of Applied Sciences, Abdelmalek Essaâdi University, BP 03, Ajdir-Al Hoceima, Morocco
e-mail: k.amghar@uae.ac.ma

absorption cooling or ejector refrigeration systems, appear to be suitable substitutes for conventional refrigeration systems [1]. The main advantage of ejector refrigeration systems over absorption cooling systems is that they operate with relatively low-quality thermal energy sources up to 80 °C [2].

The supersonic ejector (Fig. 1) is a system with several advantages, the main of which is the absence of moving mechanical parts, the engineering is very simple, and the ability to operate at very low temperatures and pressure ranges. This system is used in a wide range of applications, including refrigeration [3–5], air conditioning, heating, and desalination. It works by drawing vapors or gases from low-pressure areas to high-pressure areas, transforming the pressure energy of a primary fluid into kinetic energy to drive the sucking fluid, and then compressing the mixed fluid by converting the kinetic energy back into pressure energy. In ejector refrigeration systems, compression is possible without using mechanical energy, they can be powered by thermal energy from solar energy or thermal waste from industrial processes [6].

The most important component of ejector refrigeration systems is the ejector [3]. Because of the numerous physical processes involved, the flow regime in this device is quite complicated (the interplay of supersonic and subsonic flows, mixing, shocks, etc.). As a result, several theoretical, experimental, and numerical investigations have been carried out in order to better understand the features of the local flow inside the ejector as well as its performance [7–9].

Computational fluid dynamics (CFD) has become the most effective tool to study the flow behavior inside the ejector with speed and lower costs than those generated by experimental methods. It offers agreement with the experimental results collected by several researchers. For example, CFD mechanics, according to Hanafi et al. [10], is an effective tool for determining the hydrodynamic and thermal features of the ejector under various operating conditions. Furthermore, they discovered that the primary vapor pressure has an optimal value that results in the highest entrainment ratio. Baek et al. [11] conducted a study of the parameters of geometric variables on the entrainment performance of the two-phase high-velocity ejector using R134a liquid as a refrigerant. They discovered that the size of the recirculation bubble that
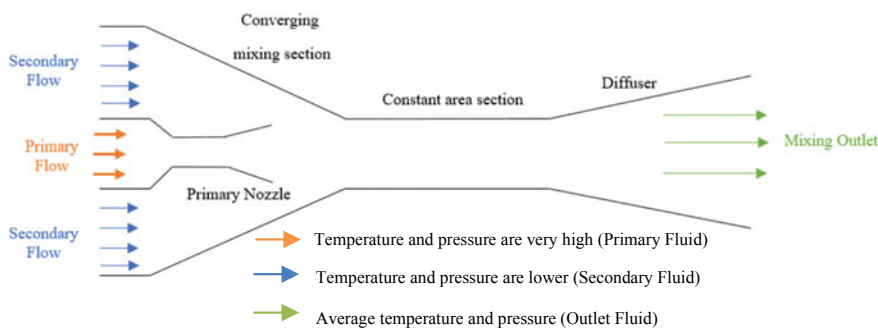

Fig. 1 2D schematic of a supersonic ejector

forms in the suction chamber is greatly influenced by the nozzle output position. They also demonstrated that the entrainment performance of the ejector is highly dependent on the diameter and length of the mixing tube. Croquer et al. [12] used the refrigerant R134a to study the type of single-phase supersonic ejector. The results show that the inlet geometry of the secondary fluid has little influence on the entrainment ratio of the secondary and primary fluids as well as on the shock position found in the mixing zone. The performance of the ejector is determined by the use of several factors, such as the entrainment ratio, ω, which is the most common index used for the evaluation of ejector performance. It is defined as the ratio of the secondary flow mass flow rate to the primary flow mass flow rate, it is given by the relationship: $\omega = \frac{\dot{m}_s}{\dot{m}_p}$.

The goal of this study is to create a numerical simulation of supersonic flow in an ejector of a refrigeration system utilizing R134a gas as the refrigerant. The findings were compared to previous research, including a numerical case study [13] and an experimental case study [14]. The findings are found to be in reasonable agreement with previous research. The effect of condensation pressure and geometrical parameters on our ejector's performance is then investigated.

2 Geometric Model

The ejector system under consideration includes all of the elements that are often found in a refrigeration system, such as a generator, a condenser, an evaporator, an expansion valve, and a pump. In our study, we used an ejector with two inlets (an inlet of the primary fluid and an inlet of the secondary fluid) and an outlet for the outlet fluid. Figure 2 shows a presentation of the ejector studied, highlighting its main dimensions. This ejector is designed for a refrigeration system that was tested in experimental measurements in the study [14].

In all numerical simulations, the boundary conditions are specified for the primary inlet pressure with a value of 2633 kPa, a secondary pressure with a value of 350 kPa, and a pressure of 720 kPa for the diffuser outlet. The temperature at the primary inlet was 100 °C, and the secondary flow temperature was 25 °C. In all CFD calculations, these values are assumed to be fixed (Table 1).
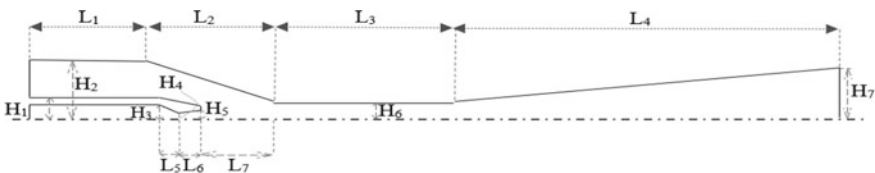

Fig. 2 Geometry of the ejector studied

Table 1 Dimensions of the ejector used in our study

L_1	L_2	L_3	L_4	L_5	L_6	L_7 (NXP)
29.79 mm	86.00 mm	89.00 mm	315.00 mm	17.20 mm	7.75 mm	29.00 mm
H_1	H_2	H_3	H_4 (H_{TH})	H_5 (H_{NE})	H_6	H_7
12.70 mm	30.00 mm	9.00 mm	3.50 mm	4.50 mm	8.31 mm	25.20 mm

3 Numerical Model

In this study, the fluid flow in the ejector was simulated using Ansys Fluent software. This program allows the prediction of fluid flows by solving the Navier–Stokes equations based on the finite volume method (FVM). The fluid used is the refrigerant R134a. With a two-dimensional, axisymmetric, and stationary computation, the working fluid flow is compressible. The SIMPLE algorithm is used in conjunction with the PRESTO! scheme for pressure–velocity coupling. The "second-order upwind" scheme is applied to discretize the advection and diffusion terms, except for the pressure term. The Soave–Redlich–Kwong (SRK) equation of state is used to calculate the supersonic flow in the ejector. The realizable k–ε model is chosen for the turbulence modeling. The solution is deemed convergent when the residuals of all algebraic equations are smaller than 10^{-6}.

4 Mesh Study

We were able to evaluate the mesh solutions independence for the four meshes (2542, 6837, 17,458, and 37,502 cells) using the RSK model of the real gas equation of state and the given operating conditions. To capture the boundary layer, these meshes are designed with a constriction at the walls. To detect the effect of the number of cells, the static pressure along the ejectors is used. By examining the results of Fig. 4, we notice that the third and fourth grids have relatively close values in terms of entrainment ratio compared to the first and second grids. As a result, the 17,458-cell array will be used for the remainder of the study (Fig. 3).
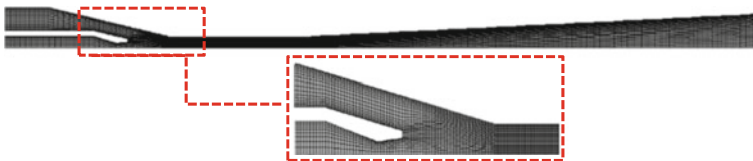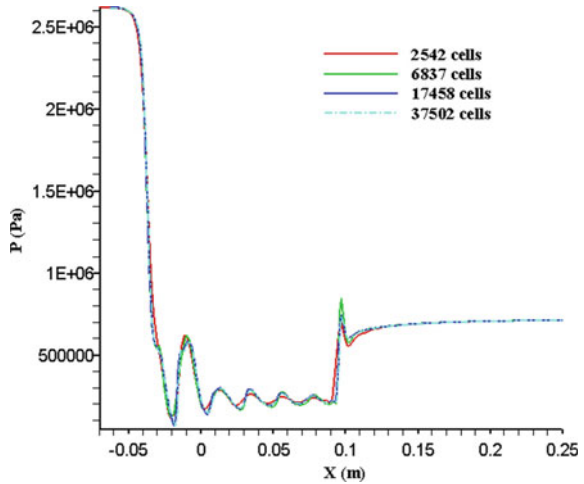

Fig. 3 Structured mesh of the ejector

**Fig. 4** Variation of the static pressure for different meshes

5 Results and Discussion

5.1 Validation

Figure 5 illustrates the variation of the calculated training ratio compared to the numerical results and experimental work depending on outlet pressure, also called condenser pressure. From this figure, there is an accordance between our simulation results and the corresponding numerical and experimental results of our work. However, one can clearly find that the current numerical results and those of Aligolzadeh and Hakkaki-Fard [13] are close, with a relative error value of 2.68%, while the predicted drive ratios in the design mode of the numerical results of Aligolzadeh and Hakkaki-Fard [13] are about 2.85%. It is concluded that the experimental results [14] are very close compared to our study.

According to Fig. 5, the results are composed of three zones. The first one is represented by a horizontal segment where the entrainment ratio is at its maximum, constant, and independent of the ejector outlet pressure (on-design regimes). This situation persists until the dividing point between the first and second regions is reached, which is called the critical point. The second region is represented by an inclined segment in which the drive ratio decreases with increasing condenser pressure (off-design regimes). In the third zone, the entrainment ratio is negative (the entrained flow is reversed).

Fig. 5 Comparison between numerical and experimental entrainment rates
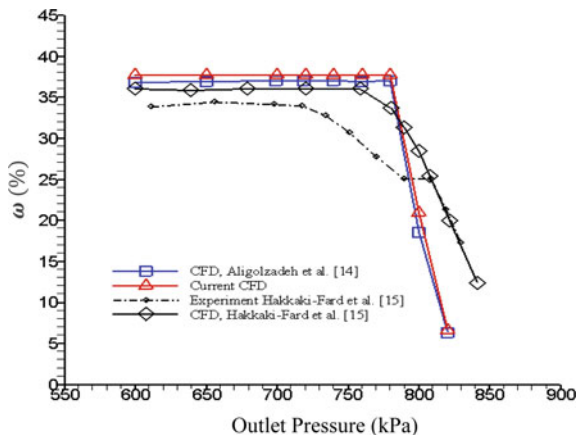
5.2 Influence of the Condensation Pressure on the Performance of the Ejector

The pressure distribution along the ejector wall, as well as the Mach number distribution, is shown in this section. The influence of the condensation pressure on the performance of the ejector is discussed. As shown in Fig. 6 which depicts the pressure distribution contours for two cases of condensing pressure, when the outlet pressure increased from 720 to 800 kPa, the pressure distribution is larger and closer to the primary nozzle. Figure 7 presents the simulation of the Mach number with the two cases of condensation pressure. We notice that in case (a), the variation of the Mach number starts just after the throat, with relatively important values at the level of the mixing chamber, which represents a succession of oblique shocks called a shock train, then at the entrance of the diffuser, a normal shock wave appears, which shows us the good functioning of the ejector (on-design regimes). Beyond the latter, the distribution of the Mach number begins to decrease along the divergent part, and thus the passage from supersonic flow to subsonic flow. According to the second case, and along with the same reasoning, one observes, at the level of the mixing chamber, a rather important difference in the Mach number, and the shock train is driven to the starting position of the mixing chamber. In this case, we note the absence of normal shock at the entrance of the diffuser, which means that we are in the zone of malfunction of the ejector (off-design regimes). We conclude that when the condensation pressure increases, the supersonic velocity decreases inversely, bringing the shock train closer to the primary nozzle.
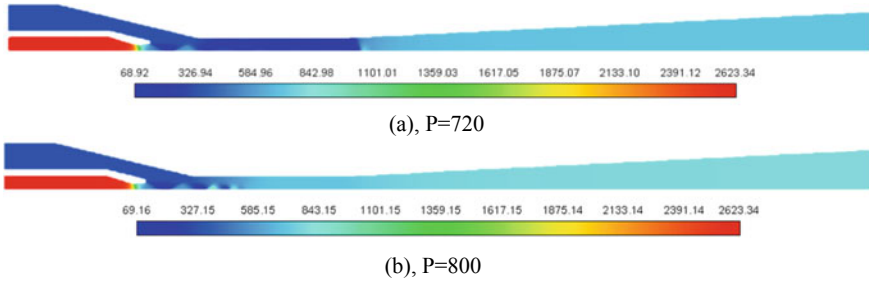
(a), P=720

(b), P=800

Fig. 6 Pressure distribution along the ejector for different values of P
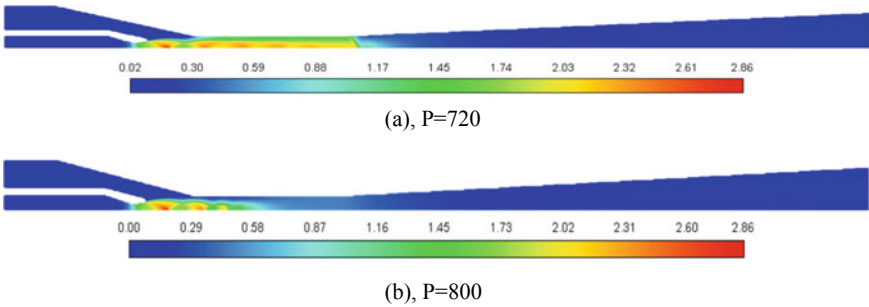

(a), P=720

(b), P=800

Fig. 7 Mach number distribution along the ejector for different values of P

5.3 The Effect of Geometrical Parameters on the Ejector's Performance

Figure 8 shows the effect of the primary nozzle exit height on the entrainment ratio. It shows that the increase in the height up to a value of $H_{NE} = 4.95$ mm causes an increase in the entrainment ratio. After exceeding this last value, the entrainment ratio starts to decrease. Through this numerical simulation, it is concluded that there is an optimal range of values for the primary nozzle exit height that equals 4.95 mm for the ejector to obtain its best performance under constant operating conditions.

The primary nozzle outlet position (NXP), which refers to the distance between the primary nozzle outlet and the mixing chamber inlet, is a critical geometric component that influences ejector performance. Different values of the ratio $N = NXP/H_{TH}$ were applied to three H_{NE} cases, using a fixed value of the H_{TH} throat height, to evaluate the influence of NXP on ejector performance (Table 2).

Figure 9 compares the results of the entrainment ratio obtained for three numerical simulation cases as a function of the ratio N. It can be seen that the entrainment ratio increased with the increase in the ratio N for all cases due to the increase in the effective surface area, until the optimum values were obtained for each case. But further increases in N resulted in a decrease in the entrainment ratio. This may be

Fig. 8 Effect of the primary
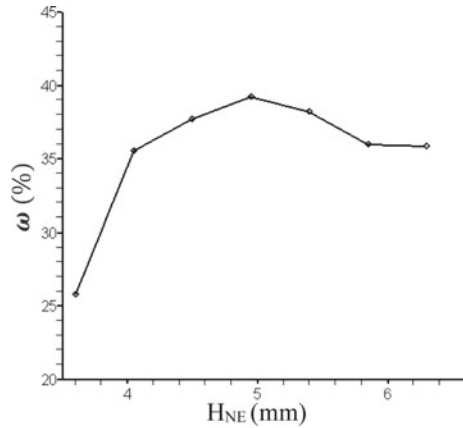nozzle exit height on the
entrainment ratio

**Table 2** The effect of NXP on ejector performance for different values of H_{NE}

	$N = NXP/H_{TH}$	7.46	8.28	9.11	9.94	10.77	11.6	12.43	13.26
$\omega(\%)$	Case 1: $H_{NE} = 4.5$	37.88	37.67	38.17	39.92	43.25	45.11	42.47	40.84
	Case 2: $H_{NE} = 4.95$	38.73	39.25	40.54	42.70	45.91	44.63	42.97	41.92
	Case 3: $H_{NE} = 5.4$	36.55	38.20	40.91	44.58	44.44	42.41	41.13	40.60

due to the formation of small recirculation zones in the converging part of the mixing
chamber, which reduces the effective secondary flow area, and thus NXP has a greater
effect on the effective secondary flow area. In conclusion, there is also an optimal
NXP for a maximum entrainment ratio. Finally, we can see that the second case, with
its optimal NXP value, has the highest entrainment ratio, indicating that it is the best
performing.

6 Conclusion

The impact of condensing pressure and geometric parameters on the performance
of the ejector of a refrigeration system using R134a as a heat transfer fluid has been
studied by numerical simulations. The flow is assumed to be stationary in a two-
dimensional model and axisymmetric with adiabatic walls. The results reveal that
the numerical model is well suited to calculating the performance of supersonic ejec-
tors under various operating conditions. Compared with the experimental data, the
relative error of the entrainment ratio was found to be less than 5.12%. Two simu-
lations were performed by adjusting the condensing pressure, and it was found that
the ejector performance is strongly influenced by the condensing pressure. Addi-
tionally, as the condensation pressure increases, the supersonic velocity drops in the
opposite direction, bringing the shock train closer to the primary nozzle. Then, we

Fig. 9 Effect of NXP on the entrainment ratio for different values of H_{NE}
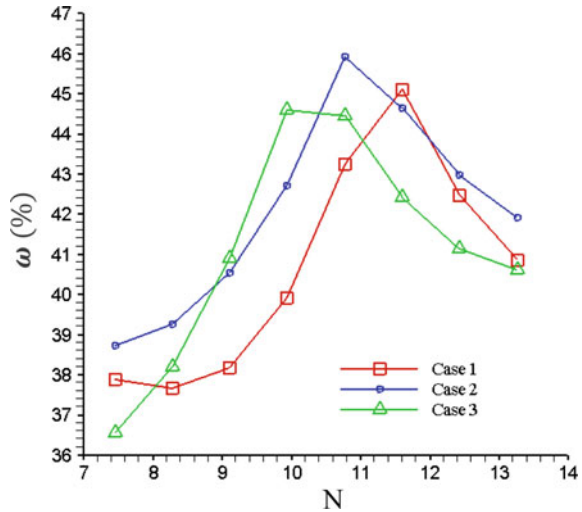
examined different values of the primary nozzle exit height and concluded that there is an optimal H_{NE} height range for the ejector to achieve its best performance under constant operating conditions. Moreover, we investigated the influence of NXP on the ejector performance in three H_{NE} cases. We discovered that there is an optimal NXP for maximum entrainment ratio, and we finally found that the second case with its optimal NXP value has the highest entrainment ratio, indicating that it performs best.

References

1. Varga S, Lebre PMS, Oliveira AC (2013) CFD study of a variable area ratio ejector using R600a and R152a refrigerants. Int J Refrig 36(1):157–165
2. Śmierciew K, Gagan J, Butrymowicz D, Karwacki J (2014) Experimental investigations of solar driven ejector air-conditioning system. Energy Build 80:260–267
3. Besagni G, Mereu R, Inzoli F (2016) Ejector refrigeration: a comprehensive review. Renew Sustain Energy Rev 53:373–407
4. Li Y, Yu J (2019) The effects of ejector geometry parameter and refrigerant charge amount on an ejector-expansion refrigeration system. Appl Therm Eng 152:402–408
5. Ruangtrakoon N, Aphornratana S (2019) Design of steam ejector in a refrigeration application based on thermodynamic performance analysis. Sustain Energy Technol Assess 31:369–382
6. Chen J, Jarall S, Havtun H, Palm B (2015) A review on versatile ejector applications in refrigeration systems. Renew Sustain Energy Rev 49:67–90
7. Thongtip T, Aphornratana S (2017) An experimental analysis of the impact of primary nozzle geometries on the ejector performance used in R141b ejector refrigerator. Appl Therm Eng 110:89–101
8. Croquer S, Poncet S, Aidoun Z (2016) Turbulence modeling of a single-phase R134a supersonic ejector. Part 2: local flow structure and exergy analysis. Int J Refrig 61:153–165
9. Besagni G, Inzoli F (2017) Computational fluid-dynamics modeling of supersonic ejectors: screening of turbulence modeling approaches. Appl Therm Eng 117:122–144

10. Hanafi AS, Mostafa GM, Waheed A, Fathy A (2015) 1-D mathematical modeling and CFD investigation on supersonic steam ejector in MED-TVC. Energy Procedia 75:3239–3252
11. Baek S, Ko S, Song S, Ryu S (2018) Numerical study of high-speed two-phase ejector performance with R134a refrigerant. Int J Heat Mass Transf 126:1071–1082
12. Croquer S, Poncet S, Aidoun Z (2015) Etude Numérique d'un Ejecteur Monophasique Utilisant Le fluide Frigorigène R-134A. In: XIIème Colloque Interuniversitaire Franco-Québécois sur la Thermique des Systèmes, Sherbrooke Québec, Canada, 8–10 June 2015, pp ART-11-10
13. Aligolzadeh F, Hakkaki-Fard A (2019) A novel methodology for designing a multi-ejector refrigeration system. Appl Therm Eng 151:26–37
14. Hakkaki-Fard A, Poirier M, Aidoun Z, Ouzzane M, Giguere D (2015) An experimental study of ejectors supported by CFD. Refrig Sci Technol 2030–2037

Optimization of Industrial Energy Efficiency by Intelligent Predictive Maintenance Tools, Case of Coupling Misalignment and Unbalance for an Industrial System

Yousra El Kihel, Amar Bakdid, and Ali El Kihel

Abstract Today's industry presents many challenges whose increasing competitiveness weighs heavily on productivity. A new way of organizing industrial processes involves the use of precise technologies and the digitization of measured data from the production apparatus. This new organization must meet the challenges of competitiveness to meet customer expectations but with short time to market and at the cost of optimized production in terms of energy consumption, reduced breakdowns, etc. In this study, we have developed a method of monitoring the main components of an industrial installation continuously and follow the evolution through data that allow us to prevent a malfunction before stopping the production (in our case misalignment and unbalance). This monitoring uses predictive algorithms and very precise technologies for the detection of anomalies, which allows to follow parameters in real time as vibration, consumed energy, temperature of the various components, etc. The method is applied on a test tank made in the laboratory and gives very satisfactory results.

Keywords Predictive maintenance · Vibration analysis · Energy · Temperature · Power consumption

1   Introduction

Today's industrial environment presents many challenges whose increasing competitiveness weighs heavily on productivity. Controlling costs while guaranteeing the desired level of quality is the key challenge of industrial production.

Y. El Kihel (✉)
IUT of Bordeaux-Bastide, Department MLT, Bordeaux University, Bordeaux, France
e-mail: elkihel.yosra@gmail.com

A. Bakdid · A. El Kihel
Laboratory of Industrial Engineering and Seismic Engineering, National School of Applied Sciences ENSA-Oujda, Mohammed Premier University, Oujda, Morocco

The Industry of the Future requires a new way of organizing industrial processes with the aim of implementing maintenance tools that are more intelligent and more capable of adapting to production. This new organization should make it possible to meet the challenges of production to meet customer expectations but with a short time to market and at the price of an optimized production in terms of energy consumption, reduced breakdowns, etc.

Predictive maintenance is of paramount importance whose strength is the anticipation of breakdowns, by executing according to the analysis and evaluation of the significant parameters of the degradation of the considered element (vibration, energy, temperature, etc.) [1].

To this end, predictive maintenance is a powerful tool for optimizing the availability of manufacturing and production factories, the profitability of maintenance work and can also be used to reduce energy consumption [2, 3].

Predictive maintenance provides factual data on the actual mechanical condition of each machine and the operating efficiency of each process [4]. In general, the techniques used for predictive maintenance are monitoring vibration analysis, infrared thermography, ultrasonic testing and oil analysis, etc. Each technique has a unique data, and its sensitivity depends on the severity of the defect [5, 6].

Our contribution is to develop a monitoring method of the main components of an industrial installation, in continuous evolution through quantifiable and qualifiable data allows the prediction of malfunctioning before production stoppage. Many authors have addressed the field of industrial monitoring, thus highlighting the growing interest shown by industry in this issue [7, 8].

This predictive maintenance, which has been developed in the form of a PSI [9] industrial monitoring platform using precise instruments, makes it possible to anticipate factory failures and thus considerably increase the availability and reliability before a breakdown occurs and increase productivity [9, 10]. On the other hand, one can highlight the energy consumption according to the state of health of the industrial equipment. Indeed, it is now common practice to add sensors to production machines, which continuously measure the state of their operation. It is a question of monitoring and analyzing the evolution of an element to find the best time to replace it. This work deals with the coupling of two failures which are misalignment and unbalance. In this work, we show how these failures impact the industrial system in terms of vibration and energy consumption [11].

2   Experimental Monitoring Protocol

2.1   *Operating Principal of the IMP*

The Industrial Monitoring Platform is an application with two components [12]:

Hardware Components: based on the implementation of a network of sensors communicating with each other via a proprietary protocol, ensuring data acquisition and transfer to the platform.

Software Components: allows users to view remotely and in real time: the energy consumed by the industrial system, the temperature and the vibration level. Considering the predefined thresholds of each parameter, the intervention time can be predicted to avoid production stops.

2.2 IMP Description

The experimental device or monitoring platform "IMP" is the latest generation of data acquisition equipment. It allows us to accurately measure, acquire, monitor and analyze data from test benches. The data can be acquired directly on a local PC or in a decentralized way via a local network.

Our platform is composed of a three-phase motor group with a variable speed drive (from 500 to 3000 tr/min).

2.3 Operating Principal of the IMP

Three main functions of the IMP: connect and transfer data from instruments to a PC, analyze and interpret results, and predict and anticipate failures.

This operation is summarized in the [13].

3 Creation of Misalignment

An experimental study on misalignment and its impact on vibration, energy consumption and temperature has already been done by us [9].

Figures 1, 2, and 3 represent the evolution of the acceleration, the power consumption and the heating temperature of the motor as a function of the thickness with different shaft speeds (from 500 to 3000 tr/min) [12].

According to this figure, it can be seen that the acceleration increases in an almost linear way as the thickness and the speed of rotation of the shaft increase.

By definition, the power consumed by the electric motor is written:

$$P_0 = C_0 \times \omega \tag{1}$$

where

P_0: Power consumption without defects (neither unbalance nor misalignment);

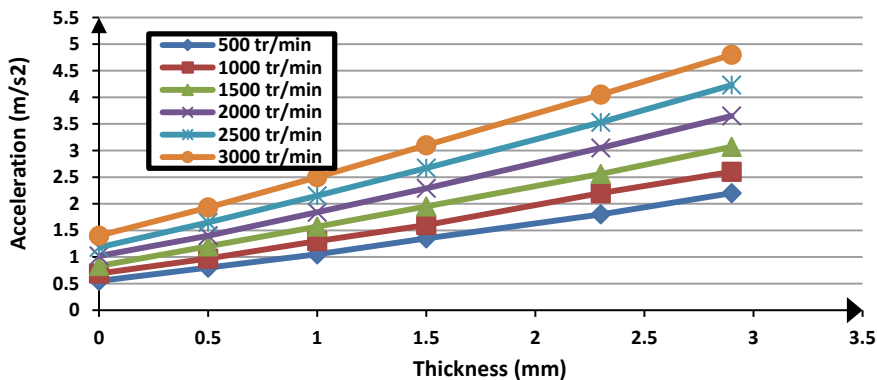C_0: The torque before the creation of the defects;

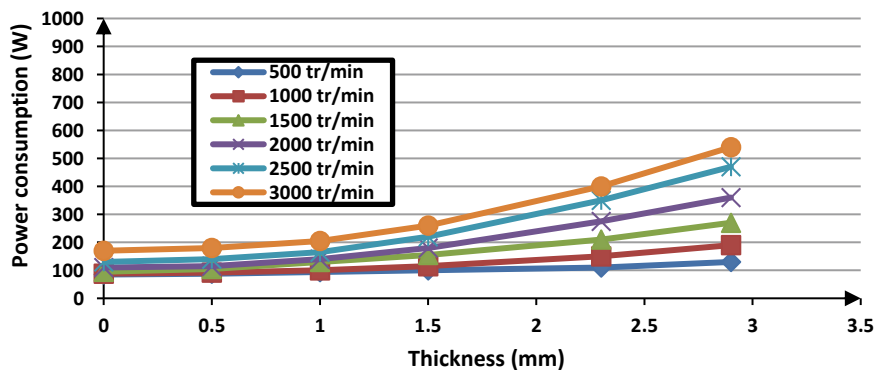**Fig. 1** Acceleration evolution according to misalignment

Fig. 2 Evolution of power consumption according to misalignment
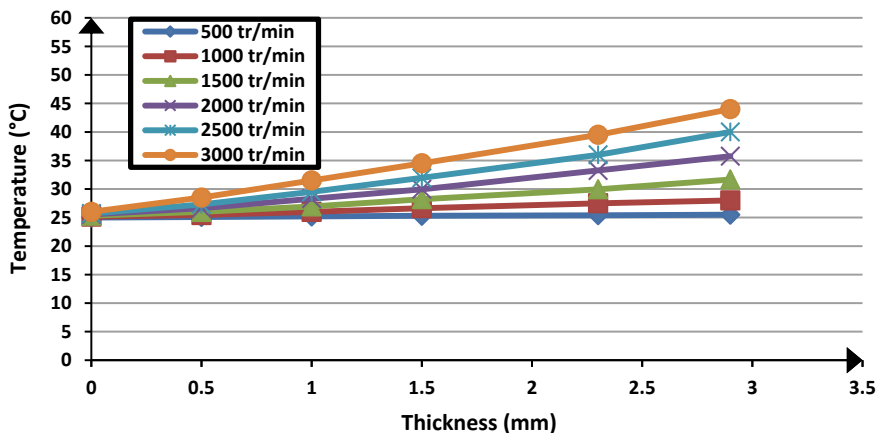

Fig. 3 Evolution of the external temperature of the engine according to the misalignment

ω: The speed of rotation.

If one wishes to express the power consumed in the present misalignment, it is advisable to arrange the relation (1), and one then obtains:

$$P_1 = (C_0 + C_d) \times \omega \tag{2}$$

where

P_1: The power consumed in the present of misalignment;

C_d: The added torque due to misalignment.

We conclude that the term misalignment introduces a torque on the system, so the power consumed by the electric motor increases with the misalignment and the speed of rotation, and this consumption starts to increase from 2000 tr/min.

When the rotation speed and misalignment increase, the temperature increases.

In the following part, we present the coupling of an electric motor and an unbalance; the study is based essentially on the analysis of the different quantities and their influences on the state of each component of the motor and its energy efficiency (Fig. 4).
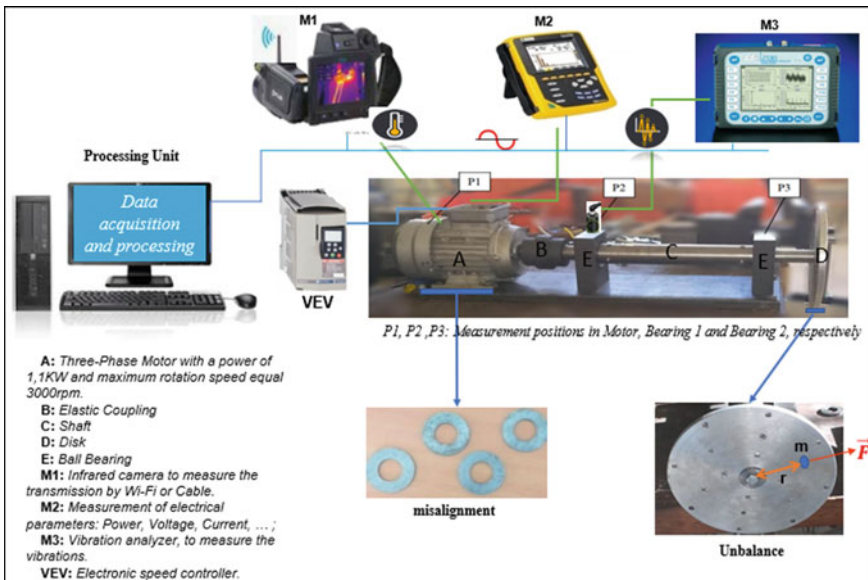

Fig. 4 Creation of misalignment and unbalance

4 Materials and Methods

4.1 Coupling of Misalignment and Unbalance

Based on the previous results, it can be seen that the effect of misalignment is readable on the system from a thickness of 3 mm. In the following section, the thickness is set to 3 mm.

Next, an unbalance is created by changing the mass "m" and setting the distance between the mass and the center ("r" $= 100$ mm $=$ cte).

The unbalance creates a rotating force, which equals:

$$F = m * r * \omega^2 * \sin \omega t \tag{3}$$

where

F: Unbalance force (N), m: Mass (kg), r: Rayon (m), ω: Rotation speed (rad/s).

Finally, the results (Figs. 5, 6 and 7) of the physical quantities are presented: acceleration, power consumption and outside temperature.

These results are measured using a vibration analyzer, an electrical analyzer and an infrared camera [13].

5 Results and Discussion

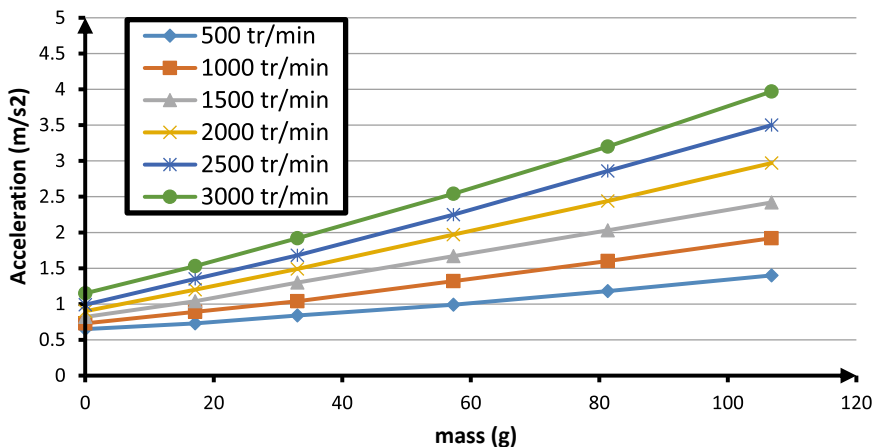The following figures represent the results obtained during the experimental study.
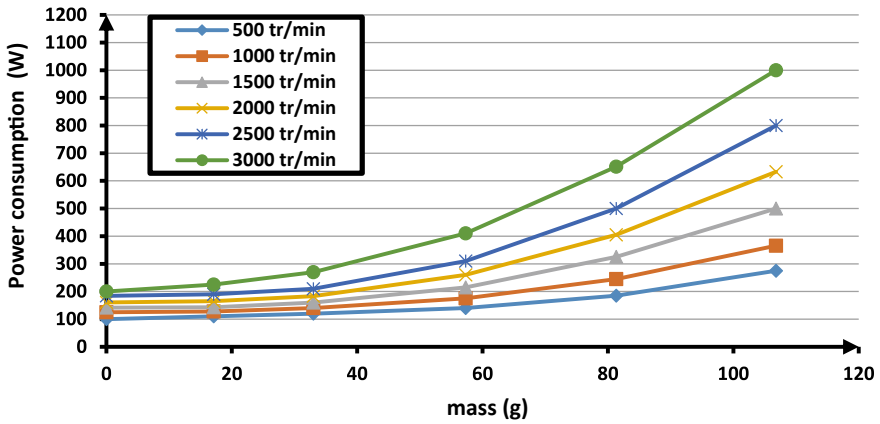

Fig. 5 Acceleration evolution

Fig. 6 Evolution of power consumption
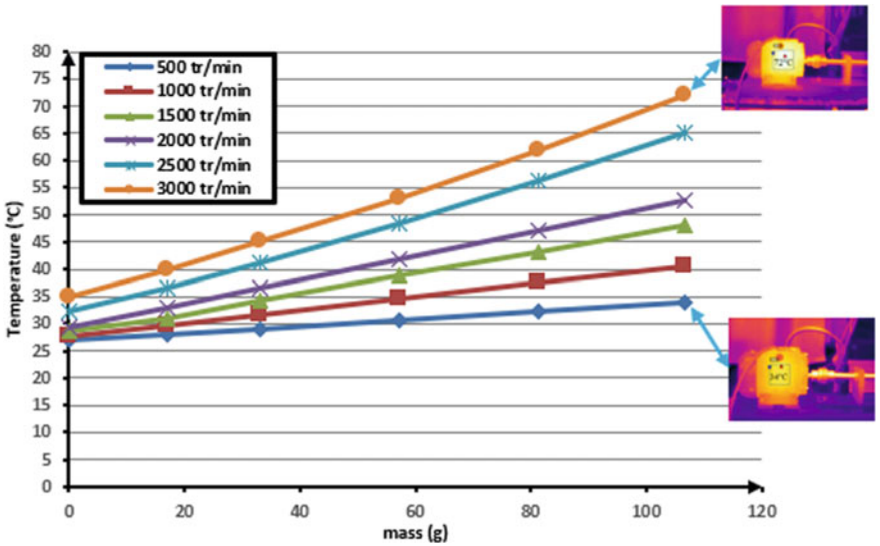
Fig. 7 Temperature outside the motor according to unbalance

5.1 The Impact of Coupling on Acceleration

Figure 5 shows the results of the different acceleration measurements with a constant thickness equal to 3 mm.

According to Fig. 5, we can see that the acceleration increases according to the mass in an almost linear way whatever the speed of rotation.

As the mass increases, the slope of the acceleration increases considerably with increasing speed [14].

5.2 The Power Consumed by the Drive Motor According to the Coupling

The power consumed by the motor is obtained by adding the unbalance torque:

$$P_2 = (C_0 + C_d + C_b) \times \omega \qquad (4)$$

where

P_2: Power consumption in the presence of faults (misalignment and unbalance);

C_b: Torque due to unbalance.

Power consumption "P_2" is then composed of the torque exerted by the unbalance and by the misalignment with the angular velocity.

Figure 6 shows the power consumption of the electric motor for the different masses.

Analyzing the results obtained, it can be seen that the power consumed by the engine increases exponentially with the increase in mass. It is also shown that the power increases considerably for the three speeds 2000, 2500 and 3000 tr/min. This phenomenon is explained by the resistant torque $C(e)$ due to the unbalance and the misalignments introduced.

5.3 The Impact of Misalignment and Unbalance on Motor Temperature

Figure 7 shows the measured temperature at the drive motor for the different unbalances applied.

The temperature of the engine increases as the mass increases at constant speed. It becomes more important at rotational speeds above 2000 tr/min.

6 Conclusion

A new method for monitoring an industrial system by analyzing and processing physical quantities has been presented in this paper. The approach is based on data acquisition, resulting in real-time processing, whose good performances have been illustrated using experimental data. Our objective is to study the impact of misalignment and unbalance on the system.

The connected measuring instruments allow following the evolution of each physical quantity (acceleration due to vibration, power consumption and engine warm-up temperature). The processing of the results is done in real time in a very precise way.

The results obtained show that increasing the misalignment thickness and/or the unbalance mass has negative effects on the industrial system. These effects reduce the

reliability and efficiency of the system and increase its energy consumption. Finally, knowing the temperature and vibration thresholds at the bearings, we can predict failures and schedule interventions to avoid risks or production stoppages.

References

1. Dalzochio J et al (2020) Machine learning and reasoning for predictive maintenance in Industry 4.0: current status and challenges. Comput Ind 123:103298
2. Jeffali F, Ouariach A, El Kihel B, Nougaoui A (2019) Diagnosis of three-phase induction motor and the impact on the kinematic chain using non-destructive technique of infrared thermography. Infrared Phys Technol 102:102970
3. Elkihel A, Abouelanouar B, Gziri H (2021) Industrial energy audit methodology for improving energy efficiency—a case study. Lecture Notes Electr Eng 681:675–681
4. Keith Mobley R (2002) An introduction to predictive maintenance, 2nd edn
5. Hajdu F, Szalai P, Mika P, Kuti R (2019) Parameter identification of a fire truck suspension for vibration analysis. Pollack Periodica 14(3):51–62
6. Duan L, Yao M, Wang J, Bai T, Zhang L (2016) Segmented infrared image analysis for rotating machinery fault diagnosis. Infrared Phys Technol 77:267–276
7. Regad Y, El Kihel B, Boushaba F, Chourak M (2019) Numerical study of the smoke plume of the Jerada thermal power plant Morocco. Pollack Periodica
8. Suguna M, Roomi M, Sanofer I (2016) Fault localisation of electrical equipments using thermal imaging technique, pp 1–3
9. Elkihel A, Abouelanouar B, Gziri H (2020) Rotating machines energy loss due to unbalance, pp 300–308
10. McMillan RB (2003) Rotating machinery: practical solutions to unbalance and misalignment, 1st edn. Fairmont Press
11. Hajdu F, Molnárka G (2020) Testing output variables for sensitivity study of nonlinear vibration systems. Pollack Periodica 15(2):70–81
12. Singh G, Naikan VNA (2017) Infrared thermography based diagnosis of inter-turn fault and cooling system failure in three phase induction motor. Infrared Phys Technol 87:134–138
13. El Kihel A, El Kihel Y, Bakdid A, Manssouri I, Amegouz D (2020) Optimization of industrial energy efficiency by intelligent Predictive Maintenance tools Case of misalignment of an industrial system. In: 2020 IEEE 2nd international conference on electronics, control, optimization and computer science, ICECOCS 2020, 9314505
14. Abouelanouar B, Elkihel A, Gziri H (2020) Experimental study on energy consumption in rotating machinery caused by misalignment. SN Appl Sci 2(7):1215

The Influence of Reynolds Number and Baffles on the Thermos-Hydrodynamic Behavior of Circular Pipe in Three Dimensions

Jamal-Eddine Salhi, Nassreddine Hmidi, Tarik Zarrouk, Abdel-illah Amrani, Ahmed Alami Merrouni, Elmiloud Chaabelasri, Merzouki Salhi, and Najim Salhi

Abstract This study looked at how steady-state three-dimensional turbulent forced convection flow and heat transfer worked in a circular pipe with baffles fixed inside the pipe. The numerical study was performed for a Reynolds number increasing from 100,000 to 140,000, a Prandtl number of 0.71, and a baffle angle of 90°. The findings demonstrate that the Nusselt number and friction factor of the circular pipe with baffles are greater than the smooth circular pipe in this study (without baffles). As a general observation, the addition of baffles and the Reynolds number plays an important role in the flow and heat transfer characteristics. Comparing the data acquired with those obtained from a smooth duct serves to verify the correctness of the results obtained.

Keywords Baffles · Forced convection · Heat exchangers · CFD

J.-E. Salhi (✉) · N. Hmidi · T. Zarrouk · A. A. Merrouni · E. Chaabelasri
Equipe des Sciences des Matériaux, Énergies Nouvelles et Applications, Laboratoire LPTPME, Département de Physique, Faculté des Sciences, Mohammed First Université, 60000 Oujda, Maroc
e-mail: j.salhi@ump.ac.ma

T. Zarrouk
e-mail: Zarrouk.tarik@ump.ac.ma

J.-E. Salhi · M. Salhi · N. Salhi
CREHEIO, Centre de Recherche de l'Ecole des Hautes Etudes d'Ingénierie, 60000 Oujda, Morocco

A. Amrani
Equipe des Energies Nouvelles et Ingénierie des Matériaux (ENIM), Laboratoire de Sciences et Techniques de L'Ingénieur (LSTI), Faculté des Sciences et Techniques, Université Mou-Lay Ismail, Boutalamine, BP 509, 52000 Errachidia, Morocco

1   Introduction

Expanding the convective heat transfer coefficient or increasing the heat transfer surface area improves heat transmission from a surface. A pump or fan is employed, or an existing one is replaced with a bigger one to improve the heat transfer coefficient. However, this may or may not be convenient. Increasing the surface area of a surface is another way to improve heat transmission from that surface. When baffles or fins are placed on the duct surfaces, the surface area for heat transmission may be enhanced [1–4].

Many technical applications use circular pipes with turbulent flows, including nuclear reactors, shell and tube heat exchangers, gas turbines and combustion chambers, and electronic cooling equipment.

A large number of experimental and computational studies have been carried out to determine the influence of various geometrical characteristics of baffles, such as form, size, spacing between baffles, attack angle, blockage ratio, and porous space, on the performance of the baffles under consideration. The geometry of baffles is a critical parameter for vortex creation [5]. As a result, a variety of forms have been devised, including the inclined baffles [6, 7], the W-baffle [8], the Z-baffle [9], and multi-V-type perforated baffles [10], and the diamond shape presented by Sripattanapipat et al. [11].

The numerical analysis of the steady-state three-dimensional turbulent flow and thermal characteristics in a tubular heat exchanger with baffles is performed in this paper. We investigate the influence of Reynolds number and baffle number on thermohydrodynamic behavior. The outcomes of this work are supposed to offer numerical data on turbulent convective heat transport in complicated geometries in three dimensions.

2   Formulation of the Problem

The computational environment and the properties of the circular pipe employed in this investigation are shown in Fig. 1. Tube: It is a cylindrical tube with a circular cross-section with a diameter of 150 mm. It has a length of 1000 mm and is crossed by a turbulent flow of air. The wall temperature is brought to Tw = 343 K, and it stays that way. All the required details on the geometrical parameters are provided in Table 1.

3   Governing Equations

Three-dimensional Navier–Stokes and energy equations explain flow and heat transfer in the computational domain. Turbulent flow is defined as incompressible
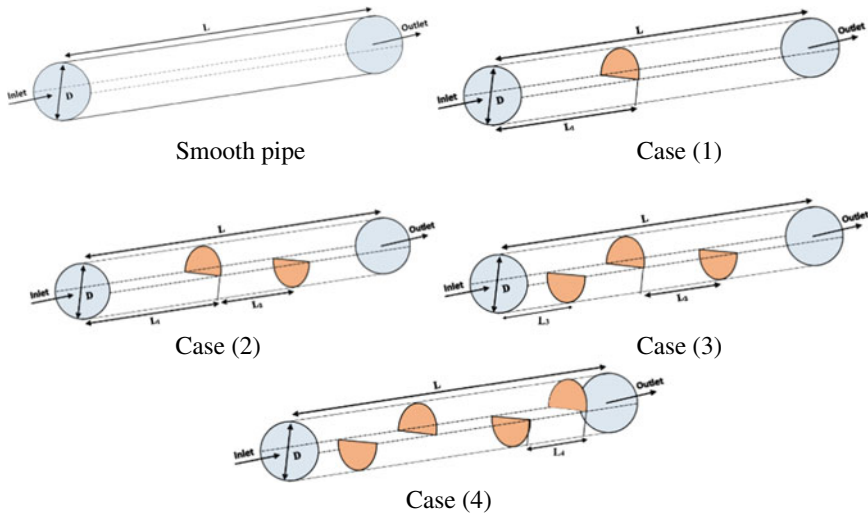
Fig. 1 Schematic view of the computational domain

Table 1 Specifics about the physical problem's size

Parameters	L (mm)	L$_1$ (mm)	L$_{2;3;4}$ (mm)	D (mm)	Number of baffles
Smooth pipe	1000	400	200	150	0
Case (1)	1000	400	200	150	1
Case (2)	1000	400	200	150	2
Case (3)	1000	400	200	150	3
Case (4)	1000	400	200	150	4

Newtonian flow in three dimensions with minimal buoyancy and viscous dissipation at a steady state. Thermophysical properties are immutable. The governing equations for turbulence in this scenario are as follows:

$$\frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0 \tag{1}$$

$$\frac{\partial(\rho u^2)}{\partial x} + \frac{\partial(\rho uv)}{\partial y} + \frac{\partial(\rho uw)}{\partial z} = -\frac{\partial P}{\partial x} + \frac{\partial}{\partial x}\left[(\mu_l + \mu_t)\left(\frac{\partial u}{\partial x}\right)\right]$$
$$+ \frac{\partial}{\partial y}\left[(\mu_l + \mu_t)\left(\frac{\partial u}{\partial y}\right)\right] + \frac{\partial}{\partial z}\left[(\mu_l + \mu_t)\left(\frac{\partial u}{\partial z}\right)\right] \tag{2}$$

$$\frac{\partial(\rho uv)}{\partial x} + \frac{\partial(\rho v^2)}{\partial y} + \frac{\partial(\rho wv)}{\partial z} = -\frac{\partial P}{\partial y} + \frac{\partial}{\partial x}\left[(\mu_l + \mu_t)\left(\frac{\partial v}{\partial x}\right)\right]$$

$$+ \frac{\partial}{\partial y}\left[(\mu_l + \mu_t)\left(\frac{\partial v}{\partial y}\right)\right] + \frac{\partial}{\partial z}\left[(\mu_l + \mu_t)\left(\frac{\partial v}{\partial z}\right)\right] \tag{3}$$

$$\frac{\partial(\rho uw)}{\partial x} + \frac{\partial(\rho vw)}{\partial y} + \frac{\partial(\rho w^2)}{\partial z} = -\frac{\partial P}{\partial z} + \frac{\partial}{\partial x}\left[(\mu_l + \mu_t)\left(\frac{\partial w}{\partial x}\right)\right]$$
$$+ \frac{\partial}{\partial y}\left[(\mu_l + \mu_t)\left(\frac{\partial w}{\partial y}\right)\right] + \frac{\partial}{\partial z}\left[(\mu_l + \mu_t)\left(\frac{\partial w}{\partial z}\right)\right] \tag{4}$$

$$\frac{\partial(\rho uT)}{\partial x} + \frac{\partial(\rho vT)}{\partial y} + \frac{\partial(\rho wT)}{\partial z} = \frac{\partial}{\partial x}\left[\left(\mu_l + \frac{\mu_t}{\sigma_T}\right)\left(\frac{\partial T}{\partial x}\right)\right]$$
$$+ \frac{\partial}{\partial y}\left[\left(\mu_l + \frac{\mu_t}{\sigma_T}\right)\left(\frac{\partial T}{\partial y}\right)\right] + \frac{\partial}{\partial z}\left[\left(\mu_l + \frac{\mu_t}{\sigma_T}\right)\left(\frac{\partial T}{\partial z}\right)\right] \tag{5}$$

The standard two-equation transport model k-ε models the turbulence phenomenon. In general, the two additional equations provided by the $k-\varepsilon$ turbulence model used:

For the kinetic energy of turbulence (k):

$$\frac{\partial(\rho uk)}{\partial x} + \frac{\partial(\rho vk)}{\partial y} + \frac{\partial(\rho wk)}{\partial z} = \frac{\partial}{\partial x}\left[\left(\mu_l + \frac{\mu_t}{\sigma_k}\right)\left(\frac{\partial k}{\partial x}\right)\right]$$
$$+ \frac{\partial}{\partial y}\left[\left(\mu_l + \frac{\mu_t}{\sigma_k}\right)\left(\frac{\partial k}{\partial y}\right)\right] + \frac{\partial}{\partial z}\left[\left(\mu_l + \frac{\mu_t}{\sigma_k}\right)\left(\frac{\partial k}{\partial z}\right)\right] \tag{6}$$

For the dissipation rate of kinetic energy (ε):

$$\frac{\partial(\rho u\varepsilon)}{\partial x} + \frac{\partial(\rho v\varepsilon)}{\partial y} + \frac{\partial(\rho w\varepsilon)}{\partial z} = \frac{\partial}{\partial x}\left[\left(\mu_l + \frac{\mu_t}{\sigma_\varepsilon}\right)\left(\frac{\partial \varepsilon}{\partial x}\right)\right]$$
$$+ \frac{\partial}{\partial y}\left[\left(\mu_l + \frac{\mu_t}{\sigma_\varepsilon}\right)\left(\frac{\partial \varepsilon}{\partial y}\right)\right] + \frac{\partial}{\partial z}\left[\left(\mu_l + \frac{\mu_t}{\sigma_\varepsilon}\right)\left(\frac{\partial \varepsilon}{\partial z}\right)\right] \tag{7}$$

where

$$\mu_t = \rho c_\mu k^2 / \varepsilon \tag{8}$$

The various empirical constants in the k-ε model are given the following values:

$$C_{1\varepsilon} = C_{3\varepsilon} = 1.44,\ C_{2\varepsilon} = 1.92,\ C_\mu = 0.09,\ \sigma_k = 1.0,\ \sigma_\varepsilon = 1.3 \tag{9}$$

4 Boundary Conditions

The boundary conditions of the computational domain are defined as follows:

- At the entrance of the channel: ($U = U_{\text{int}}$; $V = W = 0$; $T = T_{\text{int}} = 300$ K°)
- Wall (adhesive contact): ($U = V = W = 0$; $T_w = 343$ K°)

- Pipe output (output condition): $(P = P_{atm}; (\partial\varphi / \partial x) = 0$, where $\phi = U, V, W, T, k, \varepsilon)$
- Baffle: $(U = V = W = 0$; adiabatic).

The following are the primary parameters of the present survey defined by the formula:

- The friction factor (f):

$$f = \frac{2}{\left(\frac{L}{D_h}\right)} \times \frac{\Delta P}{\rho U_m^2} \tag{10}$$

- Average Nusselt Number (Nu):

$$Nu_{avg} = \frac{h D_h}{\lambda} \tag{11}$$

5 Grid Independence Study

The mesh that has been chosen is a tetrahedral mesh. Additionally, to ensure the independence of the accepted mesh size values, additional tests were conducted for the Reynolds number Re = 100,000. From the grid of 888,227 cells, the A Nusselt number and factor of friction begin to stabilize. For our further calculations, and as a trade-off between accuracy and computation time, the 990,767 cell grid is ideal (see Fig. 2).

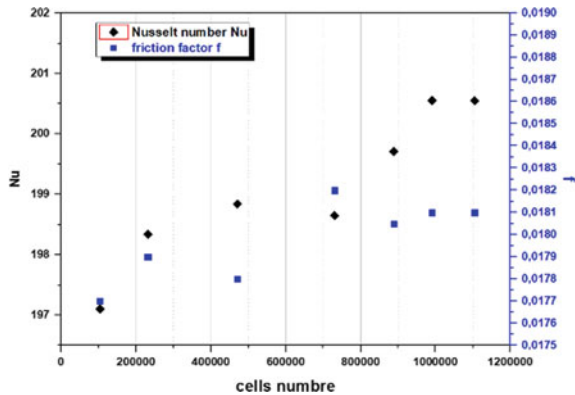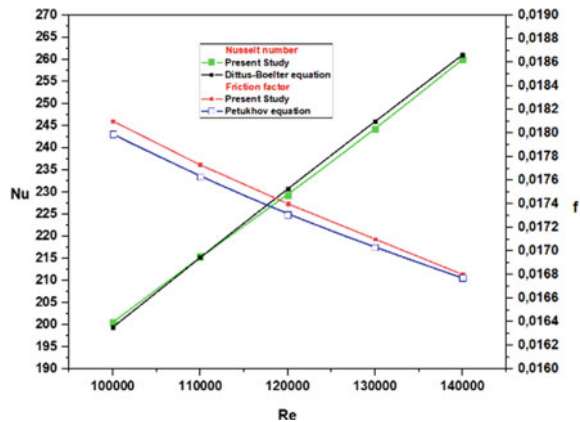**Fig. 2** Grid-independency study for Re = 100,000

Fig. 3 Variation of (Nu) and (f) as a function of Reynolds for a smooth pipe

6 Validation of the Numerical Model

An analytical correlation between Dittus-Boelter and Petukhov's correlations was used to validate our computational model. Figure 3 shows the numerical and analytical outcomes of the calculations. The Nusselt number evolution curves are practically identical, demonstrating high agreement between simulated and analytical findings. This comparison helps us to validate our simulation and validate our numerical code.

7 Results and Discussion

Several parameters can illustrate heat exchanges from one medium to another. In this paper, we have discussed four parameters that may be useful for a better understanding the heat transfers created between a heat exchanger and the air (see Fig. 4).

First, we carried out a study on the fluid velocity within the studied geometry. The figure illustrates the air velocity field within a smooth heat exchanger and circular cross-section heat exchangers with baffles.

In the first case, which characterizes a smooth heat exchanger, we can easily notice that the air velocity varies slightly. It is between 12.5 ms^{-1} and 17.5 ms^{-1}. Nevertheless, in the other cases, it is remarkable that the fluid velocity is influenced by the presence and the number of baffles incorporated in the exchanger. This influence is generally illustrated by the increase of velocity in the channel. Moreover, the maximum velocity fields are much more pronounced for a high number of baffles. Note that the observed increase is particularly during the passage of fluid particles between the baffles and the walls of the exchanger. This can be explained by an inversely proportional relationship linking the increase in velocity and the decrease in the section through which the fluid flows.

Fig. 4 Velocity contours for different case

Smooth pipe

Case (1)

Case (2)

Case (3)

Case (4)
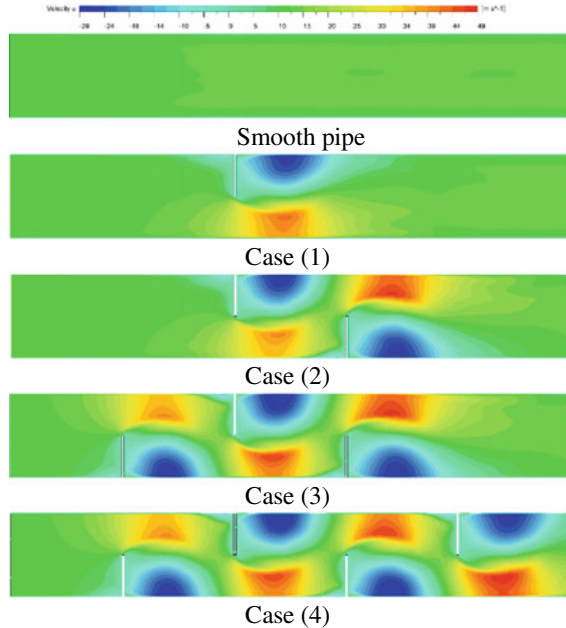
On the other hand, the figure shows that the presence of baffles causes the zones of air recirculation located in the right part of the baffles. This causes a phenomenon of air confinement and leads the fluid to follow a serpentine flow path.

To clarify the thermal behavior created inside the heat exchanger, we considered it important to carry out a study on the distribution of the temperature field within the geometry studied. Figure 5 shows the form of the isotherms produced during heat exchanges in an exchanger with and without baffles.

In the absence of the baffles, a constant temperature is observed in most of the heat exchanger volume, except in the vicinity of the walls of the channel, where there is a slight temperature variation. It should also be noted that the temperature is improving at the level of the channel walls near the outlet of the heat exchanger. The superposition of the isotherms, which are concentrated at the edge of the walls of the geometry studied, provides for the dominance of the conduction mode thermal about the convective mode.

The inclusion of baffles in the heat exchanger presents one of the ideal solutions for improving heat exchange. This is clearly illustrated by corrugated isotherms having a temperature of 13 degrees higher than the temperature observed in the center of an exchanger without baffles. The most important thing to mention is that these isotherm ripples coincide with the previously observed air recirculation zone. In other words, the residence time of the fluid particles in this place gives more time for the air to acquire a maximum amount of heat delivered by the exchanger walls. So we will get an increase in temperature in this area. Note that, this phenomenon is repeated near each baffle, so the passage of air in the vicinity of these areas will be accompanied

Fig. 5 Temperature contours for different case

Smooth pipe

Case (1)

Case (2)

Case (3)

Case (4)
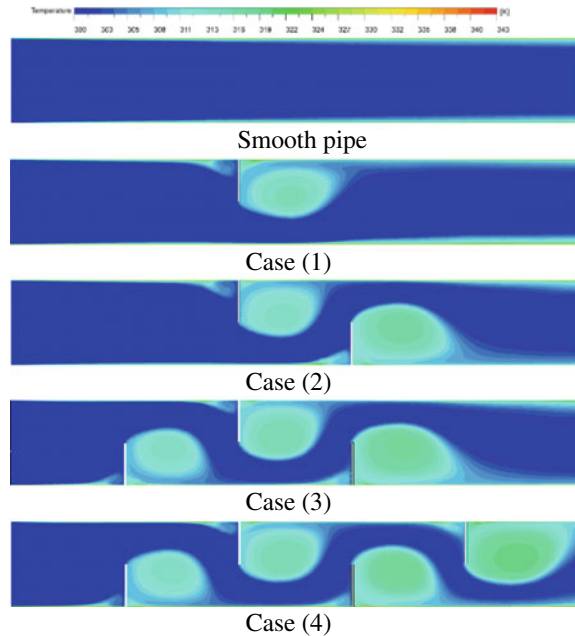
by an improvement in the temperature within the canal, from which a remarkable homogeneity of the temperature distribution field within the heat exchanger will be obtained.

In order to quantify the heat exchanges created within the heat exchanger studied, it was necessary to carry out a study on the Nusselt number. This number reflects the heat transfer rate between the exchanger and the fluid.

Figure 6a shows the variation of this number as a function of the Reynolds number for the various cases studied previously. One can easily notice that the rate of heat exchange (Nu) improves as the number of Reynolds increases. Also, it is lucid to see that for a specific number of Reynolds, the number of Nusselt increases with the number of baffles. This is mainly due to the increase of high-temperature air recirculation zones, which causes a decrease in fluid density and promotes the dominance of convection that homogenizes the temperature within the heat exchanger.

About Fig. 6b, which illustrates the variation of the friction factor as a function of the Reynolds number, the study of this factor provides us with more information on the causes of the thermal behavior produced within the exchanger.

The analysis of the histograms shows that the friction factor decreases with the increase of the Reynolds number. This is because a reciprocal formulation binds this factor with velocity. As a result, for high values of the Reynolds number, the velocity becomes increasingly high, which causes the observed decrease of the friction factor for the large Reynolds values. In other words, this can be explained by the fact that the velocity becomes low for low Re values, so the fluid particles reside more in the
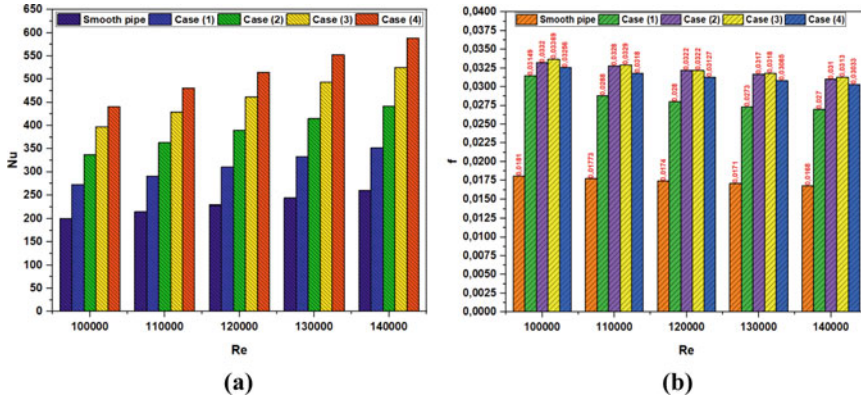
Fig. 6 **a** Evolution of (Nu), and **b** Evolution of (f) as a function of (Re) for different case

exchanger, which leads to more contact with the channel walls, hence the increase of this factor the small values of Re.

8 Conclusions

The performance of a heat exchanger with varying numbers of internal baffles is examined in this research. The study was done for four cases: one, two, three, and four baffles, respectively. All of the experiments were carried out using air as the working fluid in a turbulent flow regime with Re between 100,000 and 140,000. There are a few things that may be concluded from the results:

- In all circumstances, (Nu) rises in direct proportion to (Re).
- Reducing (Re) causes the friction factor (f) to drop.
- Increases in Nu and f are a result of baffles being added to the pipe. The bigger the number of baffles, the higher the values for both parameters.

References

1. Salhi JE, Zarrouk T, Merrouni AA, Salhi M, Salhi N (2022) Numerical investigations of the impact of a novel turbulator configuration on the performances enhancement of heat exchangers. J Energy Storage 46:103813
2. Boukhadia K (2018) Effect of the perforation design on the fluid flow and heat transfer characteristics of a plate fin heat exchanger. Int J Therm Sci 126:172–180
3. Salhi JE, Salhi M, Amghar K, Zarrouk T, Salhi N (2021) Analysis of the thermohydrodynamic behavior of a cooling system equipped with adjustable fins crossed by the turbulent flow of air in forced convection. Int J Energy Environ Eng 1–13

4. Salhi JE, Zarrouk T, Salhi N (2021) Numerical study of the thermo-energy of a tubular heat exchanger with longitudinal baffles. Mater Today Proc 45:7306–7313
5. Singh S (2012) Exergy based analysis of solar air heater having discrete V-down rib roughness on absorber plate. Energy 37:749–758
6. Salhi JE, Amghar K, Bouali H, Salhi N (2020) Combined heat and mass transfer of fluid flowing through horizontal channel by turbulent forced convection. Modell Simul Eng
7. Salhi JE, Amghar K, Filali A, Salhi N (2022) Effect of the spacing design of two alternate baffles on the performance of heat exchangers. Defect Diffus Forum 415:53–72
8. Lanjewar A (2011) Heat transfer and friction in solar air heater duct with W-shaped rib roughness on absorber plate. Energy 36:4531–4541
9. Sriromreun P (2012) Experimental and numerical study on heat transfer enhancement in a channel with Z-shaped baffles. Int Commun Heat Mass Transfer 39:945–952
10. Sharma A (2017) Optimizing discrete V obstacle parameters using a novel entropy-VIKOR approach in a solar air flow channel. Renew Energy 106:310–320
11. Sripattanapipat S, Promvonge P (2009) Numerical analysis of laminar heat transfer in a channel with Diamond-shaped baffles. Int Commun Heat Mass Transfer 36:32–38

Photovoltaic Panels End-of-Life Recycling

Michele Calì and **Alberto Acri**

Abstract The largest contribution of electricity production comes from conventional sources including coal and oil that pollute the environment. Renewable energy sources, including solar energy, wind energy and energy storage in batteries, are expected to play a progressively central role in meeting future energy needs in all sectors, largely responding to the increasing demand for energy. In particular, the use of solar energy will be considered as the main solution to global climate change and fossil fuel emissions. Although today's photovoltaic panels have an average lifespan of 25 years, their disposal is a cause for concern when photovoltaic technology is evaluated from the perspective of comprehensive life cycle analysis and End-of-Life management (EoL). We therefore need some innovative solutions that can reduce emissions of pollutants as a result of the recycling of solar panels that no longer work. This manuscript reports some of the most current efficient and effective photovoltaic (PV) panel recycling solutions and the foreseeable developments for such recycling.

Keywords Photovoltaic panels · End-of-life management · EcoDesign · Sustainability · Renewable resources

1 Introduction

The use of renewable and sustainable energy is fundamental to reduce environmental degradation and limit any significant climate change. The European Commission is currently considering the *Green Deal* by adopting a series of proposals to reduce net greenhouse gas emissions by at least 55% by 2030 compared to 1990 levels.

M. Calì
Electric, Electronics and Computer Engineering Department, University of Catania, 95125 Catania, Italy
e-mail: michele.cali@unict.it

A. Acri (✉)
Department of Engineering, University of Messina, 98158 Messina, Italy
e-mail: alberto.acri@studenti.unime.it

The European Green Deal aims to promote the efficient use of resources so to have a clean and circular economy, stop climate change, restore biodiversity loss and reduce pollution [1].

The realization of these objectives involves the use of *renewable resources*, that is, all those materials and energy sources that are present in nature such as the strength of the wind, the waterways of the rivers or the heat generated by the sun. Through this type of resources, it will be possible to create recyclable and biodegradable materials that not only have a minimal impact in environmental terms, but are also designed to be reused often without losing their original characteristics [2].

To generate a sustainable economy, it is therefore necessary to consider the link between *EcoDesign* and the *circular economy* from the moment in which the former considers the environmental impact that a given product, or production process, will have during all phases of its life cycle.

Europe has an important infrastructure for the production of solar and wind energies and for the storage of energy and the use of portable batteries. Considering, however, that this infrastructure is being replaced by more modern structures and as the maintenance cycle requires a possible replacement of parts, the application of the principles of the circular economy is the basis for exploiting the resource potential of the generated waste and minimizing the challenges of its management. Experts predict that hundreds of thousands of tons of old wind turbines, batteries and solar modules will have to be disposed of or recycled in the next decade as well as millions of tons by 2050 (Fig. 1) [3].

However, the recovery of materials from discarded devices remains impractical because they are manufactured so as not to break and to be able to carry out their work for as long as possible. Today's photovoltaic panel have a lifespan of 25 years. Therefore, materials engineers, manufacturers and waste managers are working to efficiently recover renewable energy materials but at the moment not enough of these panels have reached the end of their life cycle to make the investment in recycling plants actually worthwhile.
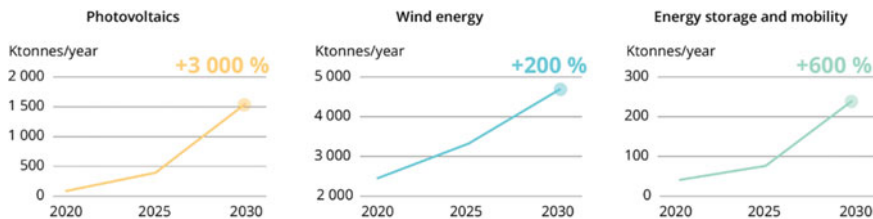

Fig. 1 Growth of waste materials from renewable resources

2 Focus on Photovoltaics

In 2018, photovoltaics became the fastest-growing energy technology in the world. According to the most recent authoritative reports [4], the use of photovoltaic panels in 2018 exceeded 100 GW (Fig. 2 [4]). This growth is due to an increasingly widespread demand leading at the end of 2018 to add further countries with a cumulative capacity of 1 GW or more, to the countries of the previous year. Of all photovoltaic panels installed globally, China has the largest share with 35.3%, followed by the European Union (19%), the USA (11.8%), Japan (10.6%) and India (6%) [5].

This continuous increase is due to a new interest in being able to generate electricity through photovoltaics without causing environmental impacts mainly related to the use of agricultural land. Therefore, the idea was born to apply photovoltaic panels in industrial basins, irrigation ponds and drinking water tanks, using the so-called *floating solar photovoltaic* (FPV) to achieve the ambitious renewable energy goals [6, 7]. These FPV installations represent new opportunities for the spread of photovoltaics, especially in all those countries with a high population density and little land available. In addition, the levelised cost of electricity (LCOE) of floating solar photovoltaic systems does not differ much from that of ground-mounted fixed-inclination systems and is therefore a good innovation.

The birth of FPV also implies opportunities to combine solar photovoltaic technology with hydroelectric power plants. In fact, in 2017, the world's first hybrid and hydroelectric FPV system (220 kW at the Alto Rabagão dam) were installed in Portugal [8]. The construction of plants based on the combination of these two technologies can limit the effects of the variability of solar resources. Indeed, such facilities can make a better use of the existing transmission resources by offering
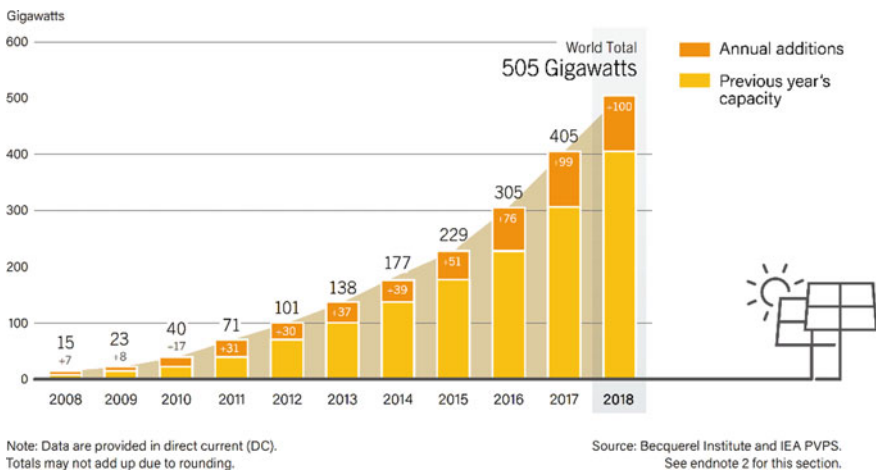

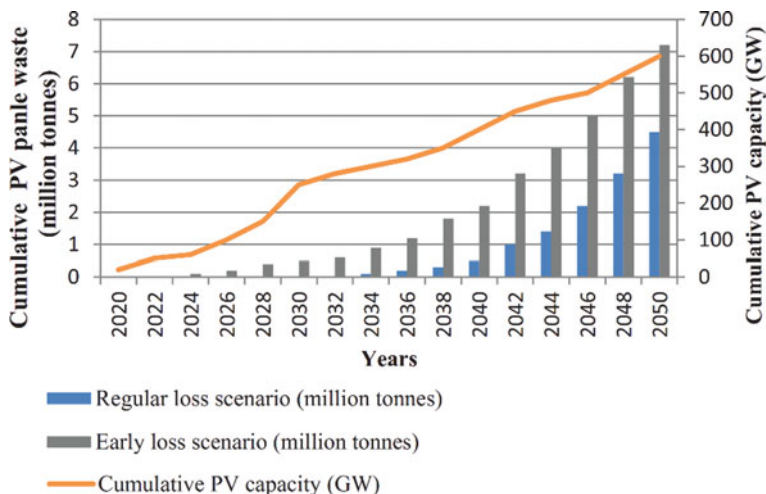**Fig. 2** Exponential growth of energy from photovoltaic panels

Fig. 3 Comparison of the amount of waste produced by solar panels in the event that they reach 25 years

opportunities in particular for those countries where networks are weak or constrained by inflexible generation technologies.

Unfortunately, the increase in the installation of photovoltaic panels signifies also an increment in the amount of waste produced when these panels reach the end of their 25-year lifespan. In short, the number of photovoltaic panels reaching the end-of-life (EoL) stage would increase exponentially as the number of photovoltaic installations increases. At the end of the useful life of these panels, these become harmful waste that threatens the environment. It is therefore essential to establish a low-cost recycling technology for the rapid commercialization and advancement of the photovoltaic industry [9]. Figure 3 quantifies the accumulation of waste due to photovoltaic solar panels in the event that solar panels resist to their useful life of 30 years [9]. In a regular loss scenario (solar panels operating until the age of 30), a dramatic increase of 6 million tons is expected by the end of 2050, while in an early loss scenario (solar panels failing before the age of 30), waste is expected to rise to 8 million tons by 2050.

3 Opportunities and Challenges for Photovoltaics

Although the European Union estimates that the region currently generates 30,000 metric tons per year of PV panel waste and that this will rise to about 500,000 tons per year over the next two decades [10], there is *an opportunity* that 95% of materials can be recycled (e.g., glass, copper, aluminum, etc.) as shown in Fig. 4. From what has been said in [5] it can be estimated that most of the recyclable materials in
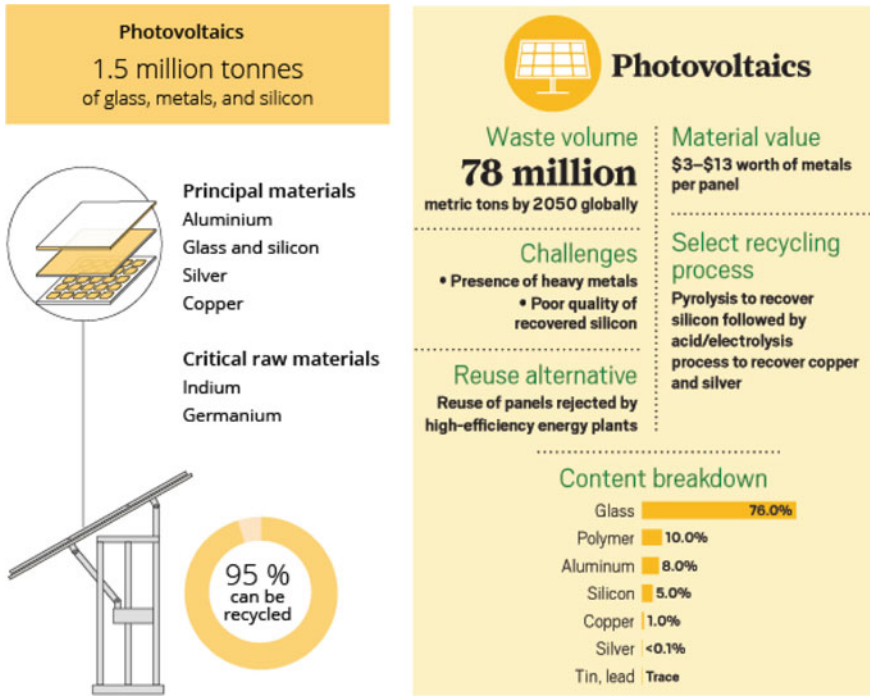
Fig. 4 Quantity of recyclable material in solar panels

photovoltaic panels are based on glass that is present with about 68% by weight, aluminum with about 15% by weight, high purity silicon with about 3% by weight and copper cables with about 1% by weight.

Among the key *challenges* in photovoltaic recycling [3], both in economic and in technological terms, are the delamination, separation and purification of silicon from glass and semiconductor thin film. According to a study conducted by the European Commission, the main problem associated with the improper disposal of waste photovoltaic panels is the possibility of recycling photovoltaic panels due to the presence of hazardous substances such as cadmium, arsenic, lead, antimony, polyvinyl fluoride (PVF) and polyvinylidene fluoride. In an effort to avoid such environmental problems and maximize material recovery, the EU-funded research, including the Full Recovery End-of-Life Photovoltaic (FRELP) project. The FRELP process recovers silicon and other metals by heating the panels in a furnace while the other metals are recovered by electrolysis. An example of a plant based on the technology developed by FRELP is the Sasil plant, inaugurated in 2015 in Italy, which can accommodate 3500 tons of photovoltaic panels per year and which is able to recover about 93% of the materials from the photovoltaic panels used [11, 12].

4   Recycling Challenges

Europe, Japan and the USA are currently conducting research on the recycling of solar panels focusing their main purpose on the recovery of important and valuable components housed inside the used solar panels. Therefore, still challenges are to be faced for photovoltaics to become one of the main sources of electricity worldwide. Since photovoltaic solar panels contain lead (Pb), cadmium (Cd) and many other harmful chemicals, recycling is the major challenge. According to [10], the average life of modern solar panels is 25 years and the most common end-of-life (EoL) technology for photovoltaic components remains their disposal in landfills. This can be quite dangerous as harmful chemicals introduced into the soil can cause contamination of drinking water. In addition, as with normal waste, the incineration process is used, which involves the release of toxic heavy metals into the atmosphere and also renouncing to the possibility of recovering raw materials. The only advantage of this method is that photovoltaic modules do not have to be separated from other commercial or industrial waste [13]. At the legislative level, only the EU has established the rules for the collection and treatment of solar panels in the Waste Electrical and Electronic Equipment (WEEE) Directive (Directive 2012/19/EU). As can be found in the literature, existing industrial technologies for recycling solar panels make it possible to obtain a recovery of 95–97% of cadmium and tellurium for thin-film solar cells, a recovery of 90% of glass and a recovery of 80% of silicon for multi-Si PV models [14]. According to the data reported in [15], the recycling of all components of 1 kW multi-Si photovoltaic panels with modern industrial thermal and chemical methods emits 6.32 kg of SO_2, 23.4 kg NO_2, 13.8 kg CO_2, as well as 0.97 kg of ammonia (NH_3), 2.59 kg of ethylene and 4.26 kg of methane.

From the point of view of real recycling, currently two types of photovoltaic panels produced and used all over the world arise: silicon-based photovoltaic panels and thin-film photovoltaic panels [5]. Both are characterized by their own recycling method since the panels are differentiated by their internal compositions. At the moment, silicon-based solar panels are the most adopted on the market (over 95%) and therefore are at the center of attention [5]. Different techniques can be indicated depending on whether we recycle zinc-based photovoltaic panels or thin-film photovoltaic panels [14].

With *silicon-based photovoltaic panels*, the glass that makes up the coating is separated from the aluminum parts that represent the frame. In particular, the glass is 95% recyclable; all the external metal parts are largely reused to form new frames for solar panels and the remaining materials are heat-treated at a temperature of 500 °C in such a way that the plastic evaporates leaving the silicon elements ready for further processing. The steam generated as a result of heat treatments is not released into the environment but is used as a heat source for further heat treatment. After this treatment, 80% of the panels are reused, while the remaining percentage is subject to additional cleaning. The silicon particles contained in cracked and scratched wafers are acid-etched and then fused to be reused to produce new silicon modules, achieving a recovery level of 85% of the silicon raw material.

Instead, *thin-film photovoltaic panels* must first be crushed to a very low particle size (4–5 mm) to obtain the removal of the foil that holds the internal materials and subsequently a separation of solid and liquid materials is performed. Liquid materials go through a process of precipitation and dehydration to allow the recovery of semiconductor materials that stands at an average recovery of 95%. Solid materials contaminated by so-called interlayer materials are instead cleaned with washes as a result the clean glass remains and must be melted and reused.

In order to be able to recycle 98% of photovoltaic panels, the PHOTORAMA project [16] has been implemented by a consortium of 13 organizations in the period 2021–2024 and funded by the EU precisely to improve the recycling of photovoltaic panels and optimize the recovery of raw materials. The prospects for this new project are of great importance both from an environmental and from an economic point of view because the developing technology will be focused on the eco-design of photovoltaic modules to create a production technique that makes the panel easily recyclable in all its components [17]. According to this project, a new delamination technology employed to separate the solar cells from the glass plate will consider innovative chemical–physical processes so to recover all the materials without resorting to those shredding techniques of the photovoltaic modules in which the cells, which constitute the valuable part of the panel, end up in a recovery chain of low economic value (the so-called down-cycling) [18]. The new technological process will result in an "up-cycling" with the recovery of high-value materials (such as aluminum from the frame and glass and polymers from the plates) and metals from the layer of solar cells (critical such as silicon, indium and gallium and precious as silver). In addition to improving the recovery of components, the recyclability of PV panels and the decrease of harmful chemicals produced by them can be improved through an optimized design aimed at increasing the EoL of the panels. The authors performed an in-depth statistical analysis on the main damage/failure causes, proposing innovative design solutions. Through panel Thermo-Mechanical Fatigue (TMF) analysis using a parametric Finite Element (FE) model, authors proved that it is possible to increase the total life cycle of the PV panel by more than 40%. As an example in Fig. 5 the PV total life cycle was evaluated with and without silicone layer. More details of this new design solution are described in the manuscript "The Design Value for Recycling End-of-Life Photovoltaic Panels" Calì M. et. al. [19].
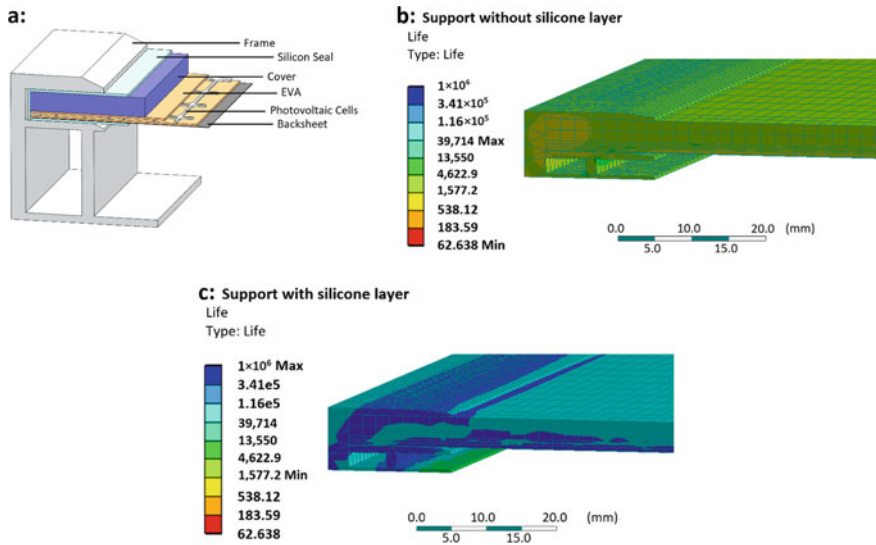
Fig. 5 (A) Parametric model of silicon-based PV panels with support; (B) Total life cycle without silicone layer; (C) Total life cycle with silicone layer

5 Conclusions

Photovoltaic technology is among the most useful technologies for the environment compared to the various technologies of power generation and electricity if this is evaluated from the point of view of the life cycle, including the management of the end of life. On the other hand, photovoltaic solar panels are considered as responsible for an increasingly high amount of waste per unit of energy than any other source of electricity generation. This means that the recycling of solar panels will be essential to mitigate the environmental problems related to EoL treatments and allows, in this way, to further increase the opportunities for economic and social development while reducing greenhouse gas emissions through the generation of electricity. Currently only Europe has effective specific regulations on the recycling of solar panels, while other countries in the world have different not fully ecological regulations. Through recycling, it is possible to reuse some materials and focus on a circular economy by exploiting projects that consider end-of-life products. Authors, after an in-depth analysis of the main damage / failure causes, propose innovative design solutions that allow to extend the EoF of the PV pannels components and to increase their recyclability. This approach will lead to positive consequences for consumers since the products will last longer and can be easily updated and repaired.

References

1. Munta M (2020) The European Green Deal: A Game Changer or Simply a Buzzword? Friedrich Ebert Stiftung: Zagreb, Croatia
2. Stahel W.R (2016) The Circular Economy. Nature 531, 435−438
3. Chowdhury S, Rahman K.S, Chowdhury T, Nuthammachot N, Techato K, Akhtaruzzaman M, Tiong S.K, Sopian K, Amin N (2020) An overview of solar photovoltaic panels' end-of-life material recycling. Energy Strat. Rev. 27, 100431. https://doi.org/10.1016/j.esr.2019.100431
4. Renewables Homepage. https://www.ren21.net/wp-content/uploads/2019/05/gsr_2019_full_report_en.pdf/. Accessed 23 Apr 2022
5. Majewski P, Al-shammari W, Dudley M, Jit J, Lee SH, Myoung-Kug K, Sung-Jim K (2021) Recycling of solar PV panels-product stewardship and regulatory approaches. Energy Policy 149:112062. https://doi.org/10.1016/j.enpol.2020.112062
6. Mittal D, Saxena BK, Rao KVS (2017) Floating solar photovoltaic systems: an overview and their feasibility at Kota in Rajasthan. In: 2017 international conference on circuits, power and computing technologies. IEEE, pp 1–7. https://doi.org/10.1109/ICCPCT.2017.8074182
7. Patil SS, Wagh MM, Shinde NN (2017) A review on floating solar photovoltaic power plants. Int J Sci Eng Res 8:789–794
8. Gadzanku S, Beshilas L, Grunwald U (2021) Enabling floating solar photovoltaic (FPV) deployment. Review of barriers to FPV deployment in Southeast Asia. Technical Report. NREL. https://doi.org/10.2172/1787553
9. Rathore N, Panwar NL (2021) Strategic overview of management of future solar photovoltaic panel waste generation in the Indian context. Waste Manag Res 1–15. https://doi.org/10.1177/0734242X211003977
10. Chowdhury S, Rahman KS, Chowdhury T, Nuthammachot N, Techato K, Akhtaruzzaman M, Tiong SK, Sopian K, Amin N (2020) An overview of solar photovoltaic panels' end-of-life material recycling. Energy Strat Rev 27:100431. https://doi.org/10.1016/j.esr.2019.100431
11. Latunussa CE, Ardente F, Blengini GA, Mancini L (2016) Life cycle assessment of an innovative recycling process for crystalline silicon photovoltaic panels. Sol Energy Mater Sol Cells 156:101–111. https://doi.org/10.1016/j.solmat.2016.03.020
12. Ardente F, Latunussa CEL, Blengini GA (2019) Resource efficient recovery of critical and precious metals from waste silicon PV panel recycling. Waste Manag 91:156–167. https://doi.org/10.1016/j.wasman.2019.04.059
13. Lunardi MM, Alvarez-Gaitan JP, Bilbao JI, Corkish R (2018) A review of recycling processes for photovoltaic modules. Sol Panels Photovolt Mater. https://doi.org/10.57722/intechopen.74390
14. Ratner S, Gomonov K, Revinova S, Lazanyuk I (2020) Eco-design of energy production systems: the problem of renewable energy capacity recycling. Appl Sci 10:4339. https://doi.org/10.3390/app10124339
15. Huang B, Zhao J, Chai J, Xue B, Zhao F, Wang X (2017) Environmental influence assessment of China's multi-crystalline silicon (multi-Si) photovoltaic modules considering recycling process. Sol Energy 143:132–141. https://doi.org/10.1016/j.solener.2016.12.038
16. Photorama Homepage. https://www.photorama-project.eu/. Accessed 23 Apr 2022
17. Paiano A (2015) Photovoltaic waste assessment in Italy. Renew. Sustain. Energy Rev. 41, 99–112
18. Korniejenko K, Kozub B, Bak A, Balamurugan P, Uthayakumar M, Furtos G (2021) Tackling the circular economy challenges—composites recycling: used tyres, wind turbine blades, and solar panels. J Compos Sci 5:243. https://doi.org/10.3390/jcs5090243
19. Calì M, Bekkay H, Nitto G, Acri A (2022) The Design Value for Recycling End-of-Life Photovoltaic Panels. Appl Sci 12(18):9092. https://doi.org/10.3390/app12189092