

Lecture Notes in Electrical Engineering 1109

Rabindra Nath Shaw · Pierluigi Siano ·
Saad Makhilef · Ankush Ghosh ·
S. L. Shimi *Editors*

Innovations in Electrical and Electronic Engineering

Proceedings of ICEEE 2023, Volume 1

 Springer

Lecture Notes in Electrical Engineering

Volume 1109

Series Editors

- Leopoldo Angrisani, Department of Electrical and Information Technologies Engineering, University of Napoli Federico II, Napoli, Italy
- Marco Arteaga, Departamento de Control y Robótica, Universidad Nacional Autónoma de México, Coyoacán, Mexico
- Samarjit Chakraborty, Fakultät für Elektrotechnik und Informationstechnik, TU München, München, Germany
- Jiming Chen, Zhejiang University, Hangzhou, Zhejiang, China
- Shanben Chen, School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
- Tan Kay Chen, Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore
- Rüdiger Dillmann, University of Karlsruhe (TH) IAIM, Karlsruhe, Baden-Württemberg, Germany
- Haibin Duan, Beijing University of Aeronautics and Astronautics, Beijing, China
- Gianluigi Ferrari, Dipartimento di Ingegneria dell'Informazione, Sede Scientifica Università degli Studi di Parma, Parma, Italy
- Manuel Ferre, Centre for Automation and Robotics CAR (UPM-CSIC), Universidad Politécnica de Madrid, Madrid, Spain
- Faryar Jabbari, Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA
- Limin Jia, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
- Janusz Kacprzyk, Intelligent Systems Laboratory, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
- Alaa Khamis, Department of Mechatronics Engineering, German University in Egypt El Tagamoa El Khames, New Cairo City, Egypt
- Torsten Kroeger, Intrinsic Innovation, Mountain View, CA, USA
- Yong Li, College of Electrical and Information Engineering, Hunan University, Changsha, Hunan, China
- Qilian Liang, Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX, USA
- Ferran Martín, Departament d'Enginyeria Electrònica, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain
- Tan Cher Ming, College of Engineering, Nanyang Technological University, Singapore, Singapore
- Wolfgang Minker, Institute of Information Technology, University of Ulm, Ulm, Germany
- Pradeep Misra, Department of Electrical Engineering, Wright State University, Dayton, OH, USA
- Subhas Mukhopadhyay, School of Engineering, Macquarie University, Sydney, NSW, Australia
- Cun-Zheng Ning, Department of Electrical Engineering, Arizona State University, Tempe, AZ, USA
- Toyoaki Nishida, Department of Intelligence Science and Technology, Kyoto University, Kyoto, Japan
- Luca Oneto, Department of Informatics, Bioengineering, Robotics and Systems Engineering, University of Genova, Genova, Genova, Italy
- Bijaya Ketan Panigrahi, Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, Delhi, India
- Federica Pascucci, Dipartimento di Ingegneria, Università degli Studi Roma Tre, Roma, Italy
- Yong Qin, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
- Gan Won Seng, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore
- Joachim Speidel, Institute of Telecommunications, University of Stuttgart, Stuttgart, Germany
- Germano Veiga, FEUP Campus, INESC Porto, Porto, Portugal
- Haitao Wu, Academy of Opto-electronics, Chinese Academy of Sciences, Haidian District Beijing, China
- Walter Zamboni, Department of Computer Engineering, Electrical Engineering and Applied Mathematics, DIEM—Università degli studi di Salerno, Fisciano, Salerno, Italy
- Junjie James Zhang, Charlotte, NC, USA
- Kay Chen Tan, Department of Computing, Hong Kong Polytechnic University, Kowloon Tong, Hong Kong

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. ****

Rabindra Nath Shaw · Pierluigi Siano ·
Saad Makhilef · Ankush Ghosh · S. L. Shimi
Editors

Innovations in Electrical and Electronic Engineering

Proceedings of ICEEE 2023, Volume 1

 Springer


Editors

Rabindra Nath Shaw
University Center for Research
and Development (UCRD)
Chandigarh University
Mohali, Punjab, India

Pierluigi Siano
Department of Management and Innovation
Systems
University of Salerno
Fisciano, Italy

Saad Makhilef
Department of Electrical Engineering
Swinburne University of Technology
Melbourne, VIC, Australia

Ankush Ghosh 
University Center for Research
and Development (UCRD)
Chandigarh University
Mohali, Punjab, India

S. L. Shimi 
Department of Electrical Engineering
Punjab Engineering College (Deemed to be
University)
Chandigarh, Punjab, India

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-99-8288-2 ISBN 978-981-99-8289-9 (eBook)
<https://doi.org/10.1007/978-981-99-8289-9>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Paper in this product is recyclable.

Preface

This book features selected high-quality papers presented at 4th International Conference on Electrical and Electronics Engineering (ICEEE 2023), jointly organized by Chitkara University, Himachal Pradesh, India, and ADSRS Education and Research, India, during August 19–20, 2023, in online mode. The conference got overwhelming response and received more than 200 papers from all around the world. All submitted papers have gone through single blind review process on an average three review per paper. The acceptance rate is less than 25%. The presented papers published in this book chapter. The book focuses on current development in the fields of electrical and electronics engineering. The book one covers electrical engineering topics—power and energy including energy distribution and transmission, renewable energy, power electronics and applications, control, robotics and automation and instrumentation and book two covers the areas of robotics, artificial intelligence and IoT, electronics devices, circuits and systems, wireless and optical communication, RF and microwaves, VLSI, and signal processing. The book is beneficial for readers from both academia and industry.

We are thankful to all the authors that have submitted papers for keeping the quality of ICEEE 2023 at high levels. The editors of this book would like to acknowledge all the authors for their contributions and the reviewers. We have received invaluable help from the members of the International Program Committee and the chairs responsible for different aspects of the workshop. We also appreciate the role of special sessions organizers. Thanks to all of them, we had been able to collect many papers on interesting topics, and during the conference, we had very interesting presentations and stimulating discussions.

We hope that these volumes will provide useful information to professors, researchers, and graduated students in the area of electrical engineering, electronics and communication engineering, and computer science engineering, along with AI

and IoT applications, and all will find this collection of papers inspiring, informative, and useful. We also hope to see you at a future ICACIS event.

Mohali, India
Fisciano, Italy
Melbourne, Australia
Mohali, India
Chandigarh, India

Rabindra Nath Shaw
Pierluigi Siano
Saad Makhilef
Ankush Ghosh
S. L. Shimi

Contents

Bidding Strategies for Generation Companies and Large Consumers in Carbon Emission Market Considering Electricity Spot Market Clearing Outcomes	1
Liyang Liu, Kun Wang, Junwei Ma, Honglin Xue, Xiaohui Wang, Wei Jiang, and Fushuan Wen	
NSTLBO-Based Approach for Optimal Scheduling of Hydrothermal Generating Units in Regulated Environment	15
Baburao Pasupulati, Bhargava Reddy Sibbala, S. Sivakumar, S. Amosedinakaran, and Rajakumar Palanisamy	
Novel Design for Power Factor Correction Using FACTS Device and Soft Computing Tools	27
Shubham Gedam, Prema Daigavane, and Tushar Guhe	
Chaos and Multistability in Fractional Order Power System: Dynamic Analysis and Implications	37
Prakash Chandra Gupta and Piyush Pratap Singh	
Identification of Critical Nodes Using Granger Causality for Strengthening Network Resilience in Electrical Distribution System	49
Divyanshi Dwivedi, D. Maneesh Reddy, Pradeep Kumar Yemula, and Mayukha Pal	
A Bald Eagle Search Optimization Approach for Congestion Alleviation in Power System Transmission Channels	61
Nisha Kumari, Kaushik Paul, and Pampa Sinha	
Optimal Power Flow Analysis Using Pelican Optimization Algorithm	73
Prakash Kumar, Md. Abul Kalam, and Kaushik Paul	

Performance Evaluation on Developed FRP Rods Used in Composite Insulators Using Rotating Wheel and Dip Facility	89
M.-Ramez Halloum, B. Subba Reddy, G. Nithin Reddy, Zoya Zaki, Febin Francis, and B. Kiran Kumar Reddy	
MFO Algorithm-Based Profit Maximization of Distribution Companies by Optimal Placement and Sizing of DGs Under Deregulated Environment	101
Ch. S. K. B. Pradeep Kumar, G. Balamurugan, Y. Butchi Raju, and Pasupulati Baburao	
Peer-To-Peer Transaction Model Among Prosumers Considering Franchise Rights of Distribution Companies	119
Zhemlin Lin, Bochun Zhan, Zihao Li, Changsen Feng, and Fushuan Wen	
Comparative Study of Different Grid Connected Wind Generator	137
Vaishnavi Pachkawade and Rutuja Hiware	
Photovoltaic with Battery and Supercapacitor Energy Storage System for Better Performance Devices and Modelling	151
Somesh Harinkhede and Prajakta Vaidya	
Optimal Planning of Hydrogen Refueling Stations Considering Balanced Utilization of Resources	163
Xuyao Meng, Fushuan Wen, Kun Ding, Changhai Yang, and Yalu Sun	
A Communication Technique to Improve the Performance of On Load Tap Changer of Transformer with Renewable Load	177
Roshni M. Virkhare, Prema Daigavane, and Hari Kumar Naidu	
Performance Analysis of Wind Power Forecasting via System Advisor Model Software	189
J. Sathyaraj and V. Sankardoss	
Performance of Fuzzy MPPT for a Distributed Photovoltaic System Under Changing Environmental Condition	201
P. Rajakumar, M. Senthil Kumar, A. Sakthidasan, R. Sundar, Baburao Pasupulati, and S. Amosedinakaran	
Enhancing Photovoltaic Connector Reliability: A Comparative Review of Studies with Practical Recommendations	219
Kartik Kapoor, Muthusamy Eswaramoorthy, Devendra Goyal, P. G. Nikhil, and Prem Prakash Bharti	
A Comprehensive Study of Power Quality Improvement Techniques in Smart Grids with Renewable Energy Systems	241
Manpreet Singh and Lakhwinder Singh	

Fault Detection at PCC Using Wavelet Theory in Grid-Tied Solar PV Battery-Based AC Microgrid 253
 Sarika S. Kanojia and Aagam Shah

Powering the Future: IoT-Enabled Smart Grids for Sustainable Energy Systems 273
 Anshu Prakash Murdan

Solar Charging Station for Electric Vehicles 287
 Ruchi Gautam and Manjeet Singh

Theoretical Analysis of Tandem Solar Cell Doped with $MASn_3$ with P3HT: PCBM Active Layer 293
 Ankit Mishra and Manoj Kumar Nigam

Sensorless PMSM Motor Control Strategies by Artificial Neural Network (ANN) with PI Controller 305
 Sagar Appaso Talke, Arshad Mansur Tamboli, Yashraj Mohan Shinde, and Dipak Suresh Yeole

A Comprehensive Review on Electric Vehicle Battery Swapping Stations 317
 Md Tahmid Hussain, Ahmad Bin Afzal, Altaf Hussain Thakurai, Ahmar Azim, Rashid Ahmed khan, Hasan Alrajhi, Mohd Rizwan Khalid, and Salman Hameed

Comparison Between PID and SMC Controller to Control the Speed of DC Separately Excited Motor 333
 Prakansha Sulakhe, V. P. Rajderkar, and Tushar Guhe

Implementation and Evaluation of a Secure Keyless Ignition and Integrated Control System for an Arduino-Based Instrumented Car 343
 Anshu Prakash Murdan and Nikhil Kissoon

Selection of Automated Guided Vehicles for Industrial Application Using Weighted Sum Method 357
 Divya Sabnis, Megha Patil, and Sagar Wankhede

Active Disturbance Rejection Control of a SEPIC Converter 367
 Priyanshu Kumar and Moina Ajmeri

Designing a Mapless Navigation Mobile Robot Using Deep Q Learning 381
 Chinmay Kolhatkar and Kranti Wagle

Implications of Location of Strain Gauges and Excitation Voltage Over the Metrological Performance of Trapezoidal-Shaped Force Transducer 393
 Richa Saxena and Harish Kumar

Improvisation in Human–Robot Interaction Using Optimized Multimodal Operational Techniques	403
Pratiksha Prakash Jawale and Shantipal S. Ohol	
A Co-occurring Rule Mining Approach to Discover Viewers Watching Behavior of OTT Platform	415
Apurva Vashist, Suchismita Mishra, and Anil Kumar Sagar	
Using Interferometer-Based Photonic Integrated Circuits to Perform Rapid Unitary Transformations	425
Sathvik Redrouthu, Jagadeepam Maddipatla, David Cao, Ryan Kim, and Steven Lu	
NASA Nearest Earth Object Classification Using Quantum Machine Learning: A Survey	439
Aman Singh Bhogal, Mausmi Sinha, and Pratiksha Meshram	
Studying the Effect of Type of Surface Passivation Layer on Performance Parameters of AlGaN MSM Detector	457
Harpreet Kaur and Manish Kumar Hooda	
Tachyon: A Programmable Optoelectronic Hardware Accelerator for Ultrafast Tensor Arithmetic	467
Sathvik Redrouthu, Jagadeepam Maddipatla, Pranav Vadde, and Anant Khandelwal	
Performance Analysis of Multi-user Cooperative Non-orthogonal Multiple Access on Time-Sharing Basis	489
Smriti Dubey, Neelesh Mehra, and Devendra Kumar Shakya	
TL-Shaped Circular Parasitic Compact Planar Antenna for 5G Microwave Applications	507
Ridhi Gupta, Manish Varun Yadav, and Swati Varun Yadav	
Systematic Literature Review on the Machine Learning Techniques for UAV-Assisted mm-Wave Communications	517
M. S. Sugesh and G. Vairavel	
Transmission of Wi-Fi Signals on 11 kV Power Transmission Lines—A Laboratory Study	535
B. Subba Reddy, L. Umanand, Vishram Pandit, Parthasarathy Ramaswamy, G. Nithin Reddy, M. Ramez Halloum, and Zoya Zaki	
An Intelligent LoRa-Based Wireless Sensor Network Mesh Architecture to Improve Precision Agriculture	545
Aman Shaikh, Maya Shelke, Satayush Rai, Pranil Ashok Rao, and Ganesh Shinde	

Low Power High Speed Modified GDI Full Adder Implemented by NCFET Technologies 559
Priyanka Kattepogu and Atul S. M. Tripathi

Design and Analysis of Non-vulnerable PRBS Generation with Internal Shift and XOR of LFSR 573
Jagadeshwaran Ramasamy and Dhandapani Samiappan

ML- and IoT-Based Crop Prediction System 583
Kushagra Sharma and Deepak Kumar

Multifocus, Infrared and Visible Light Image Fusion Using Non-sampled Shearlet Transform and SUSAN Operator 591
Dileep Kumar Sharma and Abhilasha Sharma

Leveled Approach of Context Setting in Semantic Understanding of Remote Sensing Images 599
Stuti Ahuja and Sonali Patil

Emotion Recognition: A New Tool for Healthcare Using Deep Learning Algorithms 613
Krishna Mridha, Tasnim Sarker, Rawnoqz Zaman, Madhu Shukla, Ankush Ghosh, and Rabindra Nath Shaw

A Comprehensive Review of Conventional to Modern Algorithms of Speech Enhancement 633
M. Balasubrahmanyam, R. S. Valarmathi, and C. H. Mohan Sai Kumar

Gain Improvement of PIFA Antenna Using Glass Substrate in Comparison with FR4 Substrate 649
G. Kalpana and Bhuvaneshwari Balachander

Author Index 659

About the Editors

Rabindra Nath Shaw is currently working as Adjunct Professor, Chandigarh University, Chandigarh, India and also worked as Director, International Relations, Bharath Institute of Higher Education and Research (Deemed to be University), Chennai, India. Before joining BIHER he has served also Galgotias University as Director, IR&C, He is an alumnus of the Applied Physics department, University of Calcutta, India. He is a Senior Member of IEEE Industry Application Society, USA and Fellow of Nikhil Bharat Shiksha Parishad, India. Dr. Shaw is a global leader in organizing International conferences. His brand of world leading conference series includes IEEE International Conference on Computing, Power and Communication Technologies (GUCON), IEEE International Conference on Computing, Communication and Automation (ICCCA), IEEE IAS Global Conference on Emerging Technologies (GlobConET), International Conference on Electronics and Electrical Engineering (ICEEE), International Conference on Advances in Computing and Information Technology (ICACIT) etc. He holds the position of Conference Chair, Publication Chair, and Editor for these conferences. These Conferences are held in collaboration with various international universities like Aurel Vlaicu University of Arad, University of Malaya, University of Siena. Many world leaders are working with Dr. Shaw in these conferences. Most of these conferences are fully sponsored by IEEE Industry Applications Society, USA. He is also an expert in organizing International Seminars/Webinars/ Faculty Development Programme in collaboration with leading institutes across the world.

Pierluigi Siano (M'09–SM'14) received the M.Sc. degree in electronic engineering and the Ph.D. degree in information and electrical engineering from the University of Salerno, Salerno, Italy, in 2001 and 2006, respectively. He is a Professor and Scientific Director of the Smart Grids and Smart Cities Laboratory with the Department of Management and Innovation Systems, University of Salerno. Since 2021 he has been a Distinguished Visiting Professor in the Department of Electrical and Electronic Engineering Science, University of Johannesburg. His research activities are centered on demand response, energy management, the integration of distributed energy resources in smart grids, electricity markets, and planning and management

of power systems. In these research fields, he has co-authored more than 700 articles including more than 410 international journals that received in Scopus more than 18,000 citations with an H-index equal to 65. In the period 2019–2022 he has been awarded as a Highly Cited Researcher in Engineering by Web of Science Group. He has been the Chair of the IES TC on Smart Grids. He is Editor for the *Power and Energy Society Section of IEEE Access*, *IEEE Transactions on Power Systems*, *IEEE Transactions on Industrial Informatics*, *IEEE Transactions on Industrial Electronics*, *IEEE Systems*.

Saad Makhilef is an IEEE and IET Fellow. He is a Distinguished Professor at the School of Science, Computing and Engineering Technologies, Swinburne University of Technology, Melbourne, Australia, an Honorary Professor at the Department of Electrical Engineering, University of Malaya, and a distinguished visiting professor at the Institute of Sustainable Energy, Universiti Tenaga Nasional, Malaysia. He authored and co-authored more than 600 publications in academic journals and proceedings, five books with more than 46,000 citations, and more than 85 Ph.D. students who graduated under his supervision. He serves as an editorial board member for many top journals, such as *IEEE Transactions on Power Electronics*, *IEEE Open Journal of Industrial Electronics*, *IET Renewable Power Generation*, *E-Prime*, *Journal of Power Electronics*, and *International Journal of Circuit Theory and Applications*. His research interests include Power Conversion Techniques, Control of Power Converters, Maximum Power Point Tracking (MPPT), Renewable Energy, and Energy Efficiency.

Ankush Ghosh is Senior member of IEEE, Fellow of IETE currently working as Adjunct Professor, Chandigarh University, Chandigarh, India. He has received his Ph.D. (Engg.) degree from Jadavpur University, India in 2010. He was a research fellow of the Advanced Technology Cell-DRDO, Government of India. He was awarded National Scholarship by HRD, Government of India. He has outstanding research experiences and published 6 edited books; 4 from Springer and 2 from Elsevier; 3 National and 8 International patents and more than 120 research papers indexed in Scopus/Web of Science. He is serving as an editorial board member of several international journals including Chief Editor. He has more than 15 years of experience in teaching, research as well as industry. His UG and PG teaching assignments include Microprocessor and microcontroller, AI, IOT, Embedded and real time systems etc. He has delivered Keynote/Invited lecture in a number of international seminar/conferences, refreshers courses, and FDPs. He has guided a large number of M.Tech. and Ph.D. students. Dr. Ghosh is an active member of IEEE and organized a number Seminars and workshops in association with IEEE. He is an editor and organizing committee member of the Conference series GUCON, ICCCA, ICEEE, ICACIT. He is a He is a Start-up India Mentor and Global Startup Advisor of Wadhvani NEN. He has reviewed and mentored more than 50 start-ups. He has received award for contributing in Innovate India programme from AICTE-DST, Government of India in 2019 and 2020. He has received an appreciation award from

AICTE, DST, TI, IIMB, NSRCEL, and myGOV for fostering students to strengthen the ecosystem bridging Government, Academia, and Industry in the year 2021.

S. L. Shimi is currently working as Associate Professor at Punjab Engineering College (Deemed to be University), Chandigarh. She did her Postdoc from Luleå University of Technology, Sweden during 2019 to 2021. Her research Interests are Power Electronics and Drives, Digital Control, ANN and Fuzzy Logic Applications, FACTS, Renewable Energy, Soft Computing Techniques, MATLAB with dSpace Interface, MATLAB with Arduino Interface, MATLAB/SIMULINK and SimPowerSystem toolbox, Optimization techniques such as genetic algorithm and particle swarm optimization, Renewable etc.

Bidding Strategies for Generation Companies and Large Consumers in Carbon Emission Market Considering Electricity Spot Market Clearing Outcomes



Liyang Liu , Kun Wang , Junwei Ma, Honglin Xue, Xiaohui Wang, Wei Jiang, and Fushuan Wen 

Abstract The carbon emission trading market (referred to as carbon market) is an important platform to support the goals of carbon peaking and carbon neutrality in China. In this context, considering the existing electricity spot market mechanism, it is of great significance to study the bidding strategies of generation companies and large consumers in the carbon market. First, the overall framework for generation companies and large consumers to participate in the carbon and electricity markets is proposed. The carbon market clearing is discretized to implement collaborative clearing of the carbon and electricity markets. Secondly, according to the proposed carbon-electricity market framework, a centralized carbon market clearing model is established. Then, the bidding decision-making process of carbon market participants is addressed, and optimal bidding strategies for generation companies and large consumers participating in the carbon market considering the electricity spot market clearing outcomes are presented. Finally, the feasibility and effectiveness of the proposed method are demonstrated through a numerical example based on the IEEE 24-bus power system.

Keywords Carbon emission market · Bidding strategy · Electricity spot market · Generation company · Large consumer

L. Liu (✉) · K. Wang · F. Wen
College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China
e-mail: liyong.liu@hotmail.com

J. Ma · H. Xue
Information and Communication Branch, State Grid Shanxi Electric Power Company,
Taiyuan 030021, China

X. Wang
State Grid Economic and Technological Research Institute, Beijing 102209, China

W. Jiang
Mass Entrepreneurship and Innovation Center, State Grid Zhejiang Electric Power Co., Ltd.,
Hangzhou 310052, China

1 Introduction

With the proposal of the goal of carbon peaking and carbon neutrality, China is accelerating the construction of a carbon emission trading market (referred to as carbon market). On July 16, 2021, China national carbon market was launched simultaneously in Beijing, Shanghai, and Wuhan, marking that China has achieved a transition from regional pilot markets to the construction of the national carbon market [1, 2]. In general, China's carbon market is on the stage where the secondary carbon spot market is dominant, the regional markets and the national market coexist, and the power generation and consumption entities participate together. At the same time, the construction of China's electricity spot market has made a long-term progress, and many pilot areas have entered the trial operation stage of long-term electricity spot market settlement [3].

The carbon market and electricity spot market achieve mutual coupling through the bidding strategies of market participants [4]. Based on the electricity spot market clearing outcomes, research on the bidding strategies in the carbon market can not only provide reference for the participants, but also have great significance for the design of the carbon-electricity market mechanism.

However, there are few studies on modeling the bidding strategies of carbon market participants at present. Reference [5] analyzes the relationship between the carbon market and the electricity market, and explains how the carbon market affects the electricity market. However, no specific bidding model in carbon market has been given. A decision-making model considering multi-time coupling decisions in carbon market is proposed in [6], in which an equilibrium analysis model for the carbon-electricity market is established and the equilibrium results of power producers participating in the carbon-electricity market is presented. Reference [7] proposes a coordinated bidding model of hydropower-thermal power producers in the carbon-electricity coupling market considering the China certified emission reduction (CCER) mechanism, and describes the bidding behavior of hydropower and thermal power producers in the coupled market. However, the above publications rarely consider the bidding strategies of electricity consumers in the carbon market.

In addition, the discretization of the continuous carbon market trading is the key to studying the impact of electricity spot market trading on the bidding strategies of carbon market participants. Keeping the carbon market and the electricity spot market cleared at the same granularity will help power generation and consumption entities flexibly change their carbon market bidding strategies according to the electricity spot market clearing outcomes. At the same time, in the existing research, there is a lack of research and discussion on the continuous strategies of carbon market participants in multiple clearing periods within a compliance period. It is necessary to take the trading situation during the entire compliance period, including the expectation of future trading, into the consideration of the bidding model of carbon market participants, so as to be close to the actual trading situation of carbon market.

The contributions of this work are twofold:

- The overall framework of carbon-electricity market clearing is built and a discretization method of carbon market clearing is proposed.
- Bidding strategies for generation companies and large consumers in the carbon market are established considering electricity spot market clearing outcomes.

This paper is organized as follows: The overall framework of carbon-electricity market clearing and a discretization method of carbon market clearing are presented in Sect. 2. Section 3 establishes the carbon market clearing model. The bidding model in the carbon market is established in Sect. 4. A numerical example is analyzed in Sect. 5. Conclusions derived from this study are shown in Sect. 6.

2 Overall Framework for Carbon-Electricity Market

The carbon market described in this paper refers to the market for trading carbon emission allowances (CEA). Typically, carbon markets are divided into the primary carbon market and secondary carbon market. The primary carbon market is the market for issuing CEA, which are typically distributed through government allocation or auction. The secondary carbon market is a short-term spot trading market for the circulation of CEA, where the participants can freely buy and sell CEA based on their own carbon emissions situation [8]. The participants discussed in this paper include both power generation and power consumption entities.

The carbon market compliance period is usually divided into processes such as initial CEA allocation, free trading, and end-of-period settlement [9]. During the compliance period, the mismatch between the clearing periods of the carbon market and the electricity spot market creates significant difficulties for the participants to develop timely and effective bidding strategies. A discretization method is adopted in the carbon market to solve this problem. The compliance period of the carbon market is divided into multiple trading intervals with the same granularity as the clearing periods of electricity spot market. In each interval, CEA sellers and buyers can submit transaction requests to the trading system, and the system will conduct centralized clearing at the end of each interval. This processing method essentially changes the settlement of carbon market transactions from arbitrary moments to fixed times (end of each trading interval), while the participants can still submit or modify transaction requests at any time, so the carbon market clearing outcomes remain unchanged.

In summary, the clearing process of carbon-electricity market is illustrated in Fig. 1. In a typical carbon-electricity market clearing interval, the participants first obtain bidding strategies for the electricity spot market that take into account carbon emission costs based on the traditional electricity spot market bidding model. In the early stage of this trading interval, the electricity spot market is cleared, and the clearing outcomes provide a basis for the participants to formulate carbon market bidding strategies. Then, the participants generate and update their own carbon market bidding strategies and submit transaction requests to the CEA trading system. The carbon market completes centralized clearing at the end of the trading interval.

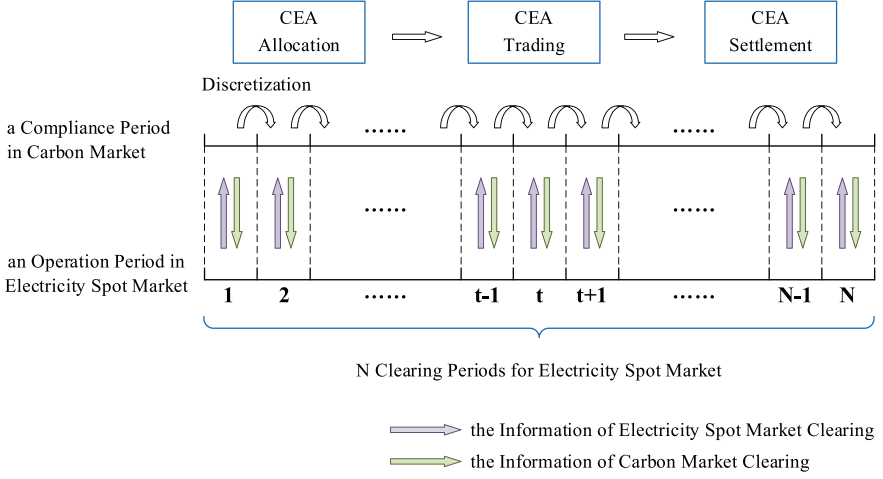


Fig. 1 The framework of carbon emission trading considering spot electricity trades

Note that the bidding model for electricity spot market that considers the results of CEA trading is already mature. Therefore, this paper adopts the bilateral bidding model in the electricity spot market proposed in [10] that considers the carbon emission costs for both power generation and consumption entities, and the model will not be described in detail in this paper. The focus of this paper is on how to determine the bidding strategies for carbon market participants under the condition that the electricity spot market clearing outcomes have been established.

3 Carbon Market Clearing

The carbon market clearing process is represented using a centralized clearing method, and the clearing model for the carbon market in trading interval t is shown below. The objective function (1) represents the maximization of social welfare (minimization form). Constraints (2) and (3) are the upper and lower bounds constraints of the bid quantities of CEA sellers and buyers, and constraint (4) represents the CEA supply–demand balance.

$$\min \sum_n \pi_{n,t}^{\text{bid},s} Q_{n,t}^{\text{CET},s} \Delta t - \sum_m \pi_{m,t}^{\text{bid},b} Q_{m,t}^{\text{CET},b} \Delta t \quad (1)$$

subject to

$$0 \leq Q_{n,t}^{\text{CET},s} \leq Q_{n,t}^{\text{CET},\text{bid},s} \quad \forall n \quad (2)$$

$$0 \leq Q_{m,t}^{\text{CET},b} \leq Q_{m,t}^{\text{CET},\text{bid},b} \quad \forall m \quad (3)$$

$$\sum_n Q_{n,t}^{\text{CET},s} = \sum_m Q_{m,t}^{\text{CET},b} \quad (4)$$

where $\pi_{n,t}^{\text{bid},s}$ and $\pi_{m,t}^{\text{bid},b}$ are the offer price of CEA seller n and bid price of CEA buyer m in trading interval t , respectively; $Q_{n,t}^{\text{CET},s}$ and $Q_{n,t}^{\text{CET},\text{bid},s}$ are the CEA sold by CEA seller n and offer quantity of CEA seller n in trading interval t , respectively; $Q_{m,t}^{\text{CET},b}$ and $Q_{m,t}^{\text{CET},\text{bid},b}$ are the CEA purchased by CEA buyer m and bid quantity of CEA buyer m in trading interval t , respectively.

4 Bidding Model

4.1 Bid Quantity

The calculation method of the CEA bid quantities of the participants in the carbon market is shown in (5). The first term represents the plan usage quantity of CEA of the market participant established at the beginning of the compliance period, which is represented by the CEA decomposition factor scalar, indicating the ideal carbon emissions of the market participant in trading interval t . The second term represents the actual usage quantity of CEA of the market participant in trading interval t , measured by the power generated or consumed by the market participant in the electricity spot market. The difference between the first and second terms reflects the deviation of the CEA plan amount and actual amount of the market participant in trading interval t . The third term represents the net accumulated quantity of CEA of the market participant in the previous trading intervals, reflecting the remaining CEA before trading interval t , and it can be calculated by (6). Since the CEA bid quantity has transferability throughout the entire compliance period, the participants do not necessarily have to complete CEA transactions immediately when CEA deviations occur. Therefore, the accumulation of CEA is more likely to occur.

$$Q_{o,t}^{\text{CET},\text{bid}} = \omega_{o,t} \cdot Q_{o,T}^{\text{CET}} - P_{o,t} \cdot e_{o,t} + Q_{o,t-1}^{\text{CET},\text{net}} \quad (5)$$

$$Q_{o,t}^{\text{CET},\text{net}} = Q_{o,t}^{\text{CET},\text{bid}} - Q_{o,t}^{\text{CET},\text{acl}} \quad (6)$$

where $Q_{o,t}^{\text{CET},\text{bid}}$, $Q_{o,t}^{\text{CET},\text{acl}}$ and $Q_{o,t}^{\text{CET},\text{net}}$ are the bid or offer quantity, actual traded CEA and net accumulated CEA of market participant o in trading interval t , respectively; $P_{o,t}$ is the power generated or consumed by market participant o in trading interval t ; $e_{o,t}$ is the carbon emission intensity of market participant o in trading interval t ; $Q_{o,T}^{\text{CET}}$ is the initial CEA of market participant o in the entire compliance period; $\omega_{o,t}$ is the CEA decomposition factor of market participant o in trading interval t .

It should be noted that the above CEA bid quantity and net accumulated quantity may not necessarily be positive values. For the CEA bid quantity, a positive value

indicates that there is a surplus of CEA in the period, and bidding in the carbon market system means selling CEA. On the contrary, a negative value indicates that there is a deficit of CEA in the period and corresponding CEA need to be purchased in the carbon market. Similarly, for the net accumulated quantity of CEA, a positive value indicates that there is a surplus of CEA, and a negative value indicates that there is a deficit of CEA. Note that the above surplus and deficit are all based on the plan quantity.

4.2 Bid Price

In the process of carbon market trading, the participants aim to maximize their own profits and decide whether to engage in CEA trading during the current trading interval. The pricing problem for each market participant can be described as setting a bid price or offer price that meets the expectation of the market participant for the current trading interval. Take a market participant as an example: if the bid quantity of CEA is traded at this price, the market participant's profit has been satisfied, so it does not need to choose to trade in future periods. Otherwise, the market participant will consider trading that portion of CEA in future periods based on its assessment of the market. Note that the price of CEA varies under the complex and random market environment, so the bid prices of all the participants will be continuously updated as the trading proceeds.

Due to the different market factors considered by the buyers and sellers, it is necessary to discuss the pricing strategies for CEA buyers and sellers separately. Taking the trading interval t as an example, the models are as follows.

4.2.1 Bid Prices of Buyers

For CEA buyers whose $Q_{o,t}^{\text{CET,bid}} < 0$, the bid prices can be formulated as follows:

$$\pi_{m,t}^{\text{bid},b} = \min\left(\theta_{m,t}^b \rho_{m,t}^b \pi_{m,t,\min}^f, c_m^e, c_m^p\right) \forall m \quad (7)$$

where $\pi_{m,t,\min}^f$ is the minimum predicted carbon price after trading interval t of CEA buyer m ; c_m^e and c_m^p are the marginal carbon abatement cost and penalty for non-compliance of CEA buyer m , respectively; $\min(x, y, z)$ represents the minimum value of x , y , and z . Equation (7) means that the bid price of the CEA buyer is the minimum value of the minimum carbon price after trading interval t , the marginal carbon abatement cost, and the penalty for non-compliance. However, CEA buyers also need to consider their own compliance pressure and the supply and demand relationship in the carbon market. Therefore, coefficient $\theta_{m,t}^b$ and $\rho_{m,t}^b$ are added in the first term of (7). $\theta_{m,t}^b$ represents the compliance pressure coefficient for CEA buyer m in trading interval t . As the compliance deadline approaches and the CEA deficit

increases, the compliance pressure on the CEA buyer increases. $\rho_{m,t}^b$ represents the excess demand coefficient for CEA buyer m in trading interval t . When there is excess demand in the carbon market, the CEA buyer should raise the bid price accordingly. $\theta_{m,t}^b$ and $\rho_{m,t}^b$ can be calculated as follows:

$$\theta_{m,t}^b = a \max\left(0, \frac{-Q_{m,t}^{\text{CET,net}}}{Q_m^{tva}} - Q_m^{tva}\right) \forall m \quad (8)$$

$$\rho_{m,t}^b = b \max\left(0, Q_t^{\text{CET,bs,net}} - Q_m^{tva}\right) \forall m \quad (9)$$

where a is the base value of compliance pressure coefficient and $a > 1$; b is the base value of excess demand coefficient and $b > 1$; $Q_t^{\text{CET,bs,net}}$ is the excess demand of CEA in trading interval t , which is the difference between the total bid quantity and total offer quantity of CEA in trading interval t ; Q_m^{tva} and Q_m^{tva} are the compliance pressure threshold and excess demand threshold of CEA buyer m , respectively. The CEA buyer will increase the bid price only if the compliance pressure and excess demand both exceed the corresponding thresholds, otherwise $\theta_{m,t}^b$ and $\rho_{m,t}^b$ are both set to 1.

$\pi_{m,t,\min}^f$ is formulated as:

$$\pi_{m,t,\min}^f = \arg \max_{\pi_{m,\tau}^f \in [t, T]} \left(-(\pi_{m,\tau}^f)^r \cdot p_{m,\tau} \right) \forall m \quad (10)$$

where r is the risk aversion coefficient of the CEA buyer to solve the problem that there may be a contradiction between the predicted carbon price and the prediction accuracy, which is set to 1 by default. The larger the value of r , the stronger the buyer's risk-bearing capacity; $\pi_{m,\tau}^f$ is the predicted carbon price in trading interval τ of CEA buyer m , and $p_{m,\tau}$ is the corresponding prediction accuracy.

4.2.2 Offer Prices of Sellers

For CEA sellers whose $Q_{o,t}^{\text{CET,bid}} > 0$, the offer prices can be formulated as follows:

$$\pi_{n,t}^{\text{bid},s} = \theta_{n,t}^s \rho_{n,t}^s \pi_{n,t,\max}^f \forall n \quad (11)$$

where $\pi_{n,t,\max}^f$ is the maximum predicted carbon price after trading interval t of CEA seller n . CEA sellers do not need to consider the marginal carbon abatement cost and the penalty for non-compliance. Similar to CEA buyers, CEA sellers are affected by the sales pressure of CEA and excess supply in the carbon market. Therefore, coefficient $\theta_{n,t}^s$ and $\rho_{n,t}^s$ are added, and they can be calculated as follows:

$$\theta_{n,t}^s = (1 - c Q_t^{\text{CET},bs,\text{net}})^{\max\left(0, \frac{Q_{n,t}^{\text{CET},\text{net}}}{T-t} - Q_n^{\text{tvc}}\right)} \forall n \quad (12)$$

$$\rho_{n,t}^s = d^{\max\left(0, -Q_t^{\text{CET},bs,\text{net}} - Q_n^{\text{tvd}}\right)} \forall n \quad (13)$$

where $\theta_{n,t}^s$ is the sales pressure coefficient of CEA seller n in trading interval t . As the compliance period approaches, CEA sellers will adjust their offer prices based on market supply and demand. $\rho_{n,t}^s$ is the excess supply coefficient of CEA seller n in trading interval t . When there is excess supply in the carbon market, CEA sellers should lower their offer prices accordingly. c is the base value of CEA sales pressure coefficient and $c > 0$; d is the base value of excess supply coefficient and $d < 1$. Q_n^{tvc} and Q_n^{tvd} are the sales pressure threshold and excess supply threshold of CEA seller n , respectively.

Similar to CEA buyers, CEA sellers also face a contradiction between the predicted carbon price and the accuracy of the prediction, which can be solved as that of CEA buyers. The formulation of $\pi_{n,t,\text{max}}^f$ is as follows:

$$\pi_{n,t,\text{max}}^f = \arg \max_{\pi_{n,t}^f \in [t,T]} \left((\pi_{n,\tau}^f)^r \cdot p_{n,\tau} \right) \forall n \quad (14)$$

where $\pi_{n,\tau}^f$ is the predicted carbon price in trading interval τ of CEA seller n , and $p_{n,\tau}$ is the corresponding prediction accuracy.

5 Case Study

5.1 Data Specifications

The IEEE 24-bus power system is employed to construct a numerical example to demonstrate the effectiveness of the proposed method and model. Eight generation companies and four large consumers participate in both the electricity spot market and carbon market simultaneously. The simulation parameters for each market participant are shown in Table 1. For the demonstration of the numerical example results, we assume that a compliance period consists of 50 carbon-electricity market trading intervals. The power demand profile for the compliance period is generated using the actual industry load curve.

Table 1 Setting of simulation parameters of generation companies and large consumers

Number	Type	Capacity (MW)	Initial CEA (t)	Emission intensity (t/MWh)
1	Coal-fired unit	130	3959.4621	0.9232
2	Coal-fired unit	350	9702.0771	0.8547
3	Coal-fired unit	430	11,707.9226	0.8696
4	Coal-fired unit	660	16,682.7778	0.8761
5	Coal-fired unit	800	18,897.2694	0.8939
6	Gas-fired unit	250	6533.7352	0.3793
7	Gas-fired unit	400	9622.8374	0.4058
8	Gas-fired unit	630	14,426.9462	0.4182
9	Power consumer	/	22,534.6495	0.7812
10	Power consumer	/	27,764.7259	0.8165
11	Power consumer	/	16,263.9231	0.6621
12	Power consumer	/	29,444.9988	0.7579

5.2 Analysis of Single Compliance Period Results

To analyze the carbon market clearing outcomes and market behaviors of generation companies and large consumers within a single compliance period, and with the total issuance of CEA as the boundary condition, the following cases are set: (1) Case 1 increases the total CEA issuance by 10%; (2) Case 2 keeps the total CEA issuance unchanged; (3) Case 3 decreases the total CEA issuance by 10%.

The carbon price and emissions variations within a single compliance period under three different cases are presented in Fig. 2. Generally, the trends of carbon price in the three cases are similar, characterized by an initial decrease, followed by a relatively stable period, and a continuous rise with significant fluctuations towards the end of the compliance period. Comparing the data from different cases, it is evident that a reduction in the total issuance of CEA increases the carbon price. Additionally, there are significant differences in carbon price changes at the end of the compliance period in different cases. The carbon price in case 1 shows a downward trend, while that in case 3 continues to rise. This is because the supply–demand relationship of CEA at the end of the compliance period varies under different cases. In case 1, there is excess supply of CEA, so the carbon price turns down. In contrast, the carbon market in case 3 faces excess demand, resulting a price hike.

The profit of different generation companies under different cases is shown in Fig. 3. As the total issuance of CEA decreases, the carbon emission cost of generation companies continues to increase, forcing them to either reduce their electricity generation or increase their bid prices in the electricity spot market, ultimately leading to a decrease in their profits. However, for gas-fired generation companies with lower carbon emission intensity, their profits are less affected. When CEA are scarce, cleaner generation companies will accumulate more CEA, which will help them

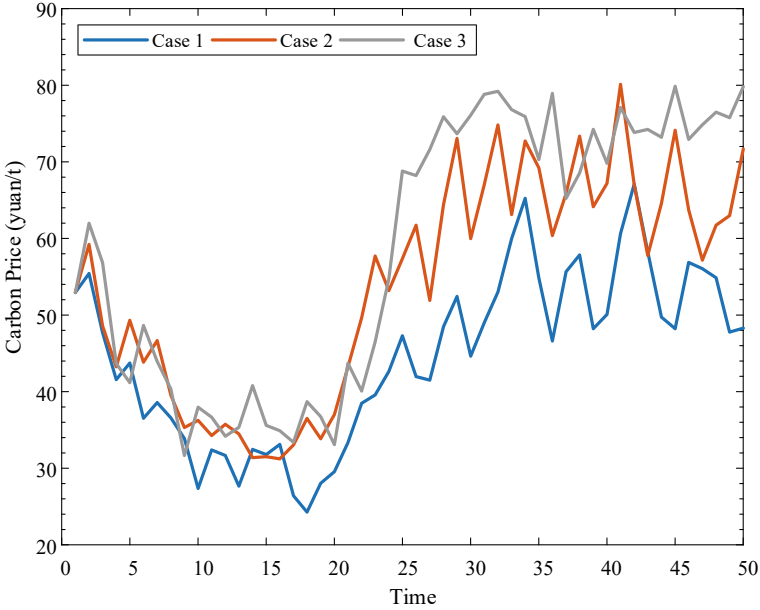


Fig. 2 Variation curve of carbon price under different cases in a single compliance period

gain a larger share of the electricity spot market, greatly reducing the impact of the reduction in CEA issuance on this part of the participants.

5.3 Analysis of Multiple Compliance Periods Results

The effectiveness of the proposed method and model for multiple compliance periods is verified through the simulation analysis of 200 carbon-electricity trading intervals (i.e., four compliance periods). The results are shown in Fig. 4.

In the first compliance period, the carbon price showed a trend of initial decreasing and then rising. At the beginning of the compliance period, the CEA are sufficient, so the transaction volume of CEA is small. Near the end of the compliance period, under the pressure of compliance, CEA buyers raised their bid prices, causing the carbon price to rise sharply and the transaction volume to surge.

In the second compliance period, based on the historical data of the surge of carbon price at the end of the previous compliance period, the participants purchase a large amount of CEA when the carbon price is relatively low at the beginning of the compliance period, causing the carbon price to rise quickly and the transaction volume to increase. Afterward, the carbon price remains at a relatively high level, while the transaction volume is relatively small. This is because the high carbon price hit the trading enthusiasm of CEA buyers, causing them to choose to appropriately

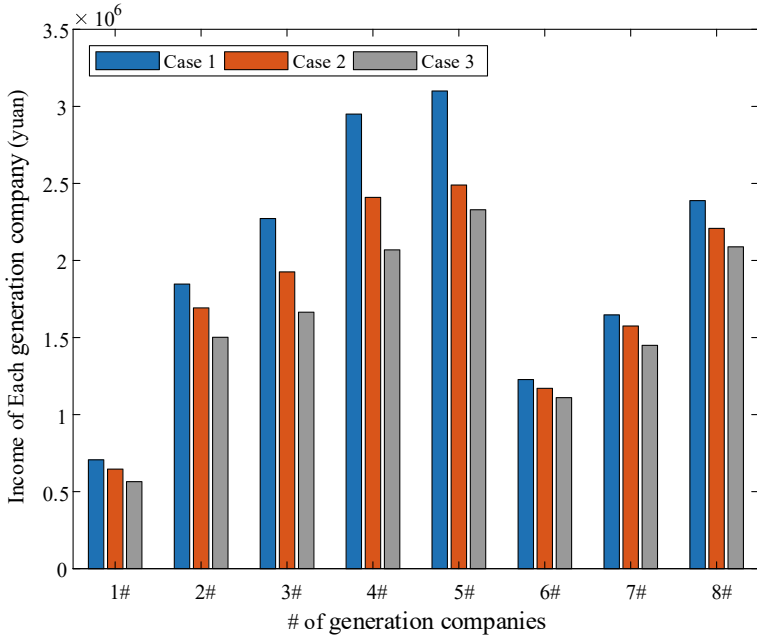


Fig. 3 Income of each generation company under different cases in a single compliance period

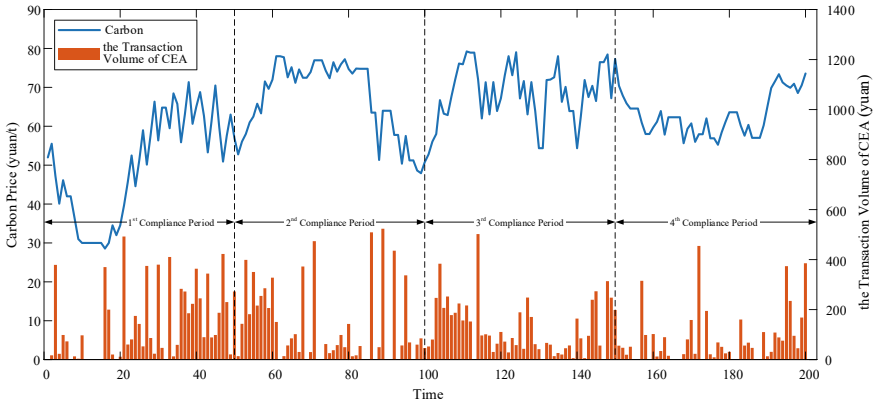


Fig. 4 Carbon market clearing outcomes in multiple compliance periods

reduce their power generation or production scale to avoid compliance risk. At the end of the compliance period, the carbon price drops sharply because the demand for CEA of buyers who reduce their power generation or production scale decreased significantly, which results in a surplus of CEA supplied by sellers who have to lower their offer prices to sell their hoarded CEA.

In the third compliance period, the carbon price fluctuates significantly, and the participants have different expectations for the future carbon price. The market transaction volume is randomly distributed during this period. At the end of the compliance period, the carbon price dropped again, and the purchasing enthusiasm of CEA buyers are not high.

At the beginning of the fourth compliance period, the carbon price continues the trend at the end of the third compliance period, fluctuating around 60 yuan/ton. At the end of the compliance period, the demand for CEA increases, causing the carbon price to rise. The transaction volume of this compliance period is mainly concentrated in the end.

6 Conclusion

Taking into account the current rules of the electricity spot market, developing scientific and reasonable bidding strategies for carbon market participants is an important issue in the construction and development of the carbon market. In this paper, the overall framework of generation companies and large consumers participating in the carbon-electricity market is introduced firstly. The discretization method of carbon market clearing is adopted to divide the compliance period into multiple trading intervals with the same granularity as the electricity spot market clearing, thus achieving coordinated clearing of the carbon-electricity market. Then, the bidding decision-making process of carbon market participants are introduced, focusing on the bidding strategies of carbon market participants considering the electricity spot market clearing outcomes. Finally, through a numerical example, we provide analysis and demonstration of the effectiveness of the proposed method and model from the perspectives of single and multiple compliance periods, and present the profit situations of each market participant. This paper provides useful reference for further research on the bidding strategies of market participants and the design and development of the carbon-electricity market mechanism.

Acknowledgements This work is financially supported by a Science and Technology Project of the State Grid Corporation of China entitled “Research on Quantification of Carbon Reduction Investment in Power Grid Projects and Cost Grooming Technology Oriented by Optimal Allocation of Resources” (Project No.: 5108-202218280A-2-425-XG; Contract No.: SGSXXT00JFJS2250157).

References

1. The world’s largest carbon emissions trading market will be unveiled, and the first 2225 electric power companies will open accounts. http://www.ce.cn/xwzx/gnsz/gdxw/202103/30/t20210330_36422393.shtml
2. Progress, problems and policy suggestions on the construction of national carbon market. <http://www.sic.gov.cn/News/455/11242.htm>

3. Song Y, Bao M, Ding Y, Shao C, Shang N (2020) Review of Chinese electricity spot market key issues and its suggestions under the new round of Chinese power system reform. *Proc CSEE* 40(10):3172–87
4. Chen H (2021) Electricity value analysis and market mechanism design under carbon-neutral goal. *Power Gener Technol* 42(02):141–150
5. Shang N, Chen Z, Lu Z, Leng Y (2023) Interaction principle and cohesive mechanism of electricity market, carbon market and green power certificate market. *Power Syst Technol* 47:142–154
6. Sun X, Ding Y, Bao M, Guo C, Liang Z, Ye C. Carbon-electricity market equilibrium analysis considering multi-time coupling decisions of power producers. *Autom Electr Power Syst*. <https://doi.org/10.7500/AEPS20220401011>
7. Zhao L, Li Y, Jin X, Cheng C. Coordinated bidding model of hydropower-thermal power in carbon-electricity coupling market considering CCER mechanism. *Autom Electr Power Syst*. <https://doi.org/10.7500/AEPS20220606002>
8. Stavins RN (2008) A meaningful US cap-and-trade system to address climate change. *SSRN Electron J* 32:293
9. Zhang X, Zhou L (2019) National carbon emissions trading system practice manual. Institute of Energy and Environmental Economics, Tsinghua University, Beijing
10. Chen L, Zhou Y, Fang C (2019) Bidding game between power generation companies and consumers considering carbon trade. *Proc CSU-ESPA* 31(10):66–72

NSTLBO-Based Approach for Optimal Scheduling of Hydrothermal Generating Units in Regulated Environment



Baburao Pasupulati, Bhargava Reddy Sibbala, S. Sivakumar,
S. Amosedinakaran, and Rajakumar Palanisamy

Abstract This research paper benevolence is a technique for short-term hydrothermal generation scheduling (STHTGS) power plant using a non-dominant sorting teaching learning-based optimization (NSTLBO) algorithm. It involves the deployment of thermal power plants in optimum operating conditions to reduce fuel costs and optimize the cost of hydroelectric power plants. The electrical energy considered in this study is assumed to be efficient. The NSTLBO algorithm has been found suitable for this problem as it reaches the minimum cost in the shortest time compared with the previous methods.

Keywords Hydro · Thermal · Economic dispatch · Valve point loading effect · NSTLBO algorithm

1 Introduction

Hydrothermal scheduling will optimize the timing of hydroelectric and thermal power plants toward diminishing the fuel costs of thermal power plants [1]. It is an important part of the energy industry and economy that provides electrical and heat

B. Pasupulati (✉) · S. Sivakumar · S. Amosedinakaran · R. Palanisamy
Department of Electrical and Electronics Engineering, Vel Tech Rangarajan Dr, Sagunthala R&D
Institute of Science and Technology, Avadi, Chennai, Tamil Nadu 600062, India
e-mail: drbaburao@veltech.edu.in

S. Sivakumar
e-mail: ssivakumar@veltech.edu.in

S. Amosedinakaran
e-mail: dramose@veltech.edu.in

R. Palanisamy
e-mail: drrajakumarp@veltech.edu.in

B. R. Sibbala
Department of Electrical and Electronics Engineering, Srinivasa Ramanujan Institute of
Technology, Andhra Pradesh, India
e-mail: bhargav.s204@gmail.com

energy within this system [2, 3]. The increasing cost of thermal power plants and the intersection of renewable energy further emphasize the importance of hydrothermal propulsion [4].

In the short-term transmission problem, some limitations such as the capacity of the hydraulic unit, the demand of the load, the hydraulic input, the flow restriction of the reservoir and the reservoir capacity should be known [5]. The stability of a hydroelectric power plant depends on the balance between thermal and hydroelectric power generation and load demand. However, generators are difficult to operate and deliver in industry due to hydraulic constraints and the need to meet load requirements [6]. Many techniques, such as Lagrangian multipliers, gradient search methods, evolutionary programming, rapid evolutionary programming, mixed evolutionary programming, simulated annealing, genetic algorithms and particle swarm optimization have been used to solve short-term problems [7–14].

However, the aforementioned algorithms all have their own limitations. For example, the Lagrange multiplier method will face the problem that the binary solution is not possible, the simulated annealing convergence speed is slow, and the problem is not easy, and the evolutionary algorithm will face the problem of slow convergence speed in multimodal optimization. Additionally, genetic algorithms may have poor search results [15], while optimization of the particle swarm may suffer from premature intersections [10, 16–19]. To solve these problems in short-term hydrothermal scheduling, this article proposes to use the non-dominant sequence-based learning-based optimization (NSTLBO) algorithm [20–22]. This optimization process was inspired by the teaching-to-learn behavior and outperforms the advanced know-how, especially in the execution period.

This method differs from other algorithms in that it does not rely on standard optimization parameters. The algorithm reduces the costs associated with thermal power plants by dividing energy consumption by hydrothermal generation. To evaluate the effectiveness of the NSTLBO method, we apply it to extensive experiments and compare the simulation results with those obtained by further approaches.

2 Problematic Construction

2.1 The Cost Minimization

$$F_1 = \sum_{t=1}^T \left\{ \sum_{i=1}^{j_s} f_s(P_s) \right\} \quad (1)$$

In the formulary, F_1 is the total operating cost of the thermal power unit.

The number of electric generators is j_s , respectively.

T is the timer for the generator. The productivity power of the thermal unit is P_s , respectively.

After delightful into account the valve point load effect, the fuel cost of the thermal unit at time T is shown as follows:

$$f_{it}(P_{sit}) = \min \sum_{t=1}^T \sum_{i=1}^{j_s} \{a_i + b_i P_{si} + c_i P_{si}^2 + |d_i \sin[e_i(P_{si}^{\min} - P_{si})]|\} \quad (2)$$

where a_i, b_i, c_i, d_i and e_i are the cost coefficients of the thermal unit and P_{si}^{\min} is the slightest power generation limit of the thermal unit.

2.2 Constraints

The STHTGS problem would gratify the subsequent equality and inequality.

Equal to the Energy Limit

$$\sum_{i=1}^{j_s} P_{si} + \sum_{g=1}^{r_h} P_{hg} + \sum_{k=1}^{l_w} P_{wk} + \sum_{m=1}^{n_{pvm}} P_{pv} - P_L = P_D \quad (3)$$

Power generation of hydro entities can be expressed as

$$P_{hg} = C_{1g}(V_{hg})^2 + C_{2g}(Q_{hg})^2 + C_{3g}V_{hg}Q_{hg} + C_{4g}V_{hg} + C_{5g}Q_{hg} + C_{6g} \quad g \in N_h, t \in T \quad (4)$$

Water Balance Constraint

$$V_{hj}^t = V_{hj}^{t-1} + I_{hj}^t - Q_{hj}^t - S_{hj}^t + \sum_{l=1}^{R_{uj}} (Q_{hl}^{t-d_{lj}} + S_{hl}^{t-d_{lj}}) \quad j \in N_h, t \in T \quad (5)$$

Limited Storage Capacity and Initial and Final Discharge Rate of the Reservoir

$$V_{hj}^{\min} \leq V_{hj}^t \leq V_{hj}^{\max} \quad j \in N_h, t \in T \quad (6)$$

$$Q_{hj}^{\min} \leq Q_{hj}^t \leq Q_{hj}^{\max} \quad j \in N_h, t \in T \quad (7)$$

Power Generation Limits

$$P_{hj}^{\min} \leq P_{hj}^t \leq P_{hj}^{\max} \quad j \in N_h, t \in T \quad (8)$$

$$P_{si}^{\min} \leq P_{si}^t \leq P_{si}^{\max} \quad i \in N_s, t \in T \quad (9)$$

3 Elucidation Procedure

3.1 Non-dominated Sorting TLBO Algorithm

The NSTLBO procedure is an improvement of the TLBO procedure, which provides a unique way to generate Pareto optimal results for multi-objective optimization difficulties. Similar to the TLBO algorithm, it uses a grading algorithm combined with the teacher–student level to manage multiple objectives. The NSTLBO algorithm uses a non-critical permutation technique and mass distance measurement to more efficiently search for space and continuously select the optimal solution along the Pareto front. With its bottleneck-free working time, it enables teachers to be selected from a wide search area and prevents premature convergence to the best locale.

The NSTLBO algorithm combines the teacher’s instruction and the learning level of the TLBO algorithm, so the student can use it quickly. Finding good solutions to individual optimization problems is easy, but it becomes more difficult when there are multiple conflicting goals. In this case, finding the best solution in the problem solving process is not an easy task. The algorithm solves this problem by comparing the solution sequences based on the congestion distance values and the non-dominant strategies to novelty the optimum solution. The process starts with startup.

The algorithm starts with an initialization step that creates an $N \times D$ matrix containing the values generated in the search space, where N represents the total size (often called ‘room size’) and D represents the size of the problem being solved given all parameters. The algorithm remains designed toward work aimed at ‘ g ’ iterations. At the beginning of each iteration, the value of the j th parameter of the i th vector is calculated using the following equation:

$$x_{(i,j)}^1 = x_j^{\min} + \text{rand}_{(i,j)} \times (x_j^{\max} - x_j^{\min}) \quad (10)$$

where $\text{rand}_{(i,j)}$ signifies a consistently disseminated arbitrary variable within the limit (0,1). The workings of the i th vector for the generation ‘ g ’ is shown by

$$X_i^g = \left[x_{(i,1)}^g, x_{(i,2)}^g, \dots, x_{(i,j)}^g, \dots, x_{(i,D)}^g \right] \quad (11)$$

In this two-objective problem, the line vector represents the main target for a generation. Line vectors correspond to two objective functions in thought. The two-objective problem defined as (a and b) can be framed as follows.

$$\begin{bmatrix} Y_{a_i^g} \\ Y_{b_i^g} \end{bmatrix} = \begin{bmatrix} fa(X_{(i)}^g) \\ fb(X_{(i)}^g) \end{bmatrix} \quad (12)$$

where $i = 1, 2, \dots, N; j = 1, 2, \dots, D$ and $g = 1, 2, \dots, G$.

Teacher Phase

The mean trajectory is computed by taking the average of the learners' values for each subject in the class. So the mean vector μ is shown as

$$M^g = \begin{bmatrix} \text{mean}([x_{(1,1)}^g, \dots, x_{(i,1)}^g, \dots, x_{(N,1)}^g]) \\ \text{mean}([x_{(1,j)}^g, \dots, x_{(i,j)}^g, \dots, x_{(N,j)}^g]) \\ \text{mean}([x_{(1,D)}^g, \dots, x_{(i,D)}^g, \dots, x_{(N,D)}^g]) \end{bmatrix}^T \quad (13)$$

Then

$$M^g = m_1^g, m_2^g, \dots, m_j^g, \dots, m_D^g \quad (14)$$

The vector with the lowest objective value is determined as the best vector and is chosen by way of the teacher for this recapitulation. The algorithm makes development by replacing the student's average by that of the teacher. This is done by combining the current mean vector with the potential mean vector in the student population, resulting in improved student level.

$$X_{\text{new}(i)}^g = X_{(i)}^g + \text{rand}^g \times (X_{\text{Teacher}}^g - T_F M^g) \quad (15)$$

Henceforth T_F is the teaching inspiration in the course of recapitulation which may be either 1 or 2.

The supplementary expert students in the matrix X_{new} relocate the inferior learners in matrix S by the non-dominated sorting algorithm.

Learner Phase

This stage is enthusiastic to interface between learners. The repetition of interface leads to the improvement of learner's expertise. Each learner works randomly with other learners, speeding up knowledge sharing. A precise student ($X_{(i)}^g$) and the other learner ($X_{(r)}^g$) has remained arbitrarily chosen ($i \neq r$). Lastly the i th vector of the matrix X_{new} in the learner phase seems

$$X_{\text{new}(i)}^g = \begin{cases} X_{(i)}^g + \text{rand}_{(i)}^g \times (X_{(i)}^g - X_{(r)}^g) & \text{if } (Y_i^g < Y_r^g) \\ X_{(i)}^g + \text{rand}_{(i)}^g \times (X_{(r)}^g - X_{(i)}^g) & \text{otherwise} \end{cases} \quad (16)$$

There is an opportunity of manifold X_{new} conditions in the learner phase. Consequently cutting-edge circumstance of a bi-objective problem of the presentation of learner phase might have preparation as

$$X_{\text{new}(i)}^g = \begin{cases} X_{(i)}^g + \text{rand}_{(i)}^g \times (X_{(i)}^g - X_{(r)}^g) & \text{if } (Y a_i^g < Y a_r^g) \\ X_{(i)}^g + \text{rand}_{(i)}^g \times (X_{(r)}^g - X_{(i)}^g) & \text{otherwise} \end{cases} \quad (17)$$

$$X_{new(i)}^g = \left\{ \begin{array}{l} X_{(i)}^g + \text{rand}_{(i)}^g \times (X_{(i)}^g - X_{(r)}^g) \text{ if } (Yb_i^g < Yb_r^g) \\ X_{(i)}^g + \text{rand}_{(i)}^g \times (X_{(r)}^g - X_{(i)}^g) \text{ otherwise} \end{array} \right\} \quad (18)$$

Lastly, the X matrix and the X_{new} matrices are treated in organized manner in the NSTLBO, which devise the ‘ N ’ best learners for the confirming iteration. The algorithm will be finished afterward ‘ G ’ number of iteration is over.

Fuzzy Membership Function

The main goal of the systems cause is to resolve the conflict by fulfilling the constraints. In many cases, the results, limitations and benefits of the proposed methods cannot be accurately predicted. Most bugs are unreachable. This may be because of confusion, inaccurate or unclear information. When we look at the decision-making process, we see that they can replace all their business goals with vague or negative ones. Fuzzy crowds are determined by equations called membership. These properties are allocated standards between 0 and 1. Through setting least and extreme performance targets and cost of ownership, the decision-maker has to make a decision. The membership function $\mu(j_i)$ in a constructive manner.

It remains measured that $\mu(j_g)$ occurred to be a linear declining and unremitting purpose and is expressed as

$$\mu(j_g) = \left\{ \begin{array}{l} 1 \quad j_g \leq j_g^{\min} \\ \frac{j_g^{\max} - j_g}{j_g^{\max} - j_g^{\min}} \quad j_g^{\min} \leq j_g \leq j_g^{\max} \\ 0 \quad j_g \geq j_g^{\max} \end{array} \right. \quad (g = 1, 2, \dots, N_{obj}) \quad (19)$$

where j_g^{\min} and j_g^{\max} remain the least and extreme standards of impartial role anywhere in the solution to be property-owning.

N_{obj} denotes the number of impartial purpose in the problem.

Regularized association values μ^k for each non-dominated resolution is intended by the subsequent equation.

$$\mu^k = \frac{\sum_{i=1}^{N_{obj}} \mu_i^k}{\sum_{k=1}^{M_{nds}} \sum_{i=1}^{N_{obj}} \mu_i^k} \quad (20)$$

where M_{nds} remains the amount of non-dominated solutions. Indicate the best contain explanation that is consuming the utmost value of μ^k .

4 Implementation

Figure 1 shows the flowchart for STHTGS problem and implementation of NSTLBO algorithm.

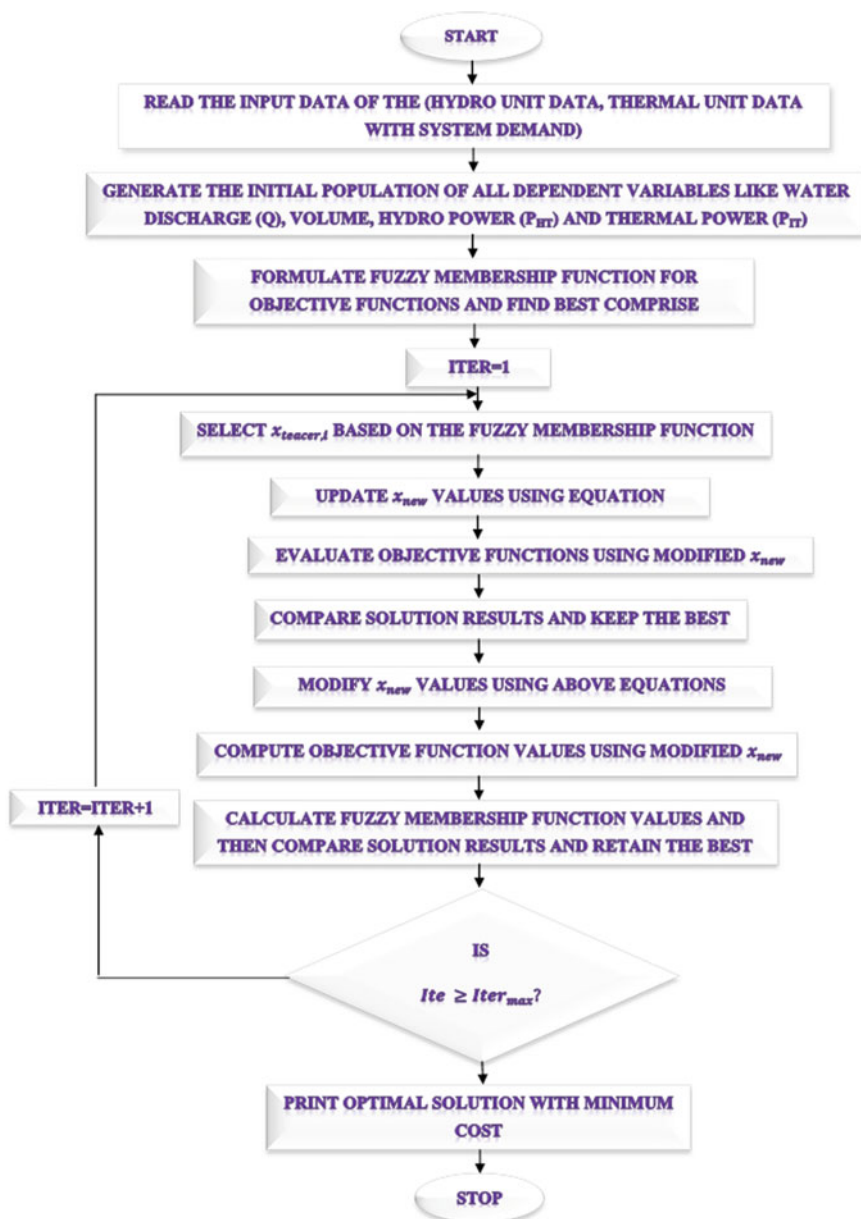


Fig. 1 Implementation of NSTLBO for SHTGS problem

5 Results and Discussion

Apply the NSTLBO algorithm to the test procedure, confirm the viability and efficacy of the NSTLBO algorithm and elucidate the STHTGS problem. The test system consists of four hydro generators and three thermal generators to achieve the best solution of the STHTGS problem, with and without the valve loading effect. The main objective of the process is the operating cost of the electric generator, captivating into explanation the effect of the valve loading point. Also, the hydraulic network of these machines is shown in Fig. 2. Total scheduling time is 24 h. The test procedure is explained in detail and the results are described below.

Test System

In this system, the best solution of the STHTGS problem is obtained by using the NSTLBO algorithm, taking into account the valve point load effect of the thermal power plant. All the input information of the hydrothermal system is taken in [21] and the NSTLBO algorithm effectively solves the STHTGS problem by finding the minimum fuel cost of the thermal unit with effectiveness, the slightest value that the NSTLBO algorithm can find. Table 1 demonstrates the hydraulic discharge for a period of 24 h in an optimal manner and also it is shown in Fig. 3.

Optimum power generation of hydro, thermal and total load demand are given in Table 2. The optimum power generation of water, steam and power demand is shown in Fig. 4 with an optimal manner.

Fig. 2 Hydraulic network

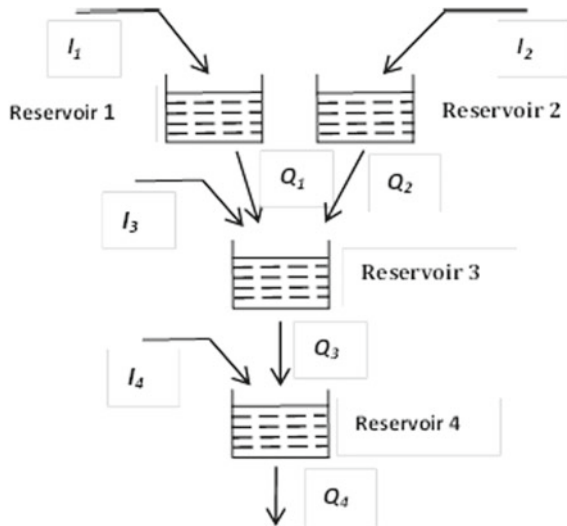


Table 1 Optimal generation of water discharge for four hydro units

Hours	Water discharge				Hours	Water discharge			
	Q_1	Q_2	Q_3	Q_4		Q_1	Q_2	Q_3	Q_4
1	12.781	14.880	17.828	11.041	13	8.5450	6.8000	24.331	18.210
2	6.1990	6.8820	22.875	17.815	14	13.379	13.734	11.177	24.123
3	14.702	14.392	12.188	23.912	15	6.1500	11.524	11.342	23.784
4	15.452	14.352	28.339	24.836	16	14.999	14.999	15.127	24.872
5	6.7224	14.981	26.521	16.567	17	14.489	12.011	17.359	24.999
6	14.999	14.553	10.780	24.748	18	14.999	15.790	10.703	21.025
7	14.961	14.363	23.672	24.989	19	12.688	10.284	28.725	24.983
8	14.574	8.8940	29.754	24.582	20	14.671	7.5870	12.842	24.791
9	14.951	12.924	11.446	14.913	21	6.9730	14.969	29.999	24.984
10	12.968	14.793	23.551	17.260	22	14.999	14.759	13.385	18.617
11	11.849	13.681	29.978	23.335	23	9.8850	10.594	10.815	23.297
12	13.859	14.548	13.527	24.069	24	11.775	14.999	17.560	24.925

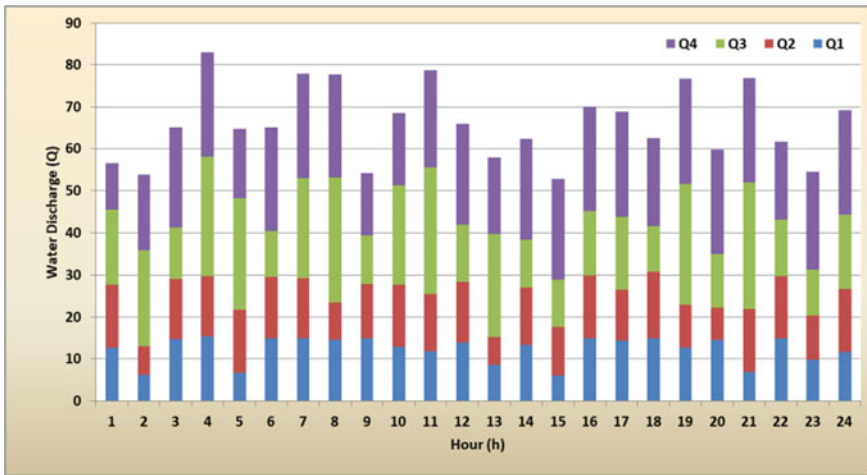


Fig. 3 Optimal generation of water discharge of four hydro units

Table 2 Optimal generation of four units of hydro, three units of thermal power and total load demand

Hour	Hydro				Thermal			Total load demand
	Ph1	Ph2	Ph3	Ph4	Ps1	Ps2	Ps3	
1	93.270	83.227	29.869	178.58	102.274	123.614	139.166	750
2	56.383	47.315	5.1850	216.99	101.77	123.887	228.47	780
3	96.363	77.525	39.374	208.17	20.910	206.958	50.700	700
4	92.260	75.829	0.0000	199.13	102.97	40.001	139.81	650
5	56.870	71.903	0.0000	259.45	102.767	40.000	139.01	670
6	86.053	70.088	37.770	325.32	99.590	40.809	140.37	800
7	87.337	71.486	0.0000	327.82	21.047	124.83	317.48	950
8	87.047	49.339	0.0000	326.61	20.904	208.53	317.57	1010
9	86.628	66.798	39.204	262.91	102.67	125.69	406.10	1090
10	84.497	72.197	0.0000	283.37	103.52	39.626	496.79	1080
11	82.018	68.992	0.0000	320.47	101.55	209.60	317.37	1100
12	86.179	71.816	44.005	323.40	102.70	292.70	229.20	1150
13	67.785	37.885	0.0000	290.68	103.17	293.19	317.29	1110
14	87.053	69.127	39.061	323.61	153.81	40.369	316.97	1030
15	52.563	61.833	39.154	321.88	101.56	292.53	140.48	1010
16	89.021	72.499	36.803	311.89	20.007	209.81	319.97	1060
17	86.612	63.69	31.388	327.84	102.63	209.64	228.20	1050
18	86.622	73.286	38.752	309.06	172.35	210.72	229.21	1120
19	83.831	56.459	0.0000	327.80	165.23	210.16	226.52	1070
20	86.640	42.486	43.407	327.24	100.67	39.827	409.73	1050
21	56.218	72.457	0.0000	327.80	21.105	292.67	139.75	910
22	87.023	71.517	38.953	293.64	101.66	39.807	227.40	860
23	73.535	57.885	38.799	320.31	102.55	206.6	50.321	850
24	96.651	80.948	55.537	302.37	173.89	39.81	50.794	800

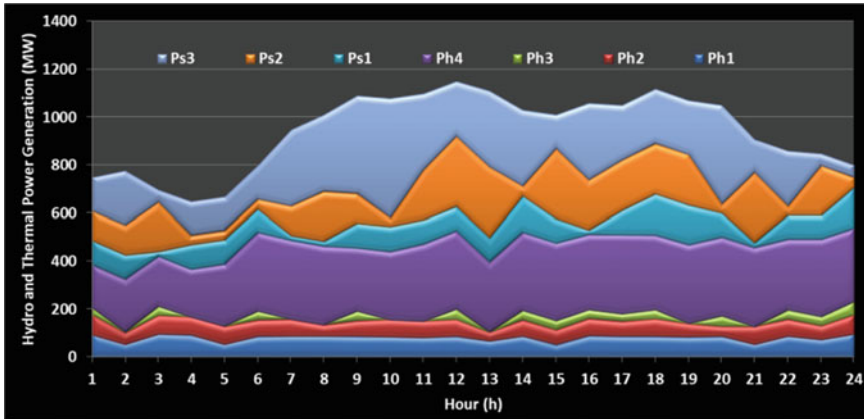


Fig. 4 Optimal generation of four hydro, three thermal units

6 Conclusion

This study presents a short-term hydrothermal power generation distribution by means of the NSTLBO algorithm. Toward verifying the efficacy of the future method, this experiment includes several consecutive hydroelectric chains and different thermal power plants to analyze and solve the SHTGS problem. The simulation outcomes of the NSTLBO technique demonstrate the validity then pre-eminence of the technique. In the future, different efficient and state-of-the-art multi-objective optimization algorithms will be considered for uncertain short-term hydrothermal production scheduling decisions.

References

1. Hossain SM, Shiblee MSAAF (2017) A short review study on problems during hydro-thermal optimal scheduling. *Int J Adv Technol Eng Explor* 4(34):142–147
2. Tahir MF, Haoyong C, Mehmood K, Ali N, Bhutto JA (2019) Integrated energy system modeling of China for 2020 by incorporating demand response, heat pump and thermal storage. *IEEE Access* 7:40095–40108
3. Tahir MF, Haoyong C, Khan A, Javed MS, Laraik NA, Mehmood K (2019) Optimizing size of variable renewable energy sources by incorporating energy storage and demand response. *IEEE Access* 7:103115–103126
4. Padmini S, Jegatheesan R, Thayyil DF (2015) A novel method for solving multi-objective hydrothermal unit commitment and scheduling for GENCO using hybrid LR-EP technique. *Procedia Comput Sci* 57:258–268
5. Das S, Bhattacharya A, Chakraborty AK (2018) Solution of short-term hydrothermal scheduling problem using quasi-reflected symbiotic organisms search algorithm considering multi-fuel cost characteristics of thermal generator. *Arab J Sci Eng* 43:2931–2960
6. Naversen CØ, Helseth A, Li B, Parvania M, Farahmand H, Catalão JP (2020) Hydrothermal scheduling in the continuous-time framework. *Electr Power Syst Res* 189:106787

7. Allen WOOD, Wollenberg BF (1984) Power generation, operation, and control. Wiley
8. Rodrigues RN, da Silva EL, Finardi EC, Takigawa FY (2012) Solving the short-term scheduling problem of hydrothermal systems via Lagrangian relaxation and augmented Lagrangian. *Math Probl Eng*
9. Zhang H, Yue D, Xie X, Dou C, Sun F (2017) Gradient decent based multi-objective cultural differential evolution for short-term hydrothermal optimal scheduling of economic emission with integrating wind power and photovoltaic power. *Energy* 122:748–766
10. Sinha N, Lai LL (2006) Meta heuristic search algorithms for short-term hydrothermal scheduling. In: 2006 international conference on machine learning and cybernetics. IEEE, pp 4050–4056
11. Hota PK, Chakrabarti R, Chattopadhyay PK (1999) Short-term hydrothermal scheduling through evolutionary programming technique. *Electr Power Syst Res* 52(2):189–196
12. Sinha N, Chakrabarti R, Chattopadhyay PK (2003) Fast evolutionary programming techniques for short-term hydrothermal scheduling. *IEEE Trans Power Syst* 18(1):214–220
13. Türkay B, Mecitoğlu F, Baran S (2011) Application of a fast evolutionary algorithm to short-term hydro-thermal generation scheduling. *Energy Sources Part B Econ Plann Policy* 6(4):395–405
14. Nallasivan C, Suman DS, Henry J, Ravichandran S (2006) A novel approach for short-term hydrothermal scheduling using hybrid technique. In: 2006 IEEE power India conference. IEEE, p 5
15. Fang N, Zhou J, Zhang R, Liu Y, Zhang Y (2014) A hybrid of real coded genetic algorithm and artificial fish swarm algorithm for short-term optimal hydrothermal scheduling. *Int J Electr Power Energy Syst* 62:617–629
16. Wong KP, Wong YW (1994) Short-term hydrothermal scheduling part. I. Simulated annealing approach. In: *IEE Proc Gener Transm Distrib* 141(5):497–501
17. Wong SYW (2001) Hybrid simulated annealing/genetic algorithm approach to short-term hydro-thermal scheduling with multiple thermal plants. *Int J Electr Power Energy Syst* 23(7):565–575
18. Kumar VS, Mohan MR (2011) A genetic algorithm solution to the optimal short-term hydrothermal scheduling. *Int J Electr Power Energy Syst* 33(4):827–835
19. Samudi C, Das GP, Ojha PC, Sreeni TS, Cherian S (2008) Hydro thermal scheduling using particle swarm optimization. In: 2008 IEEE/PES transmission and distribution conference and exposition. IEEE, pp 1–5
20. Yu B, Yuan X, Wang J (2007) Short-term hydro-thermal scheduling using particle swarm optimization method. *Energy Convers Manage* 48(7):1902–1908
21. Pasupulati B, Kumar A, Asokan K (2018) Optimal scheduling of hydrothermal system considering different environmental emissions using NSTLBO approach. *Int J Renew Energy Res (IJRER)* 8(4):1913–1925
22. Pasupulati B, Kumar RA, Asokan K (2020) A novel approach of non-dominated sorting TLBO for multi objective short-term generation scheduling of hydrothermal-wind integrated system. In: *Emerging trends in electrical, communications, and information technologies: proceedings of ICECIT-2018*. Springer, Singapore, pp 411–428

Novel Design for Power Factor Correction Using FACTS Device and Soft Computing Tools



Shubham Gedam, Prema Daigavane, and Tushar Guhe

Abstract The present study outlines the development of an innovative control strategy, wherein a PFC boost converter is integrated in reverse with a rectifying diode, to effectively recuperate current bearing harmonics to produce by a single-phase semiconductor rectifier. The line power command receives two distinct inputs, namely a regulator of voltage for the direct current (DC) link and an estimator of output power. The current controller's oscillation is employed to monitor the line current command. In the absence of a diode rectifier (a non-linear load), it can be inferred that the power factor correction (PFC) boost converter solely draws sinusoidal electricity from the source. The utilization of a rectifying diode in a PFC boost converter results in the automatic current draw, which facilitates the generation of a sinusoidal current, equivalent to the entirety of the current generated by the source. This phenomenon can be observed in a typical academic writing format. The proposed converters exhibit a near ideal power factor, enhanced power density, a streamlined control mechanism, and a reduction in the number of harmonic control constituents.

Keywords Soft computing tools · Fuzzy logic · Power quality · Power stability · Power factor

S. Gedam (✉) · P. Daigavane
Electrical Department, G.H. Raisoni College of Engineering, Nagpur, India
e-mail: shubham.gedam.mtechps@ghrce.raisoni.net

P. Daigavane
e-mail: prema.diagavane@raisoni.net

T. Guhe
Janki Electromech, Nagpur, India

1 Introduction

Power electronics converters, Switching Mode Power Supplies (SMPS), and other non-linear loads are becoming more and more common and computers, has led to the generation of significant levels of power pollution that are manifest in the distribution and transmission networks. The diode rectifier and phase controlled rectifier's main drawbacks are their subpar power factors and high levels of pulsed current flowing from the alternating current (AC) mains. The aforementioned circuits are capable of generating a substantial amount of energy. Reactive power and current harmonics are notable instances of power pollutants that result in the distortion of line voltage, elevated levels of thermal energy in transformer and electrical machinery cores, together with increased transmitting and disseminating line losses. The power factor in an AC circuit is denoted by the sine of the angle that the voltage and current made with one another. Inside of a circuit, the active or real power, denoted as $VI \cos$ and measured in units of watts or kilowatts, is present as a consequence of a V for the input voltage and an I for the resultant line current. The reactive power, which is expressed as VAR or kVAR, is commonly referred to as $VI \sin$ in academic literature. The power factor statistic serves as a gauge for a system's actual power efficiency. The measurement of in academic discourse, the line voltage and current's phase shift and distortion are encountered. By dividing the average real power by the appearance power, the power factor is calculated. The product of the root mean square values of current and voltage determines how much power is perceived. This formulation belongs to the paradigm of academic writing due to its technical and specialized use of terminology. Conventional line commutated rectifiers are associated with certain drawbacks, namely a displacement factor that lags behind the utility's voltage (Fig. 1).

- (1) They produce a large number of input current harmonics.
- (2) The input current has a lot of harmonics produced by them.

The present study involves the utilization of an AC power supply and implements rectifiers to transform pulsed DC. Furthermore, filters are used to facilitate the conversion of pulsed DC into a stable and refined form of DC. The utilization of filters may result in the distortion of our line current, characterized by the presence of current harmonics. Source current is distorted as a result of the harmonics. We are able to manage and reduce the harmonics in our system by employing specialized power devices. But some require considerable maintenance and are highly sophisticated. Because of this, two stage conversions are more popular; in this case, we picked this two stage conversion. When converting electricity from one form to another, power electronic converters are essentially necessary. Between the source and load sides, they serve as an interface. The ubiquitous implementation of single-phase power converters in recent years has notably amplified the predicaments with power quality in electrical networks. It is advised to use a high-frequency active power factor correction (PFC) circuit to handle the power factor issue. Any DC-DC converters are considered to be adequate for this application. If the input current is regulated suitably or if it inherently exhibits power factor correction (PFC) properties. Two distinct

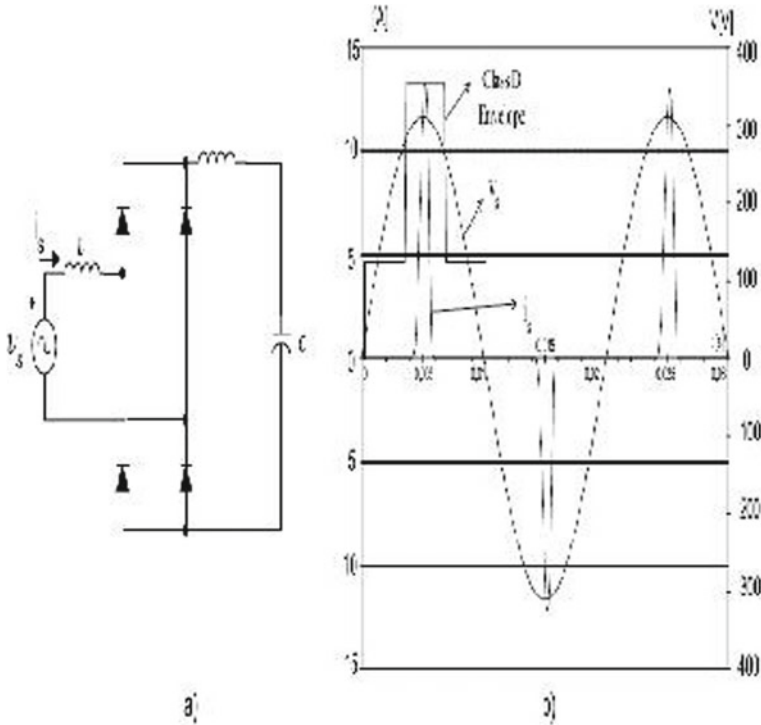


Fig. 1 Single-phase rectifier **a** circuit **b** waveforms of input voltage and current

modes of operation, Discontinuous Inductor Current Mode and Continuous Inductor Current Mode, can be employed in converting DC to DC. In the former mode, the inductor current remains non-zero throughout the switching cycle, while in the latter mode, the inductor current approaches zero periodically during each cycle. As in the past, we are employing some specialized power equipment to reduce harmonics and raise power factor. The charge's aftereffects. Changes are made to the power factor to make it more resemble an AC power supply circuit. Nothing except facts controllers and filters for custom power devices. These circuits have a high maintenance cost and are quite complicated. Employing some specialized power equipment to reduce harmonics and raise power factor. The charge's aftereffects. Changes are made to the power factor to make it more resemble an AC power supply circuit. Nothing except facts controllers and filters for custom power devices. These circuits have a high maintenance cost and are quite complicated (Figs. 2 and 3).

- (A) Passive filtering.
- (B) Active filtering.
- (C) Powerful hybrid systems.

Peak current control, mode control of the average current, and hysteresis control are commonly utilized methodologies in the field. The employment of on-cycle

Fig. 2 Active filtering

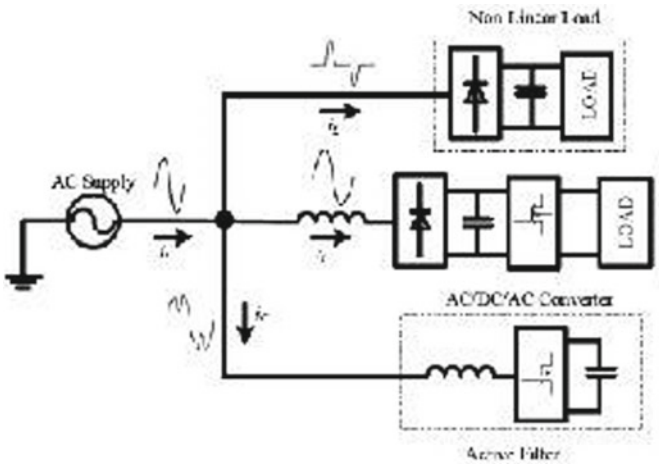
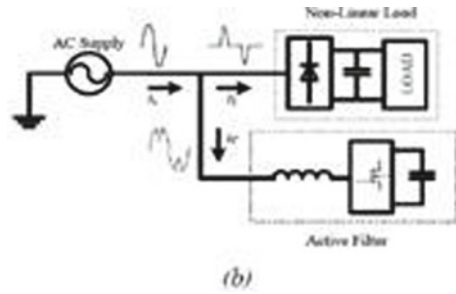
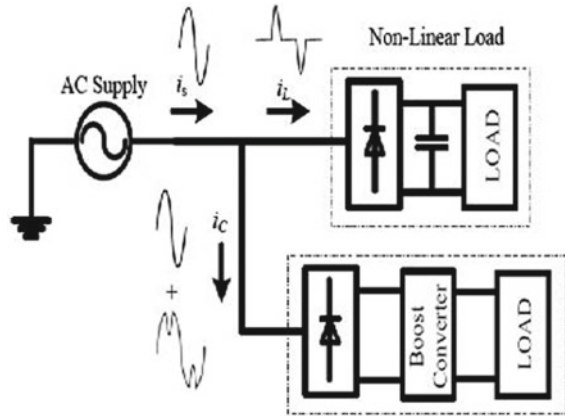


Fig. 3 Powerful hybrid

control and self-control has experienced a recent surge. In order to mitigate the deleterious effects of the diode rectifier's generated current harmonics, a three-level pulse width modulation (PWM) alternating current/direct current (AC/DC) converter has been employed by various means. In order to mitigate the production of harmonic currents stemming from non-linear loads, a number of solutions have employed the use of active power filters. The limitations of these methods encompass the need to employ a distinct converter for each non-linear load and a higher incidence of switching losses in certain approaches due to their dependence on an augmented count of switching devices. Some strategies entail highly intricate control algorithms. A novel approach to improve power factor correction efficiency by implementing a PFC boost converter is proposed as a potential solution aimed at effectively addressing the various associated issues.

Fig. 4 Schematic diagram of STATCOM



2 Suggestive Concept

The harmonic current produced when non-linear loads are used is mitigated by means of a singular converter, a limited number of switching devices, and a simplified control mechanism. The present study proposes the implementation of power factor correction methodology to mitigate the harmful effects of harmonic pollution resulting from a diode rectifier for a single phase on the power line. The suggested configuration functions as a contemporaneous source, which is connected in parallel to the non-linear load, and is fine-tuned to generate the specific harmonic currents required by the load. Through this approach, the alternating current source is only required to supply the fundamental currents. In parallel with the diode rectifier, a power factor correction (PFC) boost converter is used to reduce the negative impacts of the harmonic current caused by the non-linear load. To monitor the current command line in this configuration, hysteresis current control is employed. Consequently, the complete configuration indubitably generates a sinusoidal electric current flow from the power source. The suggested converter’s power switches are modulated in a manner that enables the generation of an almost sinusoidal line current characterized by low current distortion, minimal total harmonic distortion, and stable DC bus voltage regulation (Fig. 4).

3 Control Plan

In general, we have a wide variety of control mechanisms, including the following technique for voltage followers.

- (1) Common Current Control Method.
- (2) VFT: Only take into account load-side parameters.
- (3) AVCT: Take source and load sides into account parameter.

4 Proposed Control Scheme

This methodology yields an enhancement in the quality of input current waveform. In this scenario, a contemporary error amplifier is utilized to drive a pulse width modulation modulator and perceptibly detects and filters the current passing through the inductor. The internal current loop effectively mitigates the discrepancy between the reference and mean input current by adopting the following approach.

One common technique for establishing a reference current involves the multiplication of the output from the voltage error amplifier, which is responsible for setting the amplitude of the current reference signal (as shown in the accompanying figure), by a suitably scaled copy of the rectified line voltage V_g . This methodology involves ensuring that the reference signal remains consistently proportional to the line voltage whilst being intrinsically synchronized. This measure is deemed essential to achieving a power factor of unity. The proposed technique confers several advantages, including an unvarying switching frequency, the lack of a compensatory ramp, a reduced susceptibility to commutation-related interference, and superior input current waveforms as compared to peak current control.

In this model when STATCOM is applied on the output side of the system it can improve the power quality and it shows the decreasing of THD of the system hence it improves the power factor.

5 Simulation Modelling

See Fig. 5.

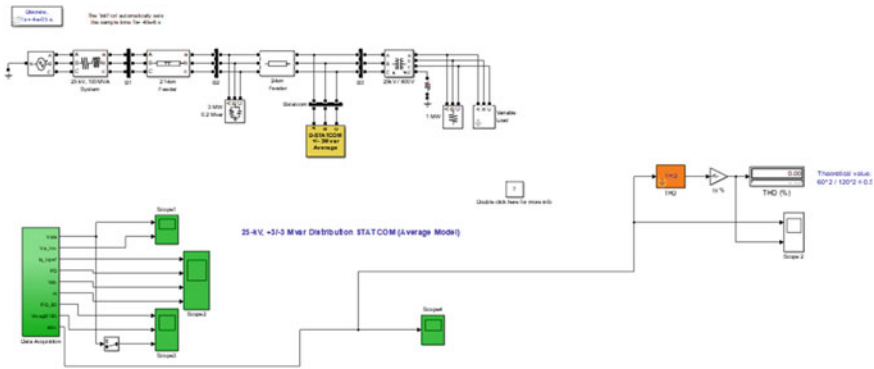


Fig. 5 Simulation modelling

6 Proposed Fuzzy Logic Control

Comparing the FLC to traditional PI controllers, the FLC is a new controller. Utilizing this controller enhanced the system’s responsiveness and stability. Here, a series of logics were employed to create FLC operation. There is no need for a mathematical strategy. Both building and using this controller are simple. The DC link voltage regulation is very important in the suggested work. The DC link voltage was managed to seamlessly control D STATCOM operation. Therefore, the voltage at the DC link is adjusted here with the aid of FLC. The rules that are applied in the proposed system are shown in the table (Fig. 6).

From this we can say that when power quality is improved in system automatically power factor in the system also improves.

When we applied D STATCOM with fuzzy logics we have seen that to improvement wave form quality.

From Fig. 09 it is seen that when we applied D STATCOM the fluctuation of graph is more wavy or we can say that the graph of the current is more transient in nature but when we applied fuzzy logic control (FLC) the nature of the current graph is less transient and they will approach near to the steady state stability as shown in Fig. 7.

From Fig. 8 shows the current nature of graph is more fluctuating in nature but when we applied a fuzzy logic control it shows Fig. 10 is the current nature of the graph is more stable.

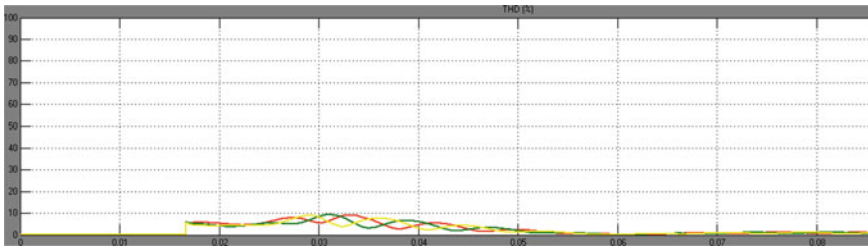


Fig. 6 Reduced THD near about 2% graph

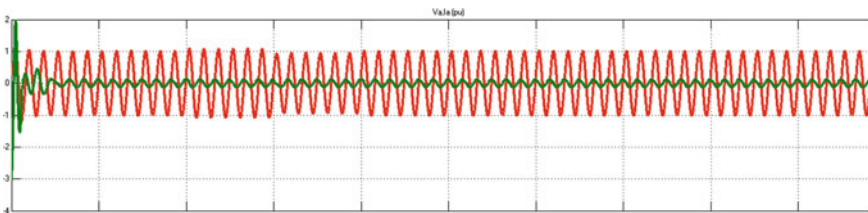


Fig. 7 Simulation result after connected fuzzy logic control (FLC)

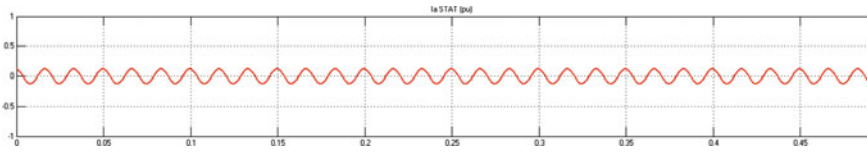


Fig. 8 Current wave form after connected fuzzy logic control

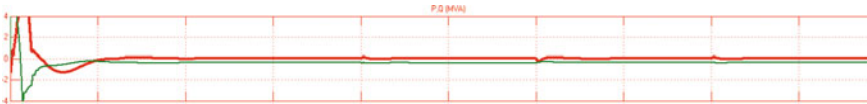


Fig. 9 Active power (P) in watt and reactive power (Q) in VAR

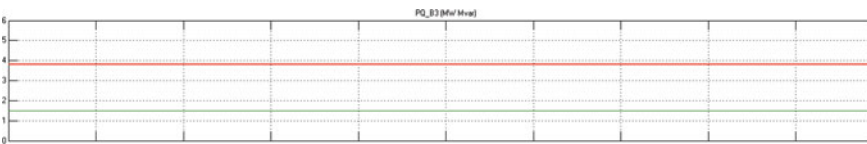


Fig. 10 Active power (P) in PU and reactive power (Q) in PU

When comparing with Figs. 9 and 11 it is seen that the active power (P) and reactive power (Q) is more stable when applied FLC control. And less stable when we applied D STATCOM.

Figure 10 shows the active power in per unit and reactive power in per unit the graph must be approach steady state stability limit.

7 Conclusion

In this work, a novel AC/DC boost-type converter for power factor correction (PFC) applications is presented. It has some intriguing and distinctive features. The harmonic current produced by the non-linear load might theoretically be reduced by a single converter, eliminating the need for a second converter. Considering the outcomes of a modelling analysis the design in question was found to completely get rid of all lower order harmonics. It is possible to get a power factor close to unity and a THD measurement about 2%. This method is only applicable in situations where the non-linear load exhibits pulsing electrical current while consistent, is quite small. Additionally, in the evaluation of related literature, a scholarly investigation of various configurations of power factor correction approaches has been conducted. But when we applied to the fuzzy logic control the nature of the graph becomes comes to more stable than D STATCOM. And is seen that the quality active and the reactive power must be approach to the steady stability. And hence from above discussion we can

conclude that when power quality is improved then automatically power factor of the system also improved.

References

1. Kapse SS, Shrawane, Daigavane MB, Daigavane PM (2020) Optimal localization and sizing of UPFC to solve the reactive power dispatch problem under unbalanced conditions. *IETE J Res* 66(3):396–413
2. Sarkar M, Subudhi B, Daigavane P (2019) Robust PI damping controller design using wide-area signal for inter-area oscillation in power system. In: 2019 8th international conference on power systems (ICPS). IEEE, pp 1–6
3. Chaudhari M, Babu K, Khubalkar SW, Daigavane P (2019) Off-grid hybrid solar power conditioning unit for critical and non-critical loads. In: 2019 international conference on intelligent computing and control systems (ICCS). IEEE, pp 969–974
4. Katole DN, Daigavane MB, Gawande SP, Daigavane PM (2018) Improved single phase instantaneous pq theory for DVR compensating nonlinear load. In: 2018 IEEE international conference on power electronics, drives and energy systems (PEDES). IEEE, pp 1–6
5. Katole DN, Daigavane MB, Gawande SP, Daigavane PM (2018) Vector based analysis for design of single phase SRF controller in dynamic voltage restorer. In: 2018 IEEE international conference on power electronics, drives and energy systems (PEDES). IEEE, pp 1–6
6. Lanjewar A, Khubalkar SW, Junghare AS (2018) Comparative analysis of two loop integer and fractional order PID controller for inverted pendulum. *ICSEDPS 2018, GHRCE, Nagpur*
7. Bose S, Khubalkar S (2022) Power quality analysis of textile industry—findings and recommendations. In: 2022 2nd Asian conference on innovation in technology (ASIANCON). IEEE, pp 1–6
8. Shende AG, Khubalkar SW, Vaidya P (2021) Hardware implementation of automatic power factor correction unit for industry. In: *Journal of physics: conference series*, vol 2089, no 1. IOP Publishing, p 012032
9. Chaudhari M, Babu K, Khubalkar SW, Talokar S (2019) Off-grid hybrid online solar power conditioning unit for domestic purposes. In: 2019 international conference on computing, power and communication technologies (GUCON). IEEE, pp 121–126
10. Shende D, Jagtap P, Hiware R (2021) Enhanced power quality using unified power flow controller systems. *J Phys Conf Ser* 2089(1):012035 (IOP Publishing)
11. Shrawane Kapse SS, Daigavane MB, Daigavane PM (2018) Improvement of ORPD algorithm for transmission loss minimization and voltage control using UPFC by HGAPSO approach. *J Inst Eng (India) Series B* 99:575–585
12. Adware R, Chandrakar V (2023) Power quality enhancement in a wind farm connected grid with a fuzzy-based STATCOM. *Eng Technol Appl Sci Res* 13(1):10021–10026
13. Peng FZ (2018) Application issues of J.T. Boys, A.W. Green. Current-forced single-phase reversible rectifier. *IEEE Proc B* 136(3):205–211

Chaos and Multistability in Fractional Order Power System: Dynamic Analysis and Implications



Prakash Chandra Gupta and Piyush Pratap Singh

Abstract This paper presents a comprehensive study about the dynamic behavior of a fractional ordered three-machine infinite bus (TMIB) power system model using Grunwald–Letnikov’s method. The study investigates nonlinear behaviors including chaos, coexisting behavior and multistability behaviors, using nonlinear tools such as phase portraits, bifurcation analysis, Lyapunov exponents and Lyapunov dimensions. The results demonstrate that the TMIB system exhibits chaos behavior, which is resulting instability in rotor angle through multiscroll chaotic attractors. Furthermore, it is found that the presence of coexisting attractors and multistability leads to undesired state switching and pose a potential threat to the stability of the TMIB power system. These findings provide valuable insights into the nonlinear behavior of TMIB power system via varying fractional order range and can be used to develop effective countermeasures to address potential stability issues arise in TMIB and similar modern power systems. The simulation is conducted in MATLAB, and the obtained results illustrate the efficacy of the work.

Keywords Power system · Fractional order · Coexisting attractor · Multistability · Chaos · Bifurcation · TMIB

1 Introduction

A multimachine power system is a complex electrical power system consisting of multiple interconnected synchronous generators, transmission lines, transformers, and loads [1]. It is a vital component of the modern electric power grid and is used to supply electrical energy to homes, businesses, and industries. The behavior of a

P. C. Gupta (✉) · P. P. Singh
Department of Electrical Engineering, National Institute of Technology Meghalaya, Shillong,
Meghalaya 793003, India
e-mail: prakashgupta@nitm.ac.in

P. P. Singh
e-mail: piyushpratap.singh@nitm.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
R. N. Shaw et al. (eds.), *Innovations in Electrical and Electronic Engineering*, Lecture
Notes in Electrical Engineering 1109, https://doi.org/10.1007/978-981-99-8289-9_4

multimachine power system is highly nonlinear and dynamic, making its analysis and control a challenging task [2].

Previous studies have investigated the chaos and bifurcation behavior of different power system models [3–6]. The SMIB system has been specifically studied for its nonlinear complex dynamic motions, including period-doubling bifurcations (PDB), chaotic movements, and unbounded motions [7, 8]. Furthermore, the SMIB power system's qualitative behavior has been investigated through the imposition of a periodic load disturbance [9]. In addition, the presence of white Gaussian noise in a system can affect its dynamical behaviors, with higher intensity of random noise potentially leading to increased chaos and instability [10, 11]. Besides single machine models, there have also been studies on multimachine power system models. The Melnikov method has been used to investigate chaos in rotor dynamics of power systems [12]. Specifically, a special case of the conservative swing equation based on a three-machine system was analyzed. The rotor dynamics of this three-machine swing equation were studied using qualitative and quantitative tools to examine its chaotic behavior. State feedback and dither signal control were also applied to transform the chaotic oscillation into a periodic oscillation [13, 14].

Multistability is a common phenomenon in complex systems, such as biological, physical, and engineering systems, and can have important implications for their behavior and stability [15, 16]. Multistability is a property of a system that can exhibit multiple stable equilibrium states or attractors. This means that the system can maintain different stable states depending on its initial conditions, and small changes or perturbations may cause it to transition from one stable state to another [2, 17]. In recent years, the analysis of chaos and multistability behavior has become increasingly popular, including its fractional order counterpart. Fractional calculus, which is a generalization of traditional calculus, has emerged as a powerful tool for modeling and analyzing the dynamics of systems [18]. It has gained attention in various fields due to its ability to capture nonlocal and memory-dependent behavior, making it suitable for describing complex systems with long-range interactions [19]. The application of fractional calculus to the study of chaos and multistability behavior has provided new insights and perspectives, leading to a deeper understanding of the underlying dynamics of complex systems [20, 21]. Despite the existence of multiple power systems in which chaos and bifurcation behavior have been discussed, the phenomenon of multistability in power system models has received limited attention. Although there have been recent studies on multistability in certain power system models [2, 17], the fractional order counterparts of such models have not yet been explored in the literature. Therefore, the investigation of the fractional order counterpart of power system models exhibiting multistability is a significant research direction that needs to be addressed. Motivated by the above literature, this paper focuses on fractional order model of TMIB power system. The contribution and novelty of this work are listed as follow:

1. A TMIB power system model is derived from the conventional N-machine power system model, and its fractional order analysis is carried out using Grunwald–Letnikov's (GL) method.

2. The TMIB power system model shows the rotor angle instability by multiscroll chaotic attractors.
3. The TMIB power system model reveals the presence of multistability (coexistence of attractors) behavior.

The rest of the paper is organized as follows: In Sect. 2, the mathematical model of fractional order TMIB power system is described. In Sect. 3, different nonlinear dynamic behaviors, including chaos, bifurcation, and multistability, in the fractional order range are discussed. Finally, in Sect. 4, the conclusions and future scope of the work are presented.

2 Mathematical Modeling of TMIB Power System

The n-machine system [12], is the basic for the classical model of the electrical network. The system comprises n internal machine nodes (1 to n) working in parallel mode, with reference node 0 acting as a neutral node. During stability analysis, the voltages $E_1, E_2, E_3, \dots, E_n$ are assumed to remain constant throughout the transient period. The rotor angles for the n-machine system are represented by $\delta_1, \delta_2, \delta_3, \dots, \delta_n$, while $r_1, r_2, r_3, \dots, r_n$ and $x'_{d1}, x'_{d2}, x'_{d3}, \dots, x'_{dn}$ denote the resistance and sub-transient reactance of the n generators. The current output from each generator unit is represented by $I_1, I_2, I_3, \dots, I_n$.

The expression for the electrical power output of the i th machine in the network is given as:

$$P_{ei} = E_i^2 G_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^n E_i E_j Y_{ij} \cos(\theta_{ij} - \delta_i + \delta_j) \quad (1)$$

Here, Y_{ii} and Y_{ij} represent the self admittance or driving point admittance at node i and mutual admittance between nodes i and j , respectively. This can be further written as in term of conductance (G) and susceptance (B) as: $Y_{ii} = G_{ii} + B_{ii}$ and $Y_{ij} = G_{ij} + B_{ij}$.

$$P_{ei} = E_i^2 G_{ii} + \sum_{j=1, j \neq i}^n E_i E_j [B_{ij} \sin(\delta_i - \delta_j) + G_{ij} \cos(\delta_i - \delta_j)] \quad (2)$$

Assuming the resistance of the transmission line is negligible and can be modeled as a pure reactive component, the expression for the electromagnetic output power is simplified as:

$$P_{ei} = E_i^2 G_{ii} + \sum_{j=1, j \neq i}^n E_i E_j B_{ij} \sin(\delta_i - \delta_j) \quad (3)$$

The equation of motion for a multimachine system can be expressed as:

$$\dot{\delta}_i = \omega_i \quad (4)$$

$$\frac{2H_i}{\omega_s} \dot{\omega}_i + D_i \omega_i = P_{mi} - P_{ei} \quad (5)$$

By utilizing Eqs. (3) and (5), we can derive the following Eq. (6):

$$\frac{2H_i}{\omega_s} \dot{\omega}_i = -D_i \omega_i + P_{mi} - E_i^2 G_{ii} - \sum_{j=1, j \neq i}^n E_i E_j B_{ij} \sin(\delta_i - \delta_j) \quad (6)$$

where ω_s represents for synchronous speed, ω_i for the deviation between the rotor angle velocity and synchronous speed for the i th bus, H_i for the inertia of the i th machine, P_{mi} and P_{ei} for the mechanical input power and electromagnetic output power respectively of the i th machine. D_i represents the damping coefficient of the i th machine. To simplify the system parameters, we can define $\frac{2H_i}{\omega_s}$ as M_i , $P_{mi} - E_i^2 G_{ii}$ as P_i , and $E_j B_{ij}$ as T_{ij} .

Assuming that an infinite bus node connects the three generators, shown in Fig. 1, a simplified swing equation for the TMIB power system that describes the generator rotors can be written by using Eq. (6).

$$\begin{cases} \dot{\delta}_i = \omega_i, \\ \dot{\omega}_i = \frac{1}{M_i} [-D_i \omega_i + P_i - E_i \sum_{j=1, j \neq i}^n T_{ij} \sin(\delta_i - \delta_j) - E_i V_B B_{iB} \sin(\delta_i - \delta_B)] \end{cases} \quad (7)$$

where $i=1,2,3$, V_B and δ_B are the infinite bus voltage and phase angle. B_{iB} represents the admittance between the i th machine and infinite bus bar.

2.1 Fractional Order Model of TMIB Power System

Although there are several available methods for fractional order computations, previous studies have primarily utilized the Caputo method for numerical solutions of fractional order systems. However, research has shown that the GL method [22] provides better smoothness in coefficients for solving fractional order systems. Therefore, we

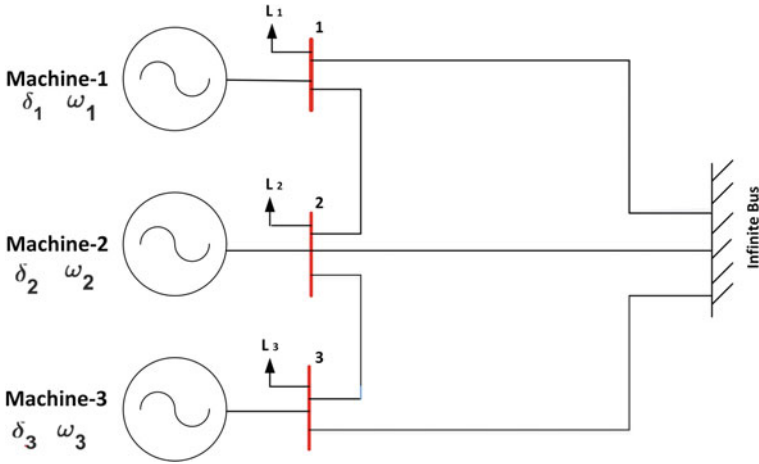


Fig. 1 Three-machine infinite bus (TMIB) power system (7)

utilize the GL method to solve the fractional order multimachine power system model.

The fractional order equation has limits of a and t and is represented by the equation ${}_a D_t^q f(t)$, where q is the fractional order of the differential equation. The calculation involves the use of the generalized difference $\Delta_h^q f(t)$, where h represents the step size. The GL derivative can be defined as:

$${}_a D_t^q f(t) = \lim_{h \rightarrow 0} \left\{ \frac{1}{h^q} \sum_{m=0}^{\lfloor \frac{t-a}{h} \rfloor} (-1)^m \binom{q}{m} f(t - mh) \right\} = \lim_{h \rightarrow 0} \left\{ \frac{1}{h^q} \Delta_h^q f(t) \right\} \quad (8)$$

To perform numerical calculations, the equation mentioned above can be modified as follows:

$$({}_{t-L}) D_t^q f(t) = \lim_{h \rightarrow 0} \left\{ h^{-q} \sum_{m=0}^{N(t)} \beta_j (f(t - mh)) \right\}. \quad (9)$$

To simulate the fractional order TMIB system using the GL method, we utilize the discretization method described in previous studies. Since the required memory for calculating binomial coefficients is theoretically infinite, we truncate the number of samples to make the calculation feasible. The discretization method is as follows:

$$\begin{aligned}
\delta_i(t_k) &= A(\delta_i(t_{k-1}), \omega_i(t_{k-1})) h^{q_{\delta_i}} - \sum_{m=1}^N \beta_m^{q_{\delta_i}} \delta_i(t_{k-m}) \\
\omega_i(t_k) &= B(\delta_i(t_{k-1}), y(t_{k-1})) h^{q_{\omega_i}} - \sum_{m=1}^N \beta_m^{q_{\omega_i}} \omega_i(t_{k-m})
\end{aligned} \tag{10}$$

In the above equation, β denotes the binomial coefficients. The value of N is determined based on the available memory. Specifically, if the available memory is not fully utilized, N is set to the truncation window size L . Otherwise, N is set to k to make the most of the available memory elements. We can define the fractional order TMIB power system derived using Eq. (7) as follows:

$$\begin{cases} D^{q_{\delta_i}} \delta_i = \omega_i, \\ D^{q_{\omega_i}} \omega_i = \frac{1}{M_i} [-D_i \omega_i + P_i - E_i \sum_{\substack{j=1 \\ j \neq i}}^n T_{ij} \sin(\delta_i - \delta_j) - E_i V_B B_{iB} \sin(\delta_i - \delta_B)] \end{cases} \tag{11}$$

The discrete version of the fractional order TMIB power system can be expressed as follows:

$$\begin{cases} \delta_i(t_k) = (\omega_{i_{k-1}}) h^{q_{\delta_i}} - \sum_{m=1}^N \beta_m^{q_{\delta_i}} x(t_{k-m}), \\ \omega_i(t_k) = [\frac{1}{M_i} \{-D_i \omega_{i_{k-1}} + P_i - E_i \sum_{\substack{j=1 \\ j \neq i}}^n T_{ij} \sin(\delta_{i_{k-1}} - \delta_j) \\ - E_i V_B B_{iB} \sin(\delta_{i_{k-1}} - \delta_B)\}] h^{q_{\omega_i}} - \sum_{m=1}^N \beta_m^{q_{\omega_i}} x(t_{k-m}) \end{cases} \tag{12}$$

The value of N is taken as the truncation window size L and as k when all the available memory elements are used. Calculation parameters for analysis of fractional order TMIB power system are given as: $D_1 = 0.003$, $D_2 = 0.0045$, $D_3 = 0.003$, $M_1 = 0.01$, $M_2 = 0.015$, $M_3 = 0.01$, $P_1 = 0.3$, $P_2 = 0.4$, $P_3 = 0.3$, $T_{12} = T_{21} = 0.1$, $T_{13} = T_{31} = 0.6$, $T_{23} = T_{32} = 1$, $B_{1B} = 2$, $B_{2B} = 1.5$, $B_{3B} = 2$, $E_1 = 1$, $E_2 = 1$, $E_3 = 1$, $V_B = 1$, and $\delta_B = 0$. To investigate the various dynamic behaviors within the fractional order range, the value of the fractional order parameter q is varied.

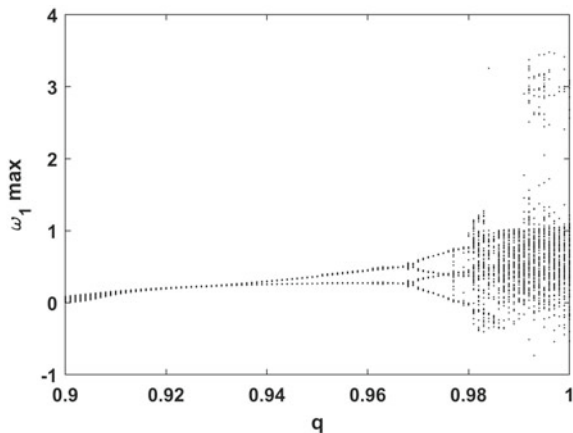
3 Nonlinear Dynamic Behavior of Fractional Order TMIB Power System

In this section, we analyzed the different dynamic behaviors of the TMIB power system by varying the order of differential equation and studying its behavior using both qualitative and quantitative analysis tools.

3.1 Dynamic Response at Varied Fractional Order (q)

Bifurcation analysis is a crucial qualitative tool used to analyze the nonlinear behavior of the system under parameter perturbations. This analysis provides information about the topological changes in the system’s behavior when it experiences small or smooth parameter variations. Nonlinear behaviors are observed using bifurcation diagram while keeping all parameter values constant and by varying the fractional order derivative within the range $q \in [0.9, 1]$. Different transitions in the dynamic behavior are observed in the bifurcation diagram as shown in Fig. 2. To further analyze these behavior the dynamic response in the phase plane is plotted at specific values to the corresponding bifurcation diagram shown in Fig. 3. Lyapunov exponents also correspond the phase plane behavior at different value of fractional order q listed the Table . At $q = 0.93$, by observing bifurcation plot, phase portrait, and the nature of maximum positive Lyapunov exponents (MLE) is negative, confirm the period-1 behavior. Similarly at $q = 0.965$, the system displays a period-2 behavior. However, at $q = 0.982$, the system exhibits aperiodic behavior, can be observed from the bifurcation plot and phase plane analysis. The Lyapunov exponents have been calculated and the nature of MLE confirms the chaos behavior. Further, dissipativity and Lyapunov dimension analysis are performed. Dissipativity analysis is an essential measure for understanding the converging and diverging nature of dynamical system. It provides a way to quantify the energy flow of a system and identify critical parameters that affect the system dynamics. A dynamical system is dissipative if the sum of all Lyapunov exponents is negative, i.e., $\sum_{i=1}^n \lambda_i < 0$, where λ_i are the individual Lyapunov exponents of the system, and n is the state space dimension. The dissipativity, at $q = 0.982$, is calculated as the sum of Lyapunov exponents ($L_1 + L_2 + L_3 + L_4 + L_5 + L_6$) = -1.4680 . The $-ve$ value indicates that the system is dissipative in nature. Lyapunov dimension or Kaplan–Yorke dimension is calculated using the following formula as:

Fig. 2 Bifurcation diagram of TMIB power system (11) when the order $q \in [0.9, 1]$



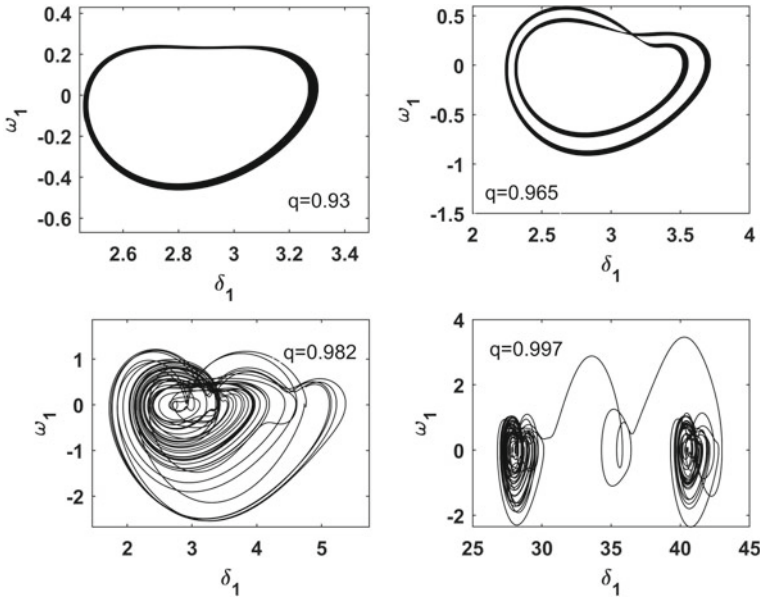


Fig. 3 Phase plane behavior (transients are removed) correspond to the bifurcation diagram of fractional order multimachine power system (11) at four different value of fractional order q

$$D_{KY} = \lim_{\epsilon \rightarrow 0} \left[\sum_{i=1}^k \lambda_i + \frac{1}{|\lambda_{k+1}|} \sum_{i=k+1}^n |\lambda_i| \right]$$

where λ_i denotes the i -th Lyapunov exponent, k is the largest index such that the sum of the first k exponents is non-positive, and n is the total number of Lyapunov exponents. The Kaplan–Yorke dimension is obtained as $D_{KY} = 2.699$ using calculated Lyapunov exponents at $q = 0.982$ which implies that the system has fractal nature and has complex attractor with a dimension between 2 and 3. At the value of fractional order $q = 0.997$, the phenomenon of chaos breaking has observed, where the MLE is positive, resulting in rotor angle instability with the emergence of multiscroll chaotic attractors. The time series and phase portrait behavior of the rotor angle and angular velocity for each machine are shown in Fig.4, under multiscroll chaotic attractors.

Obtain nonlinear behavior through the bifurcation and Lyapunov exponents in fractional order range which can be effectively applied to parameter tuning in the stability analysis of multimachine power systems experiencing chaotic behavior. Overall, fractional calculus combined with nonlinear theories is used to study chaos and chaos breaking behavior in multimachine power systems, which may provide helpful insights to aid in the development of counter measures to ensuring its stability and reliability.

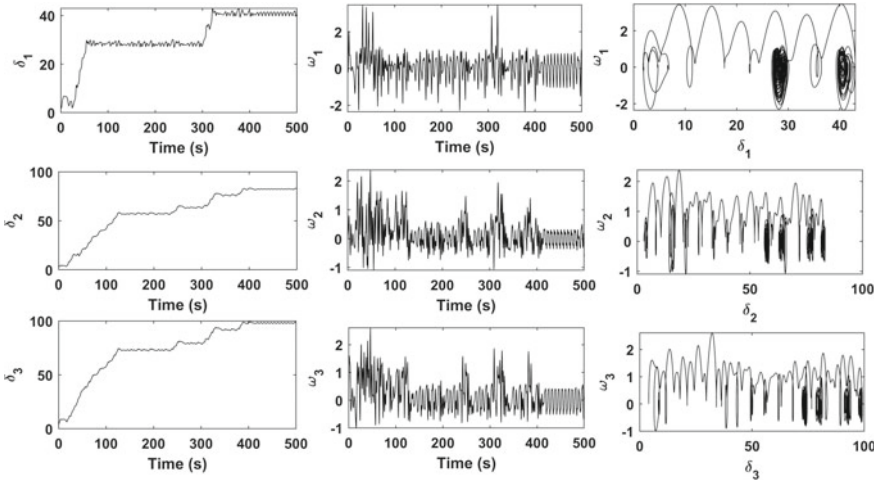


Fig. 4 Time series and phase plane behavior of fractional order multimachine power system (11), reflects angle instability through multiscroll chaotic attractors at $q = 0.997$

3.2 Coexistence of Attractor and Multistability Analysis

In the previous subsection, we analyzed the TMIB power system for a range of fractional orders. Now, we fixed the order of the differential equation at $q = 0.997$ and varied the damping coefficient parameter $d_1 = \frac{D_1}{M_1}$ while keeping all other parameters constant at their previously defined fixed values. The bifurcation diagram, shown in Fig. 5, displays the various dynamic behaviors, including PDB route to chaos, observed when the damping coefficient d_1 is varied within the range of $[0, 0.6]$. The bifurcation diagram presented in Fig. 5 not only illustrates the dynamic behavior of the TMIB power system with varying damping coefficient d_1 , but also shows the coexistence phenomenon at three distinct initial conditions. These initial conditions are $[\delta_1(0), \omega_1(0), \delta_2(0), \omega_2(0), \delta_3(0), \omega_3(0)] = [\delta_1(0), 0, 3, 0, 4, 0]$ where the value of $[\delta_1(0)]$ is considered as 2, 3, 4, and represented via *black*, *red*, *blue* colors in throughout analysis.

Fig. 6 displays the multistability behavior of the TMIB power system (11) at machine-1 in terms of its phase plane and time series response. The figure shows the response of the rotor angle and angular velocity of machine-1 at three different initial conditions for three distinct values of the parameter d_1 . Specifically, Fig. 6a corresponds to $d_1 = 0.15$, and it depicts the coexistence of three chaotic attractors at the given initial conditions. Figure 6b corresponds to $d_1 = 0.14$, and it depicts the coexistence of two chaotic attractors and one multiscroll chaotic attractor. Fig. 6c corresponds to $d_1 = 0.137$, shows the coexistence of three multiscroll chaotic attractor.

The multistability behavior of the TMIB power system (11) at machine-1 is depicted in Fig. 6, in terms of its phase plane and time series response. This Fig. 6

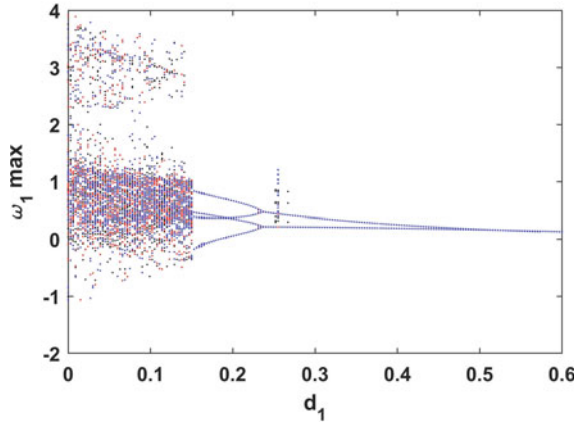


Fig. 5 Coexisting bifurcation behavior of fractional order multimachine power system (11) at three initial conditions $[2, 0, 3, 0, 4, 0]$, $[3, 0, 3, 0, 4, 0]$, $[4, 0, 3, 0, 4, 0]$; depicted via colors, respectively

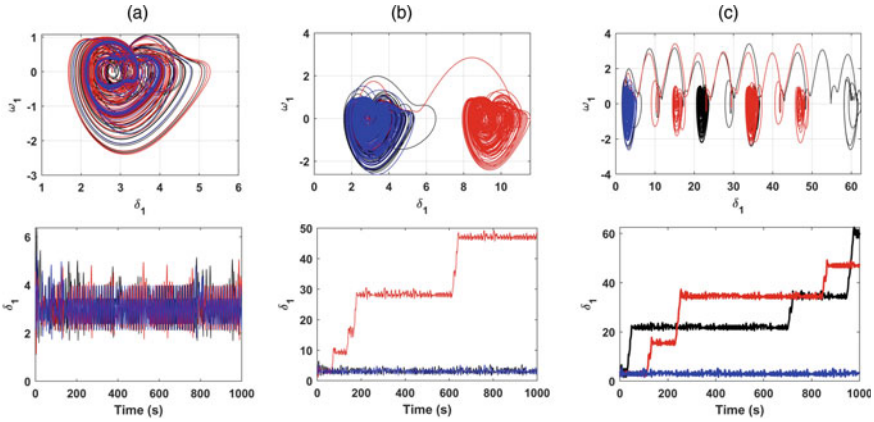


Fig. 6 Multistability (coexistence of attractors) in the fractional order multimachine power system (11), at three initial conditions depicted with different colours. **a** Phase plane and time series behavior of machine-2, exhibiting the coexistence of three chaotic attractors at $d_1 = 0.15$. **b** Coexistence of two chaotic and one multiscroll chaotic attractor of machine-2 at $d_1 = 0.14$. **c** Coexistence of one chaotic attractor with two multiscroll chaotic attractors of machine-2 at $d_1 = 0.137$

depicts the response of machine-1’s rotor angle and angular velocity under three different initial conditions for three different values of the parameter d_1 . Figure 6a, corresponds to $d_1 = 0.15$ and displays the presence of three chaotic attractors under the specified initial conditions. The coexistence of two chaotic attractors and one multiscroll chaotic attractor is depicted in Fig. 6b, which corresponds to $d_1 = 0.14$. The presence of three multiscroll chaotic attractors is shown in Fig. 6c, which corresponds to $d_1 = 0.137$.

The multistability exhibited by the TMIB power system is concerning due to its potential for unpredictable and uncontrolled behavior. Multistability can increase the risk of cascading failures in power systems. If one component of the system fails, the system may shift to a different stable state, leading to further failures and potentially a complete system collapse.

4 Conclusion

The paper presents a TMIB power system model which is derived from the classical N-Machine model power system, and applies Grunwald–Letnikov’s method to perform a fractional order analysis. The study reveals interesting behaviors, such as period-1, period-2, chaos, and angle instability by multiscroll chaotic attractor under specific parameter values and fractional order q , using both qualitative and quantitative analysis tools. Furthermore, coexisting behavior and multistability are also observed through the use of coexisting bifurcation diagram, phase plane, and time series analysis. However, this study highlights the presence of chaos and multistability in the fractional order TMIB power system, and it is concerning due to its potential for unpredictable and uncontrolled behavior, which can result in power outages, equipment damage, and safety hazards. Therefore, it is critical to manage and control the effects of chaos and multistability in practical applications such as power systems to ensure their safe and reliable operation.

References

1. Ugalde-Loo CE, Acha E, Licéaga-Castro E (2013) Multi-machine power system state-space modelling for small-signal stability assessments. *Appl Math Model* 37:10141–10161. <https://doi.org/10.1016/j.apm.2013.05.047>
2. Gupta PC, Singh PP (2022) Multistability, multiscroll chaotic attractors and angle instability in multi-machine swing dynamics. *IFAC-PapersOnLine* 55:572–578. <https://doi.org/10.1016/j.ifacol.2022.04.094>
3. Gupta PC, Banerjee A, Singh PP (2018) Analysis of global bifurcation and chaotic oscillation in distributed generation integrated novel renewable energy system. In: 2018 15th IEEE India council international conference (INDICON). IEEE, Coimbatore, India, pp 1–5
4. Gupta PC, Banerjee A, Singh PP (2019) Analysis and control of chaotic oscillation in FOSMIB power system using AISMC technique. In: 2019 IEEE students conference on engineering and systems (SCES). IEEE, Allahabad, India, pp 1–6
5. Das P, Gupta PC, Singh PP (2021) Bifurcation, chaos and PID sliding mode control of 3-bus power system. In: 2020 3rd international conference on energy, power and environment: towards clean energy technologies. IEEE, Shillong, Meghalaya, India, pp 1–6
6. Gupta PC, Singh PP (2023) Multistability, coexisting behaviours and control of fractional order dissipative small scale grid with disturbances and noise. *Eur Phys J Spec Top*. <https://doi.org/10.1140/epjs/s11734-023-00927-0>
7. Nayfeh MA, Hamdan AMA, Nayfeh AH (1990) Chaos and instability in a power system - Primary resonant case. *Nonlinear Dyn* 1:313–339. <https://doi.org/10.1007/BF01865278>

8. Chen H-K, Lin T-N, Chen J-H (2005) Dynamic analysis, controlling chaos and chaotification of a SMIB power system. *Chaos, Solitons Fractals* 24:1307–1315. <https://doi.org/10.1016/j.chaos.2004.09.081>
9. Wang X, Chen Y, Han G, Song C (2015) Nonlinear dynamic analysis of a single-machine infinite-bus power system. *Appl Math Model* 39:2951–2961. <https://doi.org/10.1016/j.apm.2014.11.018>
10. Wei DQ, Luo XS (2009) Noise-induced chaos in single-machine infinite-bus power systems. *EPL (Europhysics Letters)* 86:50008. <https://doi.org/10.1209/0295-5075/86/50008>
11. Wei DQ, Zhang B, Qiu DY, Luo XS (2010) Effect of noise on erosion of safe basin in power system. *Nonlinear Dyn* 61:477–482. <https://doi.org/10.1007/s11071-010-9663-0>
12. Kopell N, Washburn R (1982) Chaotic motions in the two-degree-of-freedom swing equations. *IEEE Trans Circuits Syst* 29:738–746. <https://doi.org/10.1109/TCS.1982.1085094>
13. Wang X, Lu Z, Song C (2019) Chaotic threshold for a class of power system model. *Shock Vib* 2019:1–7. <https://doi.org/10.1155/2019/3479239>
14. Chang S-C (2020) Stability, chaos detection, and quenching chaos in the swing equation system. *Math Prob Eng* 2020:1–12. <https://doi.org/10.1155/2020/6677084>
15. Zhusubaliyev ZT, Mosekilde E, Churilov AN, Medvedev A (2015) Multistability and hidden attractors in an impulsive Goodwin oscillator with time delay. *Eur Phys J Spec Top* 224:1519–1539. <https://doi.org/10.1140/epjst/e2015-02477-8>
16. Zhusubaliyev ZT, Mosekilde E (2015) Multistability and hidden attractors in a multilevel DC/DC converter. *Math Comput Simul* 109:32–45. <https://doi.org/10.1016/j.matcom.2014.08.001>
17. Gupta PC, Singh PP (2022) Chaos, multistability and coexisting behaviors in small-scale grid: Impact of electromagnetic power, random wind energy, periodic load and additive white Gaussian noise. *Pramana* 97:3. <https://doi.org/10.1007/s12043-022-02478-w>
18. Hilfer R (2000) Applications of fractional calculus in physics. WORLD SCIENTIFIC
19. Baleanu D (ed) (2017) Fractional calculus: models and numerical methods, 2nd edn. World Scientific, New Jersey
20. Debbouche N, Ouannas A, Momani S, Cafagna D, Pham V-T (2022) Fractional-order biological system: chaos, multistability and coexisting attractors. *Eur Phys J Spec Top* 231:1061–1070. <https://doi.org/10.1140/epjs/s11734-021-00308-5>
21. Ma C, Jun M, Cao Y, Liu T, Wang J (2020) Multistability analysis of a conformable fractional-order chaotic system. *Physica Scripta* 95:075204. <https://doi.org/10.1088/1402-4896/ab8d54>
22. Tolba MF, AbdelAty AM, Soliman NS, Said LA, Madian AH, Azar AT, Radwan AG (2017) FPGA implementation of two fractional order chaotic systems. *AEU—Int J Electron Commun* 78:162–172. <https://doi.org/10.1016/j.aeue.2017.04.028>

Identification of Critical Nodes Using Granger Causality for Strengthening Network Resilience in Electrical Distribution System



Divyanshi Dwivedi¹, D. Maneesh Reddy, Pradeep Kumar Yemula², and Mayukha Pal¹

Abstract As countries around the world commit to reducing brownfield energy generation and shifting toward clean energy, the placement of renewable energy sources (RES) optimally in the electrical distribution system remains a strenuous issue. Improperly integrating RES could have a detrimental impact on the efficient operation of the grid. This study proposes a real-time data-driven approach for optimal DERs allocation and identification of critical nodes in the electrical distribution system. A community detection clustering is performed on the IEEE 123 node feeder system to optimally cluster the nodes into two regions. Then, the Granger causal analysis is used to identify critical nodes in the system that are susceptible to failure or extreme events which may interrupt the operation of the system. Hence, strategically allocating RES to these critical nodes enhances network resilience, as validated by the computation of the percolation threshold. The findings reveal an impressive 37% boost in the system's resilience attributed to the optimized deployment of RES.

Keywords Data-driven method · Renewable energy sources · Electrical distribution system · Granger causality · Percolation threshold · Resilience

1 Introduction

In 2018, a major power disruption took place in Puerto Rico when a tree fell on a crucial power line close to Cayey. This led to a significant outage, causing around 870,000 customers to lose access to electricity, making it the most severe power outage incident experienced by the territory of the USA [1]. Instances of cascading failures occur when a critical component in the electrical distribution system (EDS) faces disruption and subsequently affects other areas of the network, resulting in power outages for a large number of customers. To avoid such scenarios, identifying

D. Dwivedi · M. Pal (✉)

ABB Ability Innovation Center, Asea Brown Boveri, Hyderabad, India
e-mail: mayukha.pal@in.abb.com

D. Dwivedi · D. M. Reddy · P. K. Yemula
IIT Hyderabad, Hyderabad, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
R. N. Shaw et al. (eds.), *Innovations in Electrical and Electronic Engineering*, Lecture Notes in Electrical Engineering 1109, https://doi.org/10.1007/978-981-99-8289-9_5

critical nodes and implementing necessary preventive measures may prove beneficial in preventing cascading failures [2]. Placing DERs at the identified critical nodes is a crucial precautionary measure that helps in maintaining continuous power flow and enhances the resilience of the system. “Resilience” refers to the ability to withstand, respond to, adapt to, and prevent disruptive events in the system during extreme events [3]. In a resilient EDS, residential solar PV panels and energy storage devices are commonly used for rapidly reconfiguring power flow and restoring electricity when extreme events happen [4]. Therefore, identifying these critical nodes and strategically placing RES in these locations could further improve the system’s resilience. In the existing literature, the researchers have not thoroughly evaluated the real-time identification of these crucial nodes due to the complexities involved [5]. Rather critical nodes have been considered only for maintaining the certainty for the supply–demand [6] and in the presence of undetectable cyber-attacks [7]. However, the incorporation of distributed energy resources (DER) at the critical node for enhancing the system’s resilience in real time was not performed.

In the distribution grid, the incorporation of RES is preferable for medium/low voltage levels as it causes reduction of system losses, and stress on generators and transmission system [8]. Various methods have been proposed for identifying optimal locations for DERs integration [9]. Placing DERs in inappropriate locations may lead to voltage deviations, power quality problems, increased power loss, and harmonic distortions. These issues have the potential to adversely impact system reliability and resilience. The current work only considers voltage stability margins and reduction in losses as key factors for optimal DER placement, without taking system reliability and resilience into account [10, 11]. Therefore, it is necessary to consider system resilience when optimally placing the DERs.

Quantifiable metrics are needed to measure the resilience of power distribution systems [12]. These metrics are based on the definition of resilience and include measures such as the aggregation of adaptive capacities of system assets, to assess the system’s ability to withstand adverse conditions [13]. Additionally, a metric has been proposed for evaluating the influence of the short-term event on the long-term resilience of the system [14]. Another probabilistic metric has been developed to assess the resilience of EDS concerning extreme incidents [15]. While these metrics evaluate the operational resilience of the system, they do not provide a comprehensive perspective on the entire infrastructure as a single quantity [16]. The metrics discussed above do not consider the progression of events and how different characteristics of the system are impacted. As a result, the evaluation of resilience becomes inaccurate which may affect system operations. To address this issue, we use the percolation threshold as a metric to evaluate resilience in a more effective manner [4]. The percolation threshold is a statistical measure utilized to track system state transitions in response to extreme events, by monitoring the progress of such events. It is essential for energy resilience planning and operational decision-making, and the data-driven analysis can be integrated into any digital substation. Conversely, existing methods for identifying critical nodes and measuring system resilience are not feasible for real-world digital substations. For the identification of the optimal locations to integrate RES into the EDS, our study employs the Granger causality technique.

Granger causality was introduced by Granger in 1969 and is proven to be an effective statistical tool for analyzing time series data. This technique enables the identification of cause-and-effect relationships in an econometric model before the effect occurs. It involves using a multivariate regression model to examine the predictability of one signal's time-series data with the assistance of another signal's history [17, 18]. Granger causality has applications in multiple domains, including finance and neuroscience [19]. Furthermore, the principle of causal inference, which is based on information fusion, has the potential to aid in predicting electrical behavior resulting from shared causal dependencies [20]. By leveraging the efficacy of this technique, we were able to identify closely dependent and correlated nodes within the network that were appropriate for DER integration.

In this paper, we introduce a novel data-driven analysis that can be integrated into digital substations for optimizing the allocation of DERs. To test and validate our framework, we simulate the standard IEEE 123 node feeder system and obtained the time-series data of active power for analysis. By computing the Granger causality between the system's nodes, we identified the correlation between nodes. In cases where a node has high dependencies on other nodes, their failures could trigger cascading failures in EDS. Critical nodes in the network were identified and selected for the integration of RES for improving their self-sustainability. Our study further demonstrated that placing DER at the identified critical nodes improves the resilience of the system by the calculation of the system's percolation threshold.

The paper is structured as follows: Sect. 2 outlines the methodology utilized, and Sect. 3 provides information about the simulation and the steps taken for implementing the proposed framework. In Sect. 4, we analyze the results of the proposed framework on IEEE 123 test system and validate its performance. Finally, Sect. 5 concludes the study and summarizes our findings.

2 Materials and Methods

2.1 Discussion on Granger Causality

Granger causality referred to as G-causality is a statistical technique that determines how the past values of one time-series, T_2 , could help for predicting the future values for another time series, T_1 , such that T_2 Granger-causes (G-causes) T_1 . This technique is established on the multivariate autoregressive model (AM) of a time-series process. Mathematically, to determine if T_2 G-causes T_1 , a bivariate autoregressive (BVAR) model is constructed for both time series, with the same time length of both processes. In the context of electrical distribution system analysis, T_1 and T_2 represent the time-series data of power consumption recorded at 1-minute intervals.

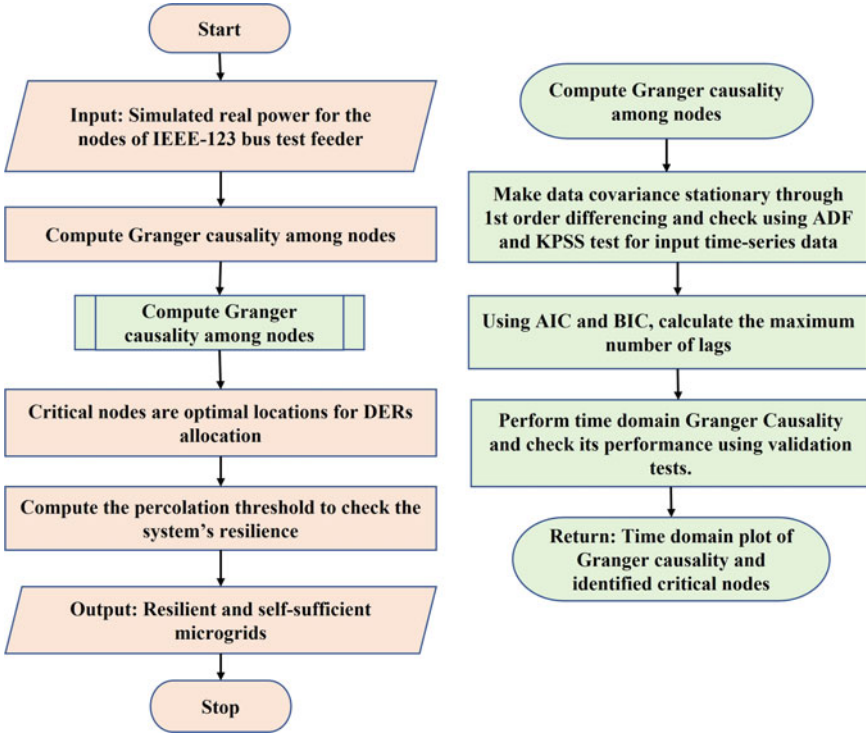


Fig. 1 Proposed framework put forward comprises distinct stages, including the application of Granger causality, the identifying critical nodes, and the assessment of the system's resilience

Model—1:

$$T_1(x) = \sum_{(i=1)}^n \alpha_{(11,i)} T_1(t-i) + \sum_{(i=1)}^n \alpha_{(12,i)} T_2(t-i) + \varepsilon_1(x) \quad (1)$$

$$T_2(x) = \sum_{(i=1)}^n \alpha_{(21,i)} T_1(t-i) + \sum_{(i=1)}^n \alpha_{(22,i)} T_2(t-i) + \varepsilon_2(x) \quad (2)$$

The value of n is the model order, representing the maximum lag integrated into our analysis to compute the Akaike Information Criterion (AIC) [21] and Bayesian Information Criterion (BIC) [22]. The Yule–Walker equations are utilized to determine the model parameters, represented by coefficient $\alpha_{(yz,i)}$ ($i = 1, 2, \dots, n$ and $y, z = 1, 2$). Furthermore, $\varepsilon_1(x)$ and $\varepsilon_2(x)$ are the error term used for loss of variances. Then, the univariate AM for T_1 and T_2 is computed by removing T_2 in Eq. (1) and T_1 in Eq. (2) in Model 2.

Model—2:

$$T_1(x) = \sum_{(i=1)}^n \alpha_{(11,i)}^* A_1(t-i) + \varepsilon_{11}(x) \quad (3)$$

$$T_2(x) = \sum_{(i=1)}^n \alpha_{(22,i)}^* A_2(t-i) + \varepsilon_{22}(x) \quad (4)$$

If the $\varepsilon_{11}, \varepsilon_{22}$ increases in comparison to error terms, $\varepsilon_1, \varepsilon_2$ then it means that T_2 G-causes T_1 and T_1 G-causes T_2 respectively. The magnitude of G-causality is determined by calculating the logarithmic ratio of the error terms. To illustrate, if T_2 G-causes T_1 , then the magnitude of G-causality is expressed as:

$$\mathcal{F}_{2 \rightarrow 1} = \log \frac{\text{var}(\varepsilon_{11}(x))}{\text{var}(\varepsilon_1(x))} \quad (5)$$

When dealing with multiple time series processes, such as T_1, T_2, T_3, T_4 , the process is the same as previously detailed, except the implementation of multivariate AM instead of BVAR, which is also known as conditional Granger-Causality [18]. An important presumption underlying models states that the time-series must exhibit covariance stationarity. Essentially, this means that the time series should possess consistent variance, mean, and covariance among two points based solely on their relative positions. The implementation of Granger causality with step-by-step instructions is depicted in Fig. 1. We use Granger causal analysis for identifying the nodes having high causal dependency in the distribution system. These nodes are recommended as the optimal allocation of RES.

2.2 Percolation for Evaluating Resilience

In order to evaluate the efficacy of the proposed approach in strengthening the electrical system's resilience at the designated DER placement, the percolation threshold is calculated as a resilience metric. The percolation threshold is a statistical tool used for the identification of variations in a system's operations. Percolation threshold offers both qualitative and quantitative measures of network resilience, making it an effective tool [4]. The percolation threshold is calculated by first determining the percolation strength, then evaluating susceptibility and estimating percolation threshold ρ_c at the point of maximum susceptibility [23].

Percolation Strength,

$$PS_{\infty}(p) = \frac{1}{NR} \cdot \sum_{q=1}^R SB_q(p) \quad (6)$$

Here, the function $SB(p)$ is dependent on the probability of bond occupation, defined as $p = v/V$, where V and v are the total edges and the edges that have been

removed from the initial structure respectively. The total nodes are denoted by N . The susceptibility can be calculated as follows:

$$\chi(p) = \frac{(1/N^2 R) \sum_{q=1}^R (SB_q(p))^2 - [PS_\infty(p)]^2}{P_\infty(p)} \quad (7)$$

$$\rho_c = \arg[\max \chi(p)] \quad (8)$$

We began by constructing a complex network for the distribution system with and without DER integration and used Eq. (6) to determine the bond occupation probability of the network. Next, we calculated the percolation strength of the network by removing edges/vertices, then evaluated the bond occupation probability at which susceptibility reaches its maximum; referring to it as the percolation threshold given in Eq. (8). It is considered as a quantifiable measure of resilience, where a higher value corresponds to a more resilient electrical system. This methodology was applied to validate the results obtained from Granger causality.

3 Data and Its Processing

The simulation was performed for the generation of the data and the proposed methodology. The IEEE 123 node test feeder is modeled in GridLAB-D which operates at 4.16 kV. The system comprises 84 constant loads and 36 other nodes (Meter nodes). The distribution system nodes were clustered into two regions using community detection clustering [24] and named them as Region-I and Region-II, as shown in Fig. 2.

Moreover, in order to determine the optimal locations for integrating DERs into the system, functional dependence between the nodes is identified using Granger causality. Nodes with high functional interdependence would significantly impact other nodes of the region and interruptions in these nodes could result in cascading failures or system outages. Therefore, identifying the highly dependent nodes and installing DERs (primarily 50 kW solar PV panels) in those locations would enhance the electrical distribution system's resilience and reliability.

In order to verify the efficacy of the optimal DER placement for enhancing system resilience, we calculated the percolation threshold for two regions under two conditions; integrated DERs and the absence of DERs. Initially, compute the Pearson correlation coefficient between the nodes, and apply a positive coefficient threshold for generating the adjacency matrix, and constructed a complex network. The percolation threshold was subsequently determined by applying the method described in Sect. 2.2 to the resulting complex networks. Figure 1 visually depicts the proposed methodology.

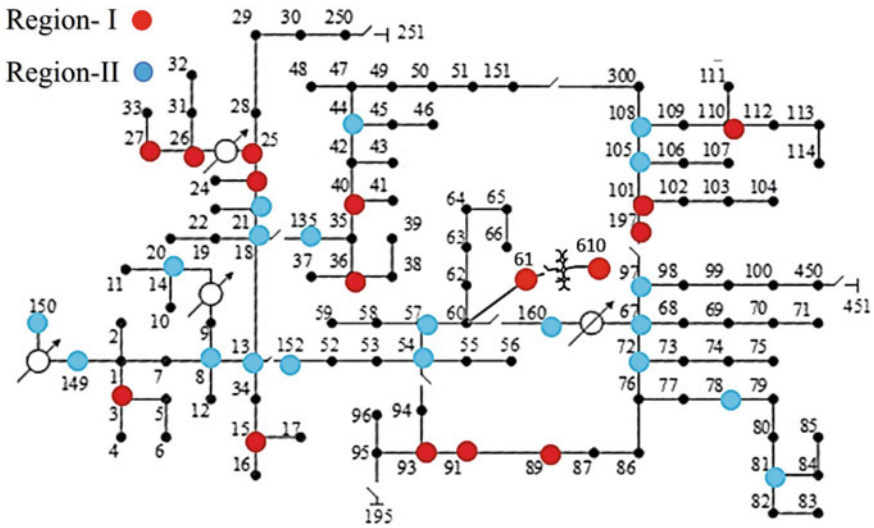


Fig. 2 IEEE-123 node test feeder clustered using community detection clustering algorithm as Region-I and Region-II

4 Results and Discussion

By implementing community detection clustering, the IEEE-123 node test feeder was optimally clustered into two regions, as depicted in Fig. 2. Granger causality was then utilized for determining the optimal locations for RES. We validated the incorporation of RES at the critical nodes in terms of the network’s resilience and reliability by assessing the percolation threshold. The proposed methodology provides an end-to-end data-driven solution that provides a resilient and reliable system, with optimal DER placement by identifying the network’s critical nodes.

4.1 Critical Nodes Identification in Region-I

Granger causality analysis is conducted on Region-I without integrating DER. As seen in Fig. 3, Region-I nodes are highly interconnected with significant magnitude. Among 16 nodes in Region-I, Meter-197, Meter-101, Meter-89, and Meter-26 were identified as critical nodes since they are strongly dependent on all other nodes in the region. Therefore, DERs such as solar panels were integrated into these nodes to enhance the system’s resilience.

Initially, solar panels were incorporated at Meter-101, resulting in a prominent decrease in dependency and magnitude between the nodes, indicating that nodes became less dependent on each other as shown in Fig.4a. Similarly, after

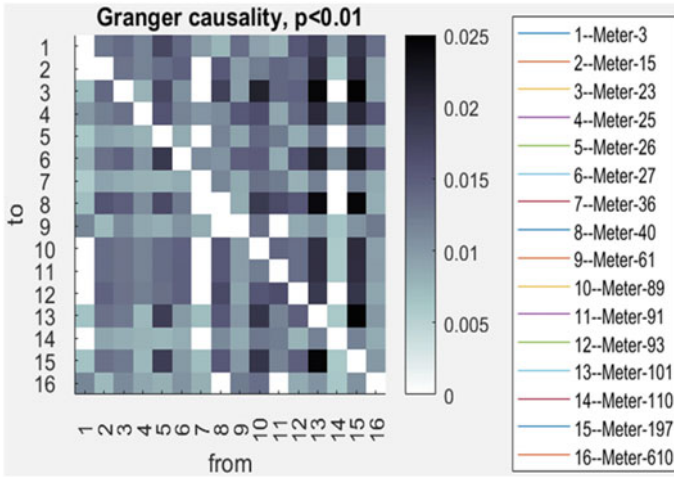


Fig. 3 Granger causality analysis for Region-I

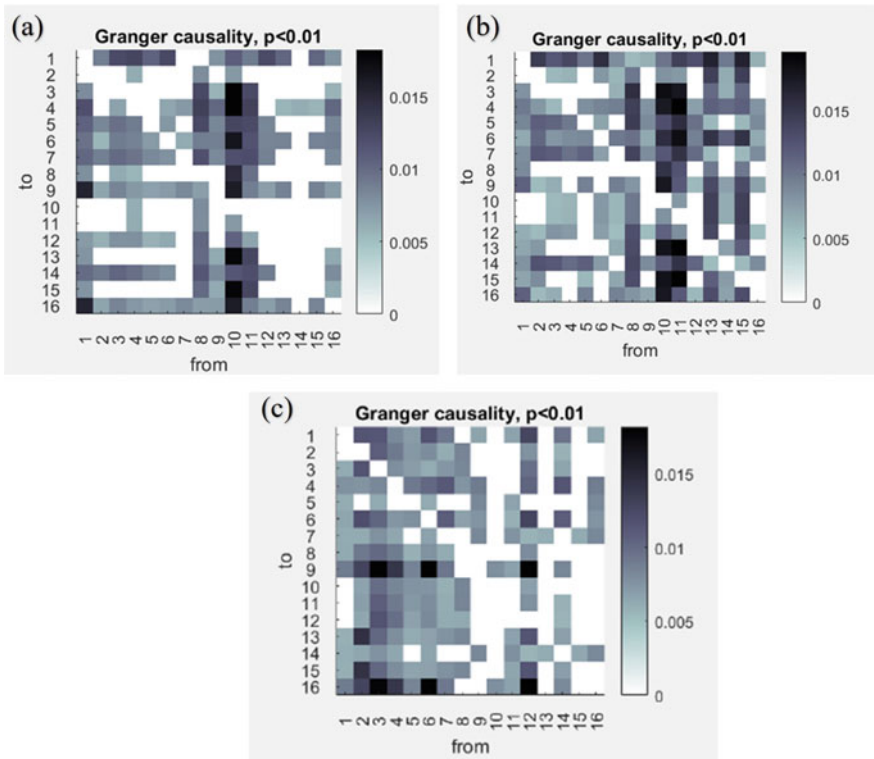


Fig. 4 Granger causal dependencies for Region-I with solar PV panels at a Meter-101, b Meter-197, c all critical nodes

incorporating solar panels at Meter-197, the causing magnitude and dependency have reduced in comparison to the Region-I operating without DERs, as seen in Fig. 4b. Finally, when solar panels were installed at all three critical nodes, the overall network dependencies were significantly reduced, as shown in Fig. 4c.

4.2 Critical Nodes Identification in Region-II

We performed the same analysis on Region-II as well. Figure 5 depicts the causal dependency between the nodes in the absence of DERs, revealing that the nodes causing effects are notably significant. Through the analysis, we were able to identify Meter-13, Meter-18, Meter-21, and Meter-44 as the critical nodes in Region-II that require the inclusion of solar PV panels to boost the region’s reliability.

The inclusion of solar panels at Meter-21 showed a significant reduction in causing magnitude, as depicted in Fig. 6a, with Meter-57 and Meter-54 becoming independent on other system nodes. Similar to that the integration of solar PV panels at Meter-13, as shown in Fig. 6b, resulted in a slightly higher magnitude than Fig. 6a, but with reduced node dependencies. Finally, we integrated solar panels at all critical nodes as shown in Fig. 6c and observed a decrease in causing magnitude as compared to Fig. 5, along with an increase in the number of independent nodes. Thus, incorporating solar panels into the Region-II distribution microgrid has resulted in a more reliable and self-sufficient system that can efficiently maintain supply and demand.

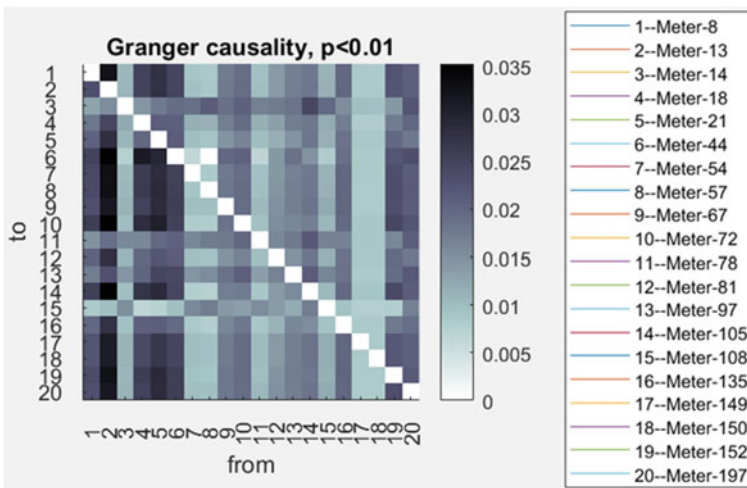


Fig. 5 Granger causality analysis for Region-II

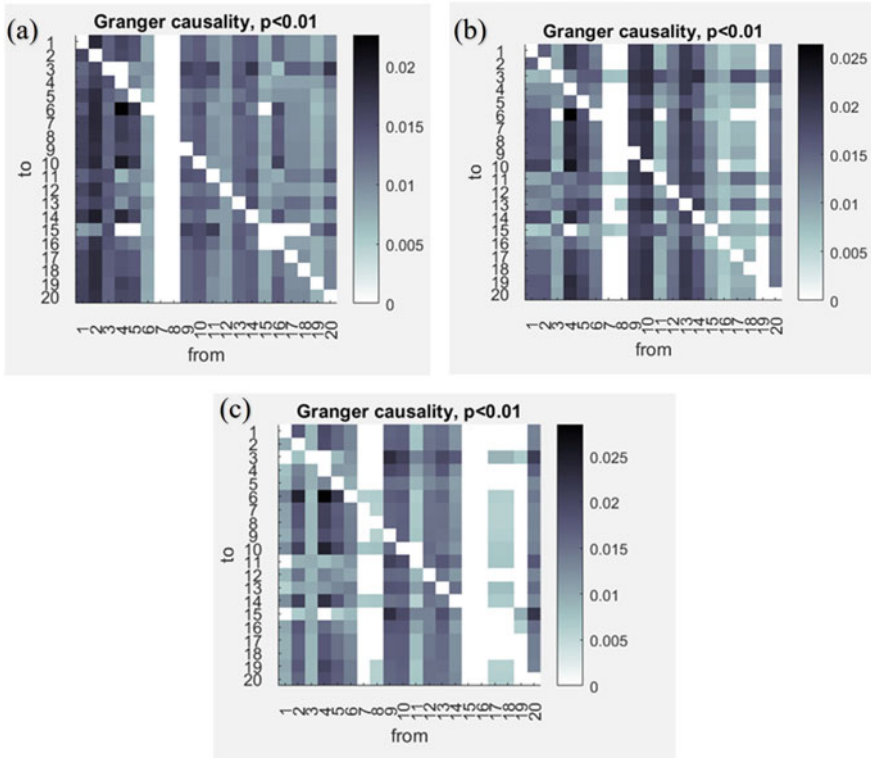


Fig. 6 Granger causal dependencies for Region-II with solar PV panels at **a** Meter-21, **b** Meter-13, **c** all critical nodes

4.3 Evaluation of System’s Resilience

This section validates the reliable and resilient response of a system with RES incorporation at critical nodes. The network’s resilience is determined by the computation of the percolation thresholds for the correlated network, for both regions under two conditions; integrated RES and the absence of RES. Higher percolation thresholds indicate greater network resilience. Results for Region-I and Region-II are shown in Table 1, the percolation threshold value increased with RES incorporation at critical nodes in both regions. Additionally, incorporating the same number of solar PV panels randomly in both regions showed less improvement in the percolation threshold when compared to RES placement at critical nodes. These findings demonstrate the effectiveness of the proposed methodology for the optimal allocation of RES, which enhances resilience and enables the system to sustain interruptions without a complete breakdown.

Table 1 Percolation threshold for both regions under various conditions for RESs integration

Region	RES placement	Percolation thresholds
Region-I	Without RES	0.06722
	With RES at RN	0.07733
	With RES at CN	0.09766
Region-II	Without RES	0.05263
	With RES at RN	0.05868
	With RES at CN	0.08349

*RN- Random Nodes, CN- Critical Nodes

5 Conclusion

The proposed methodology provides an effective approach for optimal DERs allocation in the electrical distribution grid and also quantifies and validates the system’s resilience. We grouped the nodes of the IEEE-123 test feeder system into two clusters based on similarity using a community detection clustering algorithm. We further employed Granger causality to determine the critical nodes in the system that are strongly sensitive to disruptions and critical outages. We again performed Granger causality with solar panels integrated at these locations, and it is observed that there is a significant reduction in causal dependencies and magnitudes between nodes. To validate the proposed method, we calculated the percolation threshold for both regions under two conditions; integrated DERs and the absence of DERs. The results demonstrated a 31.17% increase in the percolation threshold for Region-I and a 37% increase for Region-II when the RES got integrated at critical nodes. These findings prove the effectiveness of our proposed methodology in developing a reliable and resilient distribution system with less node dependency, and nodes could maintain supply–demand independently.

References

1. Sanchez R, Santiago L, Shah K (2022) Power returning to puerto rico after massive outage caused by fallen tree. <https://edition.cnn.com/2018/04/12/us/puerto-rico-power-outage/index.html>. 02 Sept 2022
2. Yuanyuan Z, Yiran A, Qian A (2014) Research on size and location of distributed generation with vulnerable node identification in the active distribution networ. IET Gener Trans Distrib
3. Dwivedi D, Victor SMBK, Yemula PK, Chakraborty P, Pal M (2022) Evaluation of energy resilience and cost benefit in microgrid with peer-to-peer energy trading. arXiv preprint [arXiv:2212.02318](https://arxiv.org/abs/2212.02318)
4. Dwivedi D, Yemula PK, Pal M (2022) A methodology for identifying resiliency in renewable electrical distribution system using complex network. arXiv preprint [arXiv:2208.11682](https://arxiv.org/abs/2208.11682)
5. Dou C, Hu L, Yue D, Zhang Z, Ding X, Li Y (2020) Active power distribution network vulnerable node identification method which considers new energy impact. European patent, vol 21908601. [Online]. Available: <https://patents.google.com/patent/WO2022134596A1/en>

6. Zhao T, Xu Y, Wang Y, Lin Z, Xu W, Yang Q (2016) On identifying vulnerable nodes for power systems in the presence of undetectable cyber-attacks. In: 2016 IEEE 11th conference on industrial electronics and applications (ICIEA), pp 1062–1067
7. Li M, Li S, Li L, Jia Y, Liu X, Yang Y (2013) Identifying vulnerable nodes of complex networks in cascading failures induced by node-based attacks. *Math Prob Eng.* Hindawi Publishing Corporation
8. Dulău LI, Abrudean M, Bică D (2013) Effects of distributed generation on electric power systems. *Procedia Technol* 12:681–686, (2014) In: The 7th international conference interdisciplinary in engineering, INTER-ENG 2013, 10–11 Oct 2013. Petru Maior University of Tirgu Mures, Romania
9. Saad O, Abdeljebbar C (2021) Historical literature review of optimal placement of electrical devices in power systems: Critical analysis of renewable distributed generation efforts. *IEEE Syst J* 15(3):3820–3831
10. Lee SH, Park J-W (2013) Optimal placement and sizing of multiple DGs in a practical distribution system by considering power loss. *IEEE Trans Ind Appl* 49(5):2262–2270
11. Kizito R, Li X, Sun K, Li S (2020) Optimal distributed generator placement in utility-based microgrids during a large-scale grid disturbance. *IEEE Access*, vol 8, pp 21 333–21 344
12. Vugrin E, D C, Andrea R, Silva-Monroy CA (2017) Resilience metrics for the electric power system: a performance-based approach. United States: N. p
13. Phillips T, McJunkin T, Rieger C, Gardner J, Mehrpouyan H (2020) An operational resilience metric for modern power distribution systems. In: 2020 IEEE 20th international conference on software quality, reliability and security companion (QRS-C), pp 334–342
14. Chanda S, Srivastava AK, Mohanpurkar MU, Hovsopian R (2018) Quantifying power distribution system resiliency using code-based metric. *IEEE Trans Ind Appl* 54(4):3676–3686
15. Poudel S, Dubey A, Bose A (2020) Risk-based probabilistic quantification of power distribution system operational resilience. *IEEE Syst J* 14(3):3506–3517
16. Kandaperumal G, Pandey S, Srivastava A (2022) AWR: anticipate, withstand, and recover resilience metric for operational and planning decision support in electric distribution system. *IEEE Trans Smart Grid* 13(1):179–190
17. Ghysels E, Swanson NR, Watson MW (2001) Investigating causal relations by econometric models and cross-spectral methods. *Econometric society monographs*, vol 2. Cambridge University Press, pp 31–47
18. Lin FH, Hara K, Solo V, Vangel M, Belliveau J, Stufflebeam S, Hamalainen M (2009) Dynamic granger-geweke causality modeling with application to interictal spike propagation. *NeuroImage* 47:S169 . Organization for human brain mapping 2009 annual meeting
19. Yuan Q, Jiang, T (2016) Brain efficient connectivity analysis of attention based on the granger causality method. *J Biomed Eng* 56–60
20. Rodriguez-Rivero J, Ramirez J, Martínez-Murcia F, Segovia F, Ortiz A, Salas D, Castillo-Barnes D, Illan I, Puntonet C, Jimenez-Mesa C, Leiva F, Carillo S, Suckling J, Gorris J (2020) Granger causality-based information fusion applied to electrical measurements from power transformers. *Inf Fusion* 57:59–70
21. Cavanaugh JE, Neath AA (2019) The akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computat Stat* 11(3):e1460
22. Profillidis V, Botzorris, G (2019) Trend projection and time series methods. In: Profillidis V, Botzorris G (eds) *Modeling of transport demand*. Elsevier, pp 225–270
23. Mallala B, Dwivedi D, Venkata G, Sowjan Kumar K (2023) Optimal power flow solution with current injection model of generalized interline power flow controller using ameliorated ant lion optimization. *Int J Electr Comput Eng*
24. Acharya D, Zhang H (2020) Community detection clustering via gumbel softmax. *SN Comput Sci* 1:262

A Bald Eagle Search Optimization Approach for Congestion Alleviation in Power System Transmission Channels



Nisha Kumari, Kaushik Paul, and Pampa Sinha

Abstract In the deregulated power system framework, congestion on transmission lines has become a significant concern. This study presents a congestion management strategy that focuses on efficiently adjusting the power output of generators. To determine the most effective generator for rescheduling, the generator sensitivity factor is used. The Bald Eagle Search (BES) optimization technique is employed to optimize the rescheduling of real power delivery from the generators, aiming to minimize congestion costs. The performance of this approach is analyzed using 39-bus New England test framework. A comparison with recent algorithms reveals that BES effectively minimizes congestion costs when contrasted to alternative optimization methods. Additionally, this approach improves the system voltage profile with reduction in the system losses thereby enhancing overall system stability.

Keywords Power rescheduling · Congestion management · Optimization techniques · Bald eagle search optimization · Sensitivity analysis

1 Introduction

Power system congestion occurs when the demand for electricity exceeds the transmission capacity of the power system, leading to potential power outages and blackouts. To mitigate this issue, power system operators use CM techniques to redistribute the power flows in the network and ensure reliable and efficient electricity delivery. CM involves a range of methods, including generation rescheduling, load shedding,

N. Kumari · K. Paul (✉)

Department of Electrical Engineering, BIT Sindri, Dhanbad, India

e-mail: kaushik.ee@bitsindri.ac.in

N. Kumari

e-mail: nisha250197@gmail.com

P. Sinha

School of Electrical Engineering, KIIT University, Bhubaneswar, India

e-mail: pampa.sinhafel@kiit.ac.in

and redispatching, as well as market-based approaches such as locational marginal pricing (LMP) and demand response programs. Effective CM by the system operator enhances the power transfer in the transmission channels while ensuring a consistent supply of electricity to customers.

The CM issues have been analyzed and evaluated by many power system researchers. Aditi et al. managed congestion in a wind-integrated power system by rescheduling the generation of wind and hydro units based on their combined cost bids. The approach uses hydro units with low operating costs and fast start-up times to help alleviate congestion costs and includes secured bilateral transactions [1]. Dehnavi et al. presented a new CM model that identifies candidate zones for congestion alleviation by analyzing a congestion index (CI) that measures the effectiveness of generators and loads in creating and alleviating congestion [2]. Pantos et al. proposed a market-based CM that involves minimum-cost redispatching which considered the combined influence of the generators, loads, and aggregators by the system operator to alleviate congested lines [3]. In [4], the researchers considered the power system risk in combination with the impact of renewable sources to manage congestion. In another research, locational marginal price (LMP) and transmission congestion cost (TCC) are used to manage congestion by identifying the optimal location of DG [5]. In [6] a decentralized approach for EV aggregators to collaborate with common clients for CM is proposed, which treats EVs as flexible loads and mobile distributed storage (MDS). The framework also suggests organizational support from the distribution system operator (DSO) to alleviate severe congestion.

Verma and Mukherjee proposed the use of the Firefly algorithm for CM in a deregulated environment. The algorithm aims to find optimal generation and load patterns to minimize congestion in a deregulated power system. Their study evaluates the effectiveness of the Firefly algorithm through simulation studies. Pandya and Joshi presented a CM approach that utilizes sensitivity analysis to identify critical lines and bus voltages affected by congestion. PSO has been applied to optimize the power generation and load shedding to alleviate congestion. Yesuratnam and Thukaram proposed a CM method based on relative electrical distances (RED) using voltage stability criteria that takes into account the RED between buses in a power system. The proposed method aims to optimize generation rescheduling to alleviate congestion [7]. Hazra et al. presented a CM approach that utilizes power rescheduling on the generation side in association to the scheduling of the loads at the most sensitive buses. The study investigates the effectiveness of the approach through case studies and provides insights into the optimal utilization of the generation and load shedding for congestion management [8]. Dutta and Singh proposed an optimal generator rescheduling method for CM based on PSO. The authors address CM by optimizing the generator schedules using the PSO algorithm. The proposed method aims to minimize the system operating cost while satisfying various operational constraints. Paul et al. proposed an optimal rescheduling method to mitigate congestion using the gravitational search algorithm (GSA). The authors address the issue of congestion in power systems by formulating it as an optimization problem and applying the GSA to find the optimal solution [9]. Zaeim et al. formulated a multi-objective transmission CM method that considers the management of the consumer electricity

patterns. They aimed to optimize the utilization of power consumption behavior to alleviate congestion and improve the overall performance of the power system [10]. In [11], Paul et al. proposed a CM that considered the power deliveries by the wind energy system. The work focused on the formulation of a modified whale optimization technique for congestion alleviation. The proposed technique aimed to find optimal generation patterns that minimize congestion and improve the overall performance. Deb and Goswami proposed a CM approach by generator rescheduling using the artificial bee colony (ABC) technique. The ABC algorithm is employed to find optimal generation schedules that minimize congestion [12].

This research study focuses on addressing congestion management by considering the technique of generator rescheduling for real power. The rescheduling process involves adjusting the power output of generators based on their generation shift factors (GSFs). This research study introduces a BES optimization algorithm for solving CM problem. The proposed BES incorporates effective exploration and exploitation at two stages which are the search stage (Exploration phase), and swooping stage (Exploitation phase) respectively that generated appreciable results for the CM problem. The significant contribution of this work lies in the utilization of BES to effectively optimize the rescheduling of real power output from generators, with the objective of minimizing the associated cost.

2 Problem Formulation

GSF is termed as variation/modifications in actual state of power flow through a transmission line that connects between two buses due to slight variation in the injected power. Mathematically this can be represented as:

$$\text{GSF} = \frac{\Delta P_{ij}}{\Delta P_{G_g}} \quad (1)$$

where P_{ij} = active power flow in overloaded line, P_{G_g} = shift in actual power generation by g th generator. Power flow equation for an overloaded line is expressed as;

$$P_{ij} = V_i^2 G_{ij} + V_i V_j G_{ij} \cos(\theta_i - \theta_j) + V_i V_j B_{ij} \sin(\theta_i - \theta_j) \quad (2)$$

Here P_i is voltage amplitude and θ_i is phase angle of i th bus. Further detailed derivation for GSF is explained in [13]. More sensitive generating units for CM are selected for GR process based upon un-even GSF values. The required rescheduled cost is evaluated on the basis of the price to be paid from the participating generators by resolving the following optimization problem;

The total rescheduling cost is expressed as;

$$\text{Minimize } \sum_g^N C_g(\Delta P_g) \Delta P_g \quad (3)$$

Subjected to

$$F_k^0 + \sum_{g=1}^N (\Delta P_g (\text{GSF}_g)) \leq F_k^{\max} \quad k = 1, 2, \dots, n_1 \quad (4)$$

$$P_g - P_g^{\min} = \Delta P_g^{\min} \leq \Delta P_g \leq \Delta P_g^{\max}, \quad g = 1, 2, \dots, n_1 \quad (5)$$

$$\sum_{g=1}^N \Delta P_g = 0 \quad (6)$$

where ΔP_g is the power injection and C_g is the price bids. These are the bidding prices that the generators will incur/invite to alter their actual power outputs. F_k^0 is the power transfer that is carried through the entire contracts requesting for the transmitting operation. The line's limit is F_k^{\max} . N represents quantity of generating units involved for CM, n_1 represents entire system number of transmitting lines, P_g^{\min} and P_g^{\max} stands for upper and lower limits output parameters for generators, respectively. Here Eqs. (4) and (5) represents the inequality constraints, Eq. (6) represent equality limits that balance the power in CM process.

3 Bald Eagle Search Optimizer

Alsattar and colleagues developed a metaheuristic technique that imitates the hunting behavior of bald eagles, specifically their approach to catching fish. The technique, known as the Bald Eagle Search (BES), involves three stages: selection, exploring the designated area (exploration), and swooping down to attack (exploitation). Initially in the algorithm, the eagle chooses a designated region with the highest concentration of fish, and in the second stage, it navigates within that area to identify the optimal position. Finally, it moves from that position to attack its prey.

3.1 Selection Stage

During this particular stage, the bald eagles choose the area to search by evaluating the quantity of food available. This selection process greatly enhances their ability to capture fish:

$$Z_{\text{new},i} = Z_{\text{best}} + \alpha * r(Z_{\text{mean}} - Z_i) \quad (7)$$

The parameter α , which has a value between 1.5 and 2, is responsible for monitoring the changes in the eagle's position. Z_{best} is the current search area, while r varies within $[0,1]$.

3.2 Search Stage (Exploration Phase)

In this scenario, the exploration behavior is conducted by the eagles to locate fish. They move in a spiral pattern to increase their search efficiency. The optimal location for swooping down to catch the fish can be determined by:

$$Z_{i,\text{new}} = Z_i + h(k) * (Z_i + Z_{i+1}) + m(k) * (Z_i - Z_{\text{mean}}) \quad (8)$$

$$m(k) = \frac{mr(i)}{\max(|mr|)}, \quad h(i) = \frac{hr(i)}{\max(|hr|)} \quad (9)$$

$$hr(k) = r(k) * \cos(\theta(k)); \quad mr(k) = r(k) * \sin(\theta(k)) \quad (10)$$

$$r(k) = \theta(k) + R * \text{rand} \quad (11)$$

$$\theta(k) = a * \pi * \text{rand} \quad (12)$$

The location of the bald eagle is decided using $m(k)$ and $h(k)$, which have values ranging from 0 to 1 and are used to calculate the eagle's polar coordinates. The parameters $r(i)$ and $\theta(k)$ represent spiral flight, respectively. β ranges between 5 and 10. R ranges from 0.2 to 2. The random number, rand , falls between 0 and 1.

3.3 Swooping Stage (Exploitation Phase)

In this case, the eagle traverses toward the prey identified from the best location. Additionally, all the solutions make its movement toward the best location as well. This process can be described as follows:

$$Z_{i,\text{new}} = \text{rand} * Z_{\text{best}} + m1(k) * (Z_i - c1 * Z_{\text{mean}}) + h1(k) * (Z_i - c2 * Z_{\text{best}}) \quad (13)$$

$$m1(k) = \frac{mr(k)}{\max(|mr|)}, \quad h1(k) = \frac{hr(k)}{\max(|hr|)} \quad (14)$$

$$hr(k) = r(k) * \cosh[\cos^{-1} \theta(k)] \quad (15)$$

$$mr(k) = r(k) * \cosh[\theta(k)], mr(k) = r(k) * \sinh[\theta(k)] \quad (16)$$

$$\theta(k) = a * \pi * \text{rand} \quad \text{and} \quad r(k) = \theta(k)$$

where $c1, c2 \in [1, 2]$.

The movement concentration of the eagles toward the central locations is increased by the parameters $c1$ and $c2$, which have values between 1 and 2. A flowchart of the BES optimizer is presented in Fig. 1.

4 Results and Discussion

The application of BES to manage congestion by rescheduling the generators has been demonstrated on 39 bus New England Test System which includes 10 generators and 29 load buses [14] and is shown in Fig. 2

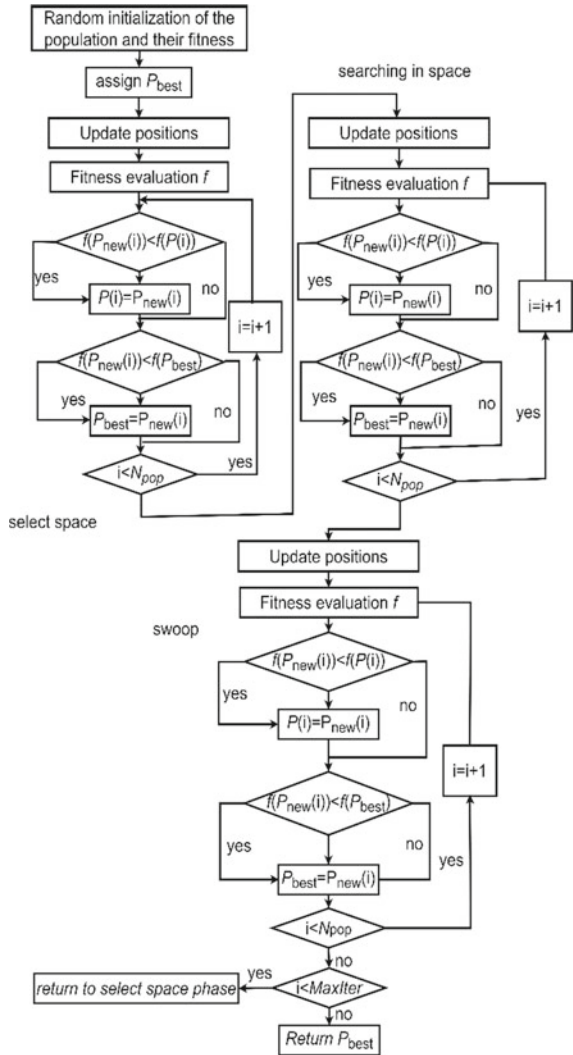
The line L14-34 has a power flow of 262.3 MVA, which is well below its flow limit of 600 MVA. The congestion in the system is created when the L14-34 has been tripped, the flow on line L15-16 between buses 16 and 15 exceeds its limit, reaching 628.6 MVA compared to the limit of 500 MVA. Load flow analysis has been used to detect this issue of overloading.

Table 1 presents the GSF values. The GSF of generators G4 to G7 are uniformly set to -0.35 , indicating that these four generators do not contribute to congestion management thus the contributing generators for the CM are the G2, G3, G8, G9, G10.

Table 2 highlights the outcomes achieved with the application of BES for the CM problem. The outcomes of BES are then compared to the findings with RED [7], GSA [9], ABC [12], PSO [15]. The congestion cost achieved with BES is 7943.91 \$/h and is minimum when compared to RED, PSO, ABC, and GSA. The congestion cost comparison is shown in Fig. 3. The rescheduling quantities for generators 3, 8, and 10 have been compared with different methods, as illustrated in Fig. 4. The convergence characteristics for the congestion cost minimization with BES has been shown in Fig. 5.

Table 3 reveals that rescheduling using BES leads to enhancements in the minimum voltage of the system and reductions in active power losses. The voltage profile of the system after BES-based rescheduling is comparatively better than the results reported in RED, ABC, PSO. It is also observed that the system losses have been reduced. The system loss achieved with BES post CM is 57.96 MW and it is noted that the system losses achieved with BES is comparatively lower than the other techniques. The voltage at the buses after CM with BES is represented in Fig. 6.

Fig. 1 BES optimizer flowchart for CM



5 Conclusion

In this study, an approach of rescheduling generator’s active power is considered for CM. The generators for CM are selected based on GSF. The selected generators are rescheduled using BES to mitigate congestion. The utilization of the GSF helps to minimize the number of participating generators. The BES-based solution yields superior results, and the solution demonstrates system stability with reduced redispatch costs. The BES algorithm requires tuning of very few parameters, namely the

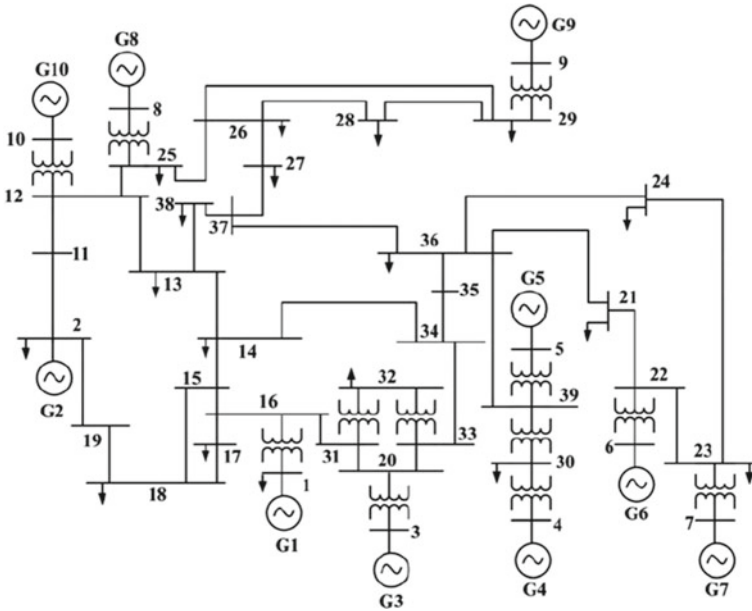


Fig. 2 Representation of 39 bus system

Table 1 GSF data for 39-bus New England system

Gen. No	01	02	03	04	05
GSF	0.00	- 0.47	- 0.04	- 0.35	- 0.035
Gen. No	06	07	08	09	10
GSF	- 0.35	- 0.35	- 0.49	- 0.44	- 0.51

colony and this simplicity in parameter tuning increases the likelihood of finding the most effective values compared to other metaheuristic algorithms.

Table 2 Comparison evaluation result with proposed method for 39-bus New England system

Parameters	Method				
	RED [16]	PSO [17]	ABC [12]	GSA [9]	BES [proposed]
Rescheduling cost(\$/h)	8639.17	8872.9	8456	8033	7944.28
Power flow of L 15–16 (MW) after CM	510	490	499.50	499.10	498.32
ΔP_1 (MW)	– 99.59	– 149.1	– 131.0	– 130.82	– 137.68
ΔP_2 (MW)	98.75	65.6	63.2	46.43	72.34
ΔP_3 (MW)	– 159.64	– 129	– 132.0	– 124.69	– 115.60
ΔP_4 (MW)	12.34	NP	NP	NP	NP
ΔP_5 (MW)	24.69	NP	NP	NP	NP
ΔP_6 (MW)	24.69	NP	NP	NP	NP
ΔP_7 (MW)	12.34	NP	NP	NP	NP
ΔP_8 (MW)	24.69	75.4	72.2	88.90	46.18
ΔP_9 (MW)	12.34	52.1	49.1	47.95	93.78
ΔP_{10} (MW)	49.38	83.0	78.8	72.26	36.54
Total cost (MW)	518.45	554.2	526.3	518.45	502.1

Fig. 3 Congestion cost comparison with BES

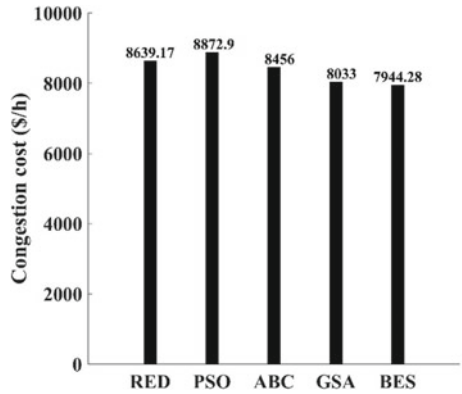


Fig. 4 Rescheduled power comparison with BES

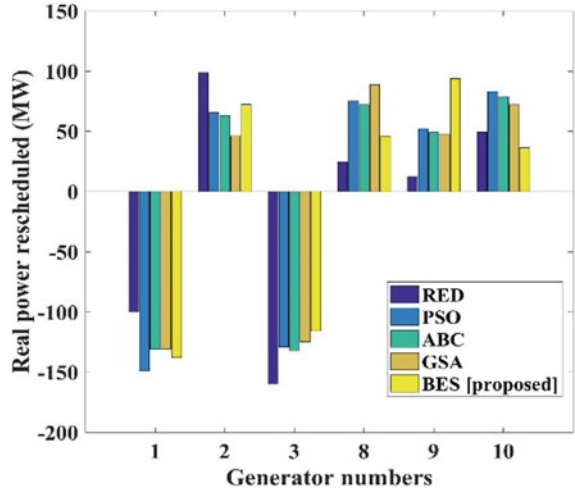


Fig. 5 Convergence profile of BES for CM

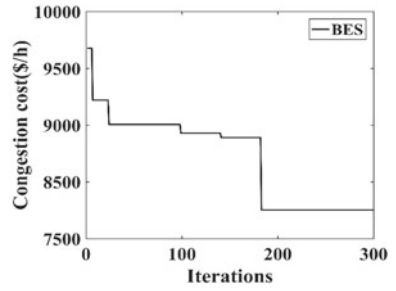
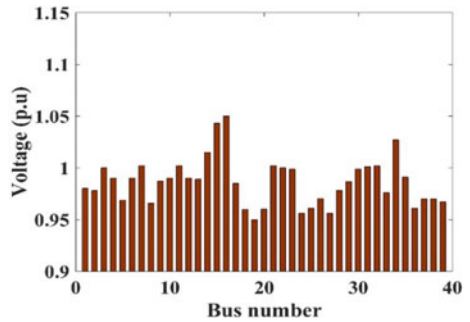


Table 3 Representation of system voltage and losses

	Congested state	RED [7]	PSO [8]	ABC [9]	BES [Proposed]
V_{min} (p.u.)	0.941	0.945	0.932	0.940	0.955
P_{loss} (MW)	59.64	57.31	58.00	59.00	57.96

Fig. 6 Voltage at the buses post CM with BES



References

1. Gupta A, Verma YP, Chauhan A (2023) Wind-hydro combined bidding approach for congestion management under secured bilateral transactions in hybrid power system. *IETE J Res* 69:354–367
2. Dehnavi E, Afsharnia S, Akmal AAS, Moeini-Aghaie M (2022) A novel congestion management method through power system partitioning. *Electr Power Syst Res* 213:108672
3. Pantoš M (2020) Market-based congestion management in electric power systems with exploitation of aggregators. *Int J Electr Power Energy Syst* 121:106101
4. Paul K (2022) Multi-objective risk-based optimal power system operation with renewable energy resources and battery energy storage system: a novel hybrid modified grey wolf optimization–Sine Cosine Algorithm approach. *Trans Inst Meas Contr* 01423312221079962
5. Sarwar M, Siddiqui AS, Ghoneim SS, Mahmoud K, Darwish MM (2022) Effective transmission congestion management via optimal DG capacity using hybrid swarm optimization for contemporary power system operations. *IEEE Access* 10:71091–71106
6. Asrari A, Ansari M, Khazaei J, Fajri P (2019) A market framework for decentralized congestion management in smart distribution grids considering collaboration among electric vehicle aggregators. *IEEE Trans Smart Grid* 11:1147–1158
7. Yesuratnam G, Thukaram D (2007) Congestion management in open access based on relative electrical distances using voltage stability criteria. *Electr Power Syst Res* 77:1608–1618
8. Hazra J, Sinha A, Phulpin Y (2009) Congestion management using generation rescheduling and/or load shedding of sensitive buses. In: 2009 international conference on power systems, pp 1–5
9. Paul K, Kumar N, Agrawal S, Paul K (2019) Optimal rescheduling of real power to mitigate congestion using gravitational search algorithm. *Turk J Electr Eng Comput Sci* 27:2213–2225
10. Zaeim-Kohan F, Razmi H, Doagou-Mojarrad H (2018) Multi-objective transmission congestion management considering demand response programs and generation rescheduling. *Appl Soft Comput* 70:169–181
11. Charles Raja S, Prakash S, Jeslin Drusila Nesamalar J (2021) Effective power congestion management technique using hybrid Nelder–Mead–grey wolf optimizer (HNMGWO) in deregulated power system. *IETE J Res* 1–12
12. Deb S, Goswami AK (2012) Congestion management by generator rescheduling using artificial bee colony optimization technique. In: 2012 annual IEEE India conference (INDICON), pp 909–914
13. Paul K, Sinha P, Bouteraa Y, Skruch P, Mobayen S (2023) A novel improved Manta Ray foraging optimization approach for mitigating power system congestion in transmission network. *IEEE Access* 11:10288–10307
14. Padiyar K (2010) Power system dynamics: stability & control. BS Publications
15. Dutta S, Singh S (2008) Optimal rescheduling of generators for congestion management based on particle swarm optimization. *IEEE Trans Power Syst* 23:1560–1569
16. Yesuratnam G, Thukaram D (2007) Congestion management in open access based on relative electrical distances using voltage stability criteria, vol 77, pp 1608–1618
17. Dutta S, Singh SP (2008) Optimal rescheduling of generators for congestion management based on particle swarm optimization, vol 23, pp 1560–1569

Optimal Power Flow Analysis Using Pelican Optimization Algorithm



Prakash Kumar, Md. Abul Kalam, and Kaushik Paul

Abstract This research paper introduces a highly effective and dependable method for addressing the optimal power flow (OPF) issues with application of an evolutionary-based approach. In this research work, Pelican Optimization Algorithm (POA) has been applied to compute the optimal values and parameters related to the OPF problem. The efficient exploration and exploitation phases of POA enhances the evaluation of the OPF fitness function without getting stuck in the local optima. The effectiveness of POA for the OPF has been evaluated on IEEE 30-bus system, considering various goals likes minimizing fuel cost, improving voltage profile, and enhancing voltage stability. The outcomes highlighted that POA demonstrated promising results in comparison to the latest optimization techniques that has been implemented for OPF solutions.

Keywords Optimal power flow · Fuel cost · Control variables · Pelican optimization algorithm · Voltage stability

1 Introduction

In power system operation, the aspect of optimal power flow (OPF) plays a vital role that enhances the operating state of the electric power system. Power systems are complex networks consisting of generation sources, transmission lines, and distribution networks that deliver electricity to consumers [1]. The efficient operation of these systems is vital to ensure reliability, minimize costs, and maximize the utilization of available resources. OPF algorithms and techniques play a pivotal role in

P. Kumar · Md. Abul Kalam · K. Paul (✉)
Department of Electrical Engineering, BIT Sindri, Dhanbad, India
e-mail: kaushik.ee@bitsindri.ac.in

P. Kumar
e-mail: kumarprakashsharma2018@gmail.com

Md. Abul Kalam
e-mail: kalam.ee@bitsindri.ac.in

optimizing the generation and transmission of electric power, taking into account various operational constraints and objectives [2]. The primary goal of OPF is to find out the optimal settings for control variables like generator output levels, transformer tap positions, and power flow in transmission lines, among others, to achieve the best possible system performance.

The optimization problem in OPF involves minimizing or maximizing an objective function while adhering to a set of restrictions. The objective function typically represents an economic measure, such as minimizing the total generation cost or maximizing the social welfare, by considering factors such as fuel costs, power losses, and environmental impact [3]. The constraints encompass a range of physical and operational limitations, including generator capacity limits, voltage limits, line flow limits, and reserve requirements.

OPF algorithms tackle the optimization issue by using mathematical optimization methods such mixed-integer programming, nonlinear programming, and linear programming. These algorithms utilize mathematical models that represent the power system's behavior, including power flow equations, generator cost curves, and network constraints [4]. By solving the OPF problem, engineers and system operators can obtain optimal operating conditions that improve system efficiency, reliability, and cost-effectiveness. The application of OPF is widespread in power system planning, real-time operation, and market operations. In the planning phase, OPF is utilized to determine optimal expansion plans, generator siting, and transmission network reinforcement, considering factors like load growth, generation mix, and environmental regulations. In real-time operation, OPF is used to schedule generation dispatch, regulate voltages, and manage transmission congestion [5]. OPF is also critical in electricity markets, where it helps optimize power flows and determine optimal generation offers and market clearing prices.

There has been significant progress in the field of OPF by the power system researchers. Marcelino et al. proposed a cross-entropy OPF approach for smart grids, aiming to enhance the efficiency of power distribution system [6]. Ali et al. addressed the multi-objective OPF problem with constraints by employing a hybrid multi-objective evolutionary algorithm. Their approach provides a solution that incorporates multiple objectives [7]. Esteban-Pérez and Morales introduced a distributionally robust optimal power flow model with contextual information to enhance the robustness of power flow optimization in uncertain scenarios [8].

Avvari et al. considered a hybrid power system with renewable energy sources and formulated a multi-objective-based optimization approach for OPF offering an optimized solution considering multiple objectives [9]. Fathabad et al. proposed asymptotically tight conic approximations to solve chance-constrained AC optimal power flow problems, enabling efficient power flow optimization with probabilistic constraints [4]. Ramesh et al. developed an OPF solution using meta-heuristic algorithms for the deregulated electricity power market, considering load congestion and aiming for improved market operation [10]. Zhang et al. tackled OPF problem in hierarchical partition-managed power systems by utilizing distributed power flow calculations, enabling efficient management of power systems [11]. Jithendranath

and Das focused on OPF with solar PV generation in islanded microgrids, and applied PSO to optimize power flow considering multiple objectives [12].

This research paper introduces a POA-based solution methodology for the OPF problem. Three objectives are taken into concern for the OPF issue which are the fuel cost minimization, enhancement in the voltage profile, and improvement in the power system voltage stability. The IEEE 30-bus standard system is used to assess the suggested method. Comparative analysis with existing literature results is also presented.

2 Problem Formulation of Optimal Power Flow

The OPF is a nonlinear and multifaceted optimization problem with the goal of achieving the ideal operative condition for a power system. Mathematically the OPF problem is stated in Eq. (1) [13].

$$\text{Minimize } f(c, d) \quad (1)$$

$$\text{Subject to: } g(c, d) = 0 \quad (2)$$

$$h(c, d) \leq 0 \quad (3)$$

where $f(c, d)$ is the objective function that can be minimized in the OPF problem; $g(c, d)$ is equality constraints and $h(c, d)$ is the inequality constraints. c is the dependent variables. It is governed by the control variables and alter when the control parameters vary. The state variables are computed after the control variables are determined.

$$c = [P_{G_1}, Q_{G_1}, \dots, Q_{G_{nG}}, V_{L_1}, \dots, V_{L_{nL}}, S_{l_1}, \dots, S_{l_{nl}}] \quad (4)$$

where $P_{G_1}, Q_{G_1}, \dots, Q_{G_{nG}}, V_{L_1}, \dots, V_{L_{nL}}$, and $S_{l_1}, \dots, S_{l_{nl}}$ are the swing bus power, the reactive power of the generator bus, the load node voltages and the transmitted power through the line, respectively. nG, nl , and nL are number of generators, transmission lines, and load bus, respectively. d is the control variables also known as the independent variables.

$$d = [P_{G_2}, \dots, P_{G_{nG}}, Q_{C_1}, \dots, Q_{C_{nC}}, V_{G_1}, \dots, V_{G_{nG}}, T_1, \dots, T_{nT}] \quad (5)$$

where $P_{G_2}, \dots, P_{G_{nG}}, Q_{C_1}, \dots, Q_{C_{nC}}, V_{G_1}, \dots, V_{G_{nG}}$, and T_1, \dots, T_{nT} and are the active power generated by the generator, the capacity of the parallel reactive capacitor, node voltage, and the tap location of the changeable transformer, respectively. nT and nC are number of transformer and number of reactive VAR capacitor, respectively. Active and reactive power balance can be represented as:

$$\sum_{m=1}^{nGB} P_{G_m} - P_D - P_L = 0, \quad m = 1, \dots, nGB \quad (6)$$

$$\sum_{m=1}^{nGB} Q_{G_m} - Q_D - Q_L = 0, \quad m = 1, \dots, nGB \quad (7)$$

where P_D and P_L are active power demand and loss, Q_D is the reactive power demand and Q_L represents the reactive power loss of the specified system. Generator bus count is expressed as nGB.

(a) Generation boundaries:

$$P_{G_m}^{\min} \leq P_{G_m} \leq P_{G_m}^{\max}, \quad m = 1, \dots, nG \quad (8)$$

$$Q_{G_m}^{\min} \leq Q_{G_m} \leq Q_{G_m}^{\max}, \quad m = 1, \dots, nG \quad (9)$$

$$V_{G_m}^{\min} \leq V_{G_m} \leq V_{G_m}^{\max}, \quad m = 1, \dots, nG \quad (10)$$

where $P_{G_m}^{\min}$, $Q_{G_m}^{\min}$, and $V_{G_m}^{\min}$ are corresponding lower bounds; $P_{G_m}^{\max}$, $Q_{G_m}^{\max}$, and $V_{G_m}^{\max}$ are corresponding upper bounds.

(b) Transformer boundaries:

$$T_m^{\min} \leq T_m \leq T_m^{\max}, \quad m = 1, \dots, nT \quad (11)$$

(c) Shunt compensator capacity boundaries:

$$Q_{C_m}^{\min} \leq Q_{C_m} \leq Q_{C_m}^{\max}, \quad m = 1, \dots, nC \quad (12)$$

(d) Safety boundaries:

$$V_{L_m}^{\min} \leq V_{L_m} \leq V_{L_m}^{\max}, \quad m = 1, \dots, nL \quad (13)$$

$$S_{l_m} \leq S_{L_m}^{\max}, \quad m = 1, \dots, nl \quad (14)$$

The objective function with the penalty factors λ_p , λ_q , λ_v , and λ_s can be written as given below:

$$\begin{aligned} f_{\text{Aug}} = & f + \lambda_p (P_{G_1} - P_{G_1}^{\text{Lim}})^2 + \lambda_v \sum_{m=1}^{nL} (V_{L_m} - V_{L_m}^{\text{Lim}})^2 \\ & + \lambda_q \sum_{m=1}^{nG} (Q_{G_m} - Q_{G_m}^{\text{Lim}})^2 + \lambda_s \sum_{m=1}^{nl} (S_{l_m} - S_{l_m}^{\text{Max}})^2 \end{aligned} \quad (15)$$

3 Pelican Optimization Algorithm (POA)

The POA has been formulated by Trojovsky and Dehghani to solve the optimization problem in different scientific field [14]. The hunting behavior and strategic approach of pelicans are intelligent processes that have honed their skills as hunters. The POA takes inspiration from this specific strategy, utilizing it as a model to guide its development. The suggested swarm-based POA is provided in this section along with its mathematical model and source of inspiration.

Proposed POA's mathematical model

In POA, the pelicans represent the members of the population and can be expressed as in Eq. (16):

$$z_{i,j} = l_j + \text{rand} \cdot (u_j - l_j), \quad i = 1, 2, \dots, m \quad (16)$$

where l_j and u_j denotes upper and lower bounds, respectively. N is a count of population members, rand is a random number in the range [0,1], and $z_{i,j}$ represents the candidate solution.

The population matrix for POA can be represented as in Eq. (17):

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_i \\ \vdots \\ Z_N \end{bmatrix}_{N \times m} = \begin{bmatrix} z_{1,1} & \cdots & z_{1,j} & \cdots & z_{1,m} \\ \vdots & \ddots & \vdots & & \vdots \\ z_{i,1} & \cdots & z_{i,j} & \cdots & z_{i,m} \\ \vdots & & \vdots & \ddots & \vdots \\ z_{N,1} & \cdots & z_{N,j} & \cdots & z_{N,m} \end{bmatrix}_{N \times m} \quad (17)$$

Each pelican in POA represents a population member and a possible solution to the problem. Assessment of the goal function is done based on the comparative feasible solution achieved from the POA. The goal function represented according to the feasible solution can be represented as in Eq. (18):

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N \times 1} = \begin{bmatrix} F(Z_1) \\ \vdots \\ F(Z_i) \\ \vdots \\ F(Z_N) \end{bmatrix}_{N \times 1} \quad (18)$$

where F_i is the value of the goal function for the i th candidate solution.

There are two parts to this hunting technique as follows:

Approaching prey (exploration phase)

In POA approach, the pelicans locate and approach the prey's position. This pelican strategy serves as a model for inspecting the region of the search area that improves the exploration capabilities. Equation (19) mathematically represents the aforementioned concept.

$$z_{i,j}^{P_1} = \begin{cases} z_{i,j} + \text{rand} \cdot (p_j - I \cdot z_{i,j}), & F_P < F_i \\ z_{i,j} + \text{rand} \cdot (z_{i,j} - p_j), & \text{else} \end{cases} \quad (19)$$

where p_j is the location of the prey, F_P resembles the objective function, and $z_{i,j}^{P_1}$ is current computed position.

If the value of the fitness function is enhanced in the new position for a pelican, the POA updates itself with the most optimal solution corresponding to the new position and restricts within the search boundary using Eq. (20).

$$X_i = \begin{cases} X_i^{P_1}, & F_i^{P_1} < F_i \\ X_i, & \text{else} \end{cases} \quad (20)$$

where $X_i^{P_1}$ represents the current position.

Winging on the surface of the water (exploitation phase)

In this step, the pelicans stretch their wings on the water's surface to lift the fish upward and then scoop them up in their throat pouches. More fish are taken by pelicans using this tactic in the region that is being targeted. This procedure improves POA's capability for exploitation and local search. Equation (21) uses mathematics to model the pelican's hunting activity.

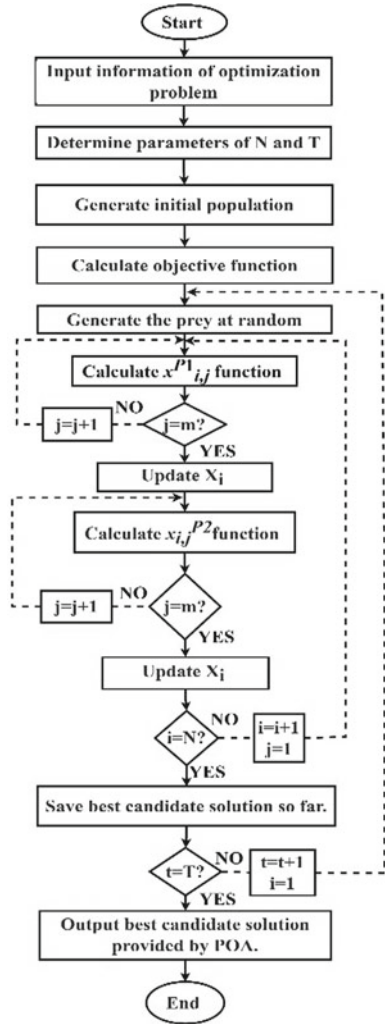
$$Z_{i,j}^{P_2} = z_{i,j} + R \cdot \left(1 - \frac{t}{T}\right) \cdot (2 \cdot \text{rand} - 1) \cdot z_{i,j} \quad (21)$$

where $z_{i,j}^{P_2}$ is the updated status, $R = 0.2$, $R \cdot \left(1 - \frac{t}{T}\right)$ coefficient helps the POA's exploitation power approach the best overall solution. As a result of this coefficient's initial high value, each member's immediate surroundings are taken into account in greater detail. The " $R \cdot \left(1 - \frac{t}{T}\right)$ " coefficient drops as the method replicates more, leading to decreasing radii of each member's neighborhood. Effective updating of the candidate solution is represented by Eq. (22).

$$X_i = \begin{cases} X_i^{P_2}, & F_i^{P_2} < F_i \\ X_i, & \text{else} \end{cases} \quad (22)$$

In Eq. (22) $X_i^{P_2}$ is latest position and $F_i^{P_2}$ is the fitness function (The POA flowchart has been given in Fig. 1).

Fig. 1 POA flowchart



4 Results and Discussion

The POA has been applied to IEEE 30 bus system to compute and evaluate the fitness function related to fuel cost, voltage stability index, and voltage deviation. The population count for POA has been taken as 30 and the maximum iteration considered is 300. The conventional IEEE 30-bus system depicted in Fig. 2 [13]. Table 1 represents the generators cost for IEEE 30 bus system. The bus systems consist of a connected bus, 24 PQ loads, and 6 generators.

The following cases have been taken into consideration in order to illustrate the efficacy and durability of the technique:

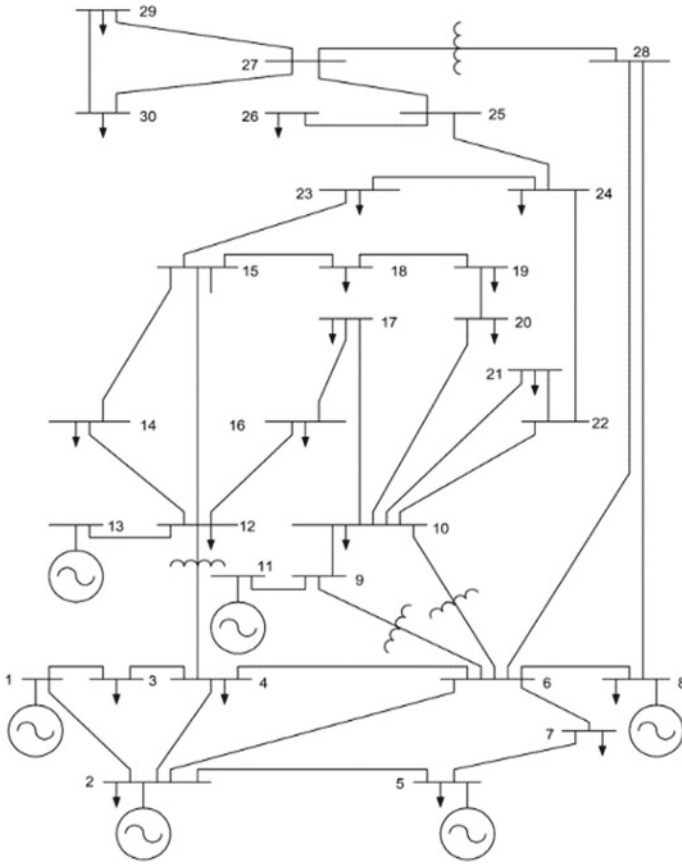


Fig. 2 IEEE 30-bus system (IEEE 30 bus)

Table 1 Generator cost coefficient

Bus no.	Generator cost coefficient		
	<i>a</i>	<i>b</i>	<i>c</i>
1	0.00375	2	0
2	0.0175	1.75	0
5	0.0625	1	0
8	0.00834	3.25	0
11	0.025	3	0
13	0.025	3	0

Case 1: Generations Cost Minimization

Here, the whole expense of fuel is taken into account as the objective function f , i.e.,

$$f = \sum_{m=1}^{nG} f_m \text{ (\$/h)} \tag{23}$$

where f_m is the m th generator fuel cost.

The quadratic functions used to illustrate the generator expense curves as

$$f_m = c_m + b_m P_{G_m} + a P_{G_m}^2 \text{ (\$/h)} \tag{24}$$

where c_m , b_m , and a_m are they m th generator cost coefficient is represented Table 1. Figure 3 the variance in the overall fuel cost is depicted. Table 2 provides the control variables most suitable settings. By using the POA method, a cost of 792.40\$/h is obtained. It is obvious that fuel costs have decreased significantly. A comparative analysis of the fuel cost with other optimization techniques of the OPF problem has been compared in Table 3. The cost achieved with POA has been better in comparison to the cost achieved with the optimization used in the referred literature.

Case 2: Strengthening of the Voltage Profile

The system bus voltages are the most vital parameters that signify the safety and reliability of power system operations. Only taking cost-based goals into account in an optimization problem may lead to a workable solution with an undesirable voltage profile. By reducing the voltage across the load bus variations from 1.0 per unit, a two-fold aim function is put forward in this situation to reduce fuel costs and enhance voltage profiles. The fitness function is considered as:

Fig. 3 Variance of overall fuel cost

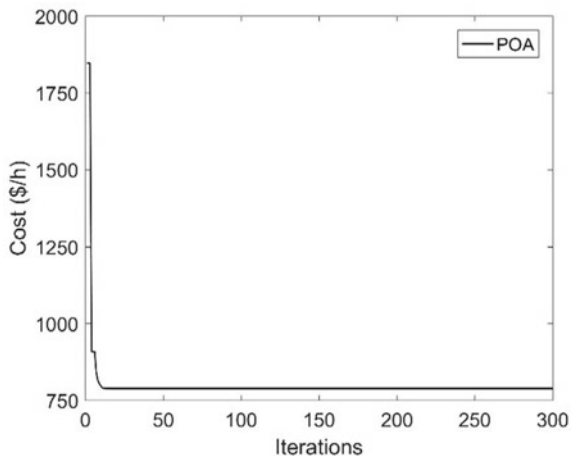


Table 2 The IEEE-30 bus system network's optimal control variable setting

Control variables	Initial [13]	Case 1 (cost)	Case 2 (cost + voltage profile improvement)	Case 3 (cost + voltage stability enhancement)
P_{G_1}	0.9921	0.2092	0.5500	0.4898
P_{G_2}	0.8000	0.2450	0.2415	0.2130
P_{G_5}	0.5000	0.1151	0.3500	0.2119
P_{G_8}	0.2000	0.1200	0.1851	0.1197
$P_{G_{11}}$	0.2000	1.7553	0.1779	1.7553
$P_{G_{13}}$	0.2000	0.1200	1.4000	0.4798
V_{G_1}	1.0500	1.0744	1.0236	1.0799
V_{G_2}	1.0400	1.0858	1.0054	1.0738
V_{G_5}	1.0100	1.0493	1.0216	1.0376
V_{G_8}	1.0100	1.0486	1.0350	1.0405
$V_{G_{11}}$	1.0500	1.0790	1.0496	1.0342
$V_{G_{13}}$	1.0500	1.0421	1.0175	1.0200
T_{4-12}	1.0233	0.9694	1.0702	1.0780
T_{6-9}	0.9557	0.9238	0.9000	1.0690
T_{6-10}	0.9724	0.9467	0.9954	1.0320
T_{27-28}	0.9728	0.9820	0.9703	1.0680
$Q_{C_{10}}$	0.0	0.0403	0.9573	0.0198
$Q_{C_{12}}$	0.0	0.0369	1.0120	0.0315
$Q_{C_{15}}$	0.0	0.0500	0.0162	0.0454
$Q_{C_{17}}$	0.0	0.0198	0.0424	0.0381
$Q_{C_{20}}$	0.0	0.0315	0.0256	0.0500
$Q_{C_{21}}$	0.0	0.0454	0.0465	0.0500
$Q_{C_{23}}$	0.0	0.0488	0.0335	0.0403
$Q_{C_{24}}$	0.0	0.0500	0.0220	0.0369
$Q_{C_{29}}$	0.0	0.0500	0.0198	0.0500
Fuel cost (\$/h)	901.88	792.40	796.92	793.15
$\sum V.D$	1.1554	0.8593	0.0867	0.9514
L_{Max}	0.1681	0.1274	0.1381	0.1230

$$f = \sum_{m=1}^{nG} f_m + w \sum_{m \in nL} |V_m - 1.0| \quad (25)$$

where w represents the weighting components. Figure 4 depicts the variation in the overall fuel costs. The cost ached with POA in case 2 is 796.92 \$/h. The voltage at each of the buses for the considered case is given in Fig. 5. The voltage patterns are appreciable for the considered optimization problem with POA. Particularly, as

Table 3 Comparison of fuel cost with POA

Applied optimization approaches	Fuel cost (\$/h)	Ref.
POA [proposed]	792.40	
EACS	800.79	[15]
BHOA	800.89	[16]
GA	800.96	[17]
PSO	801.66	[18]
AGSO	801.75	[19]
ICA	801.78	[20]
MSFL	802.29	[21]
Modified DE	802.37	[22]
EP	802.62	[23]
IPSO	802.63	[23]

shown in Table 2 the total amount of deviation in voltage has decreased from 0.8593 in Case 1 to 0.0867 in Case 2.

Case 3: The Improvement of Voltage Stability

A key feature of the power system is its capacity to continuously maintain a suitable node voltage during the normal operating condition along with the small increment in the load, change in the power system topology or when the network is experiencing a minor disturbance. The fitness function formulated in this case aims toward fuel cost minimization while considering the improvement in the voltage stability. L-index has been taken to measure the voltage stability that ranges from 0 (without loss) to 1 (with loss).

Fig. 4 Variation of overall fuel cost with voltage profile improvement

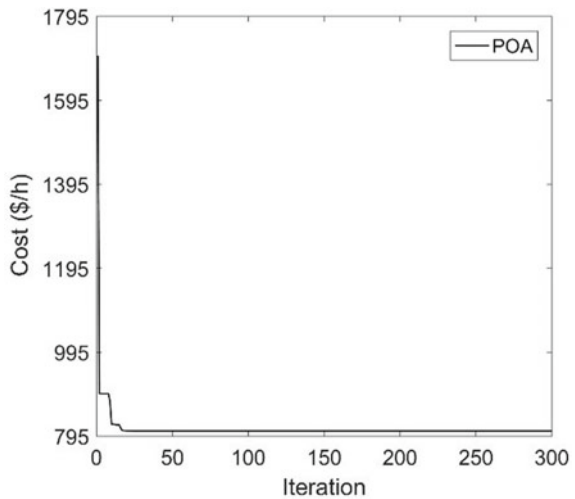
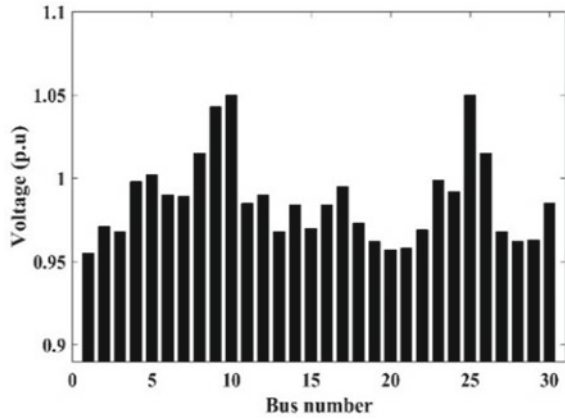


Fig. 5 Voltage at the buses



The subsequent fitness function is suggested in direction to boost the stability of voltage and transfer the entire system out of the voltage collapse position as

$$f = \sum_{m=1}^{nG} f_m + wL_{Max} \tag{26}$$

where w represents the weighting component and L_{Max} is represented as:

$$L_{Max} = \text{Max}\{L_m, m = 1, \dots, nL\} \tag{27}$$

Table 2 provides the control variables' ideal settings. The fuel cost achieved in this case is 793.15 \$/h. The fuel cost convergence is depicted in Fig. 6. The table shows that in this situation, the readings of L_{Max} at the load buses are drastically decreased. As a result, the voltage stability's separation from downfall has wide. The L_{Max} value is 0.1230 which is appreciable for the considered cases (as stated in Table 4).

Fig. 6 Variation of overall fuel cost with voltage enhancement

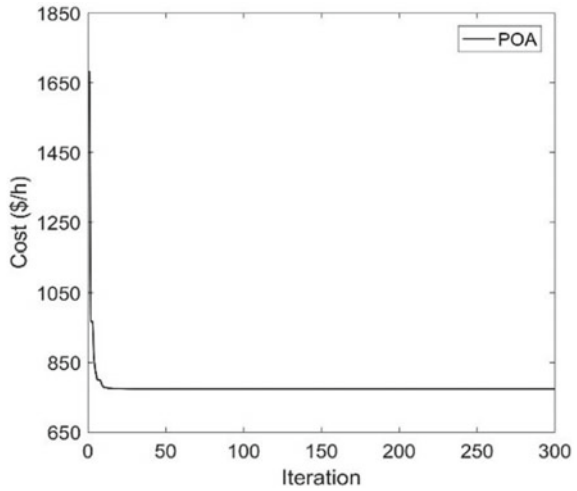


Table 4 Comparison of voltage stability index with POA

Optimization	Voltage stability index (L-index)	Ref.
POA [proposed]	0.1230	Proposed
PSO-SSO	0.124	[24]
AMTP-Jaya	0.1240	[25]
DE	0.1243	[26]
TLBO	0.12444	[25]
PSO	0.125	[24]
NISSO	0.12547	[27]
GEM	0.1257	[28]
SSO	0.1267	[27]
ECHT-DE	0.13632	[29]
SF-DE	0.13671	[29]
SP-DE	0.13743	[29]
HFPS	0.144	[30]

5 Conclusion

This work presents a revolutionary particle optimization-based solution to the OPF problem. The suggested POA method makes use of global and local search capabilities to look for the best control variable settings in order to reduce fuel costs, boost voltage stability, and improve voltage profile. The efficacy and robustness of the suggested strategy have been evaluated and examined with regard to various objectives. The outcomes highlight the impact of the proposed approach and is efficient in analyzing to some of the recent meta-heuristic algorithms.

References

1. Maheshwari A, Sood YR, Jaiswal S (2023) Flow direction algorithm-based optimal power flow analysis in the presence of stochastic renewable energy sources. *Electr Power Syst Res* 216:109087
2. Gao M, Yu J, Yang Z, Zhao J (2023) A physics-guided graph convolution neural network for optimal power flow. *IEEE Trans Power Syst*
3. Paul K (2022) Modified grey wolf optimization approach for power system transmission line congestion management based on the influence of solar photovoltaic system. *Int J Energy Environ Eng* 13(2):751–767
4. Fathabad AM, Cheng J, Pan K, Yang B (2023) Asymptotically tight conic approximations for chance-constrained AC optimal power flow. *Eur J Oper Res* 305(2):738–753
5. Paul K (2022) Multi-objective risk-based optimal power system operation with renewable energy resources and battery energy storage system: a novel hybrid modified Grey Wolf optimization–Sine Cosine Algorithm approach. *Trans Inst Meas Contr* 01423312221079962
6. Marcelino C, Pérez-Aracil J, Wanner E, Jiménez-Fernández S, Leite G, Salcedo-Sanz S (2023) Cross-entropy boosted CRO-SL for optimal power flow in smart grids. *Soft Comput* 27(10):6549–6572
7. Ali A, Abbas G, Keerio MU, Koondhar MA, Chandni K, Mirsaeidi S (2023) Solution of constrained mixed-integer multi-objective optimal power flow problem considering the hybrid multi-objective evolutionary algorithm. *IET Gener Transm Distrib* 17(1):66–90
8. Esteban-Pérez A, Morales JM (2023) Distributionally robust optimal power flow with contextual information. *Eur J Oper Res* 306(3):1047–1058
9. Avvari R, DM VK (2023) A novel hybrid multi-objective evolutionary algorithm for optimal power flow in wind, PV, and PEV systems. *J Oper Autom Power Eng* 11(2):130–143
10. Ramesh S, Verdú E, Karunanithi K, Raja S (2023) An optimal power flow solution to deregulated electricity power market using meta-heuristic algorithms considering load congestion environment. *Electr Power Syst Res* 214:108867
11. Zhang J, Cai J, Wang S, Li P (2023) Many-objective optimal power flow problems based on distributed power flow calculations for hierarchical partition-managed power systems. *Int J Electr Power Energy Syst* 148:108945
12. Jithendranath J, Das D (2023) Multi-objective optimal power flow in islanded microgrids with solar PV generation by NLTVMOPSO. *IETE J Res* 69(4):2130–2143
13. Abido MA (2002) Optimal power flow using particle swarm optimization. *Int J Electr Power Energy Syst* 24(7):563–571
14. Trojovský P, Dehghani M (2022) Pelican optimization algorithm: a novel nature-inspired algorithm for engineering applications. *Sensors (Basel)* 22(3)
15. Abdulasool AQ, Al-Bahrani LT (2021) Multi-objective constrained optimal power flow based on enhanced ant colony system algorithm. In: Presented at the 2021 12th international symposium on advanced topics in electrical engineering (ATEE)
16. Hatamlou A (2013) Black hole: a new heuristic optimization approach for data clustering. *Inf Sci* 222:175–184
17. Rotaru M, Sykulski J (2007) A study of evolutionary based optimal power flow techniques. *IEEE Xplore*, pp 1–6
18. Al-Bahrani L, Dumbrava V (2016) Optimal power flow based on particle swarm optimization. *78:253–264*
19. Daryani N, Hagh MT, Teimourzadeh S (2016) Adaptive group search optimization algorithm for multi-objective optimal power flow problem. *Appl Soft Comput* 38:1012–1024
20. Ghasemi M, Ghavidel S, Rahmani S, Roosta A, Falah H (2014) A novel hybrid algorithm of imperialist competitive algorithm and teaching learning algorithm for optimal power flow problem with non-smooth cost functions. *Eng Appl Artif Intell* 29:54–69
21. Niknam T, Narimani MR, Jabbari M, Malekpour AR (2011) A modified shuffle frog leaping algorithm for multi-objective optimal power flow. *Energy* 36(11):6420–6432

22. Sayah S, Zehar K (2008) Modified differential evolution algorithm for optimal power flow with non-smooth cost functions. *Energy Convers Manage* 49(11):3036–3042
23. Yang B, Chen Y, Zhao Z, Han Q (2006) Solving optimal power flow problems with improved particle swarm optimization. In: 2006 6th World Congress on intelligent control and automation, vol 2. IEEE, pp 7457–7461
24. El Sehiemy RA, Selim F, Bentouati B, Abido MJE (2020) A novel multi-objective hybrid particle swarm and salp optimization algorithm for technical-economical-environmental operation in power systems, vol 193, p 116817
25. Warid W (2020) Optimal power flow using the AMTPG-Jaya algorithm, vol 91, p 106252
26. Shaheen AM, El-Sehiemy RA, Farrag SM (2016) Transmission, and distribution, solving multi-objective optimal power flow problem via forced initialised differential evolution algorithm, vol 10, no 7, pp 1634–1647
27. Nguyen TT (2019) A high performance social spider optimization algorithm for optimal power flow solution with single objective optimization, vol 171, pp 218–240
28. Bouchekara H, Chaib A, Abido MA (2016) Multiobjective optimal power flow using a fuzzy based grenade explosion method, vol 7, pp 699–721
29. Biswas PP, Suganthan PN, Mallipeddi R, Amaratunga GA (2018) Optimal power flow solutions using differential evolution algorithm integrated with effective constraint handling techniques, vol 68, pp 81–100
30. Khan A, Hizam H, bin Abdul Wahab NI, Lutfi Othman MJ (2020) Optimal power flow using hybrid firefly and particle swarm optimization algorithm, vol 15, no 8, p e0235668

Performance Evaluation on Developed FRP Rods Used in Composite Insulators Using Rotating Wheel and Dip Facility



M.-Ramez Halloum , B. Subba Reddy , G. Nithin Reddy , Zoya Zaki, Febin Francis, and B. Kiran Kumar Reddy

Abstract The use of composite insulators in distribution and transmission power systems is increasing rapidly because of their features over conventional glass and ceramic insulators. However, many incidents of failure/fracture on composite insulators have been reported. Composite insulators being organic are more susceptible to degradation under the influence of electrical discharges, currents, and ultraviolet radiation. The fiber-reinforced plastic (FRP) rod is the essential element in composite insulators as it provides the required mechanical strength and internal electrical insulation. In this investigation, the performance of newly developed FRP rods is evaluated using the rotating wheel and dip experiment. This study is conducted to understand the degradation process of FRP rods under damp conditions with electrical stress (arcs, scintillations, and leakage current). We used two different types of FRP rods (normal and tracking-resistant) received from manufacturers. Visual inspection analysis shows the superior performance of the tracking-resistant FRP rod. In addition, detailed microscopic physical and chemical analyses such as Fourier transform infrared spectroscopy (FTIR), scanning electron microscope (SEM), and energy dispersive X-ray spectroscopy (EDS) are conducted on both FRP rod samples. Promising results are obtained and discussed.

Keywords Composite insulator · FRP rod · Electric field · Degradation · SEM · EDS · FTIR

M.-R. Halloum (✉) · B. Subba Reddy · G. Nithin Reddy · Z. Zaki · F. Francis · B. Kiran Kumar Reddy
Indian Institute of Science, Bengaluru 560012, India
e-mail: mramezh@iisc.ac.in

B. Subba Reddy
e-mail: sreddy@iisc.ac.in

G. Nithin Reddy
e-mail: nithinreddy@iisc.ac.in

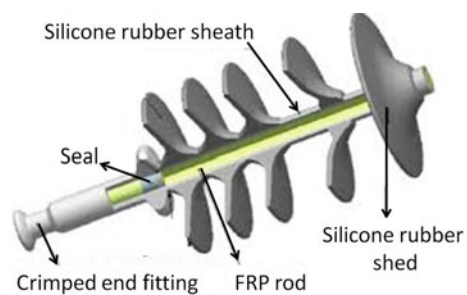
F. Francis
e-mail: febinfrancis@iisc.ac.in

1 Introduction

Composite outdoor insulators are presently used for extra-high voltage and ultra-high voltage power transmission systems due to their improved pollution characteristics. Composite insulators enjoy many advantages, such as low price, ease of installation, superior mechanical strength, and excellent hydrophobic nature [1]. Composite outdoor insulators consist of three main parts, the fiber-reinforced plastic (FRP) rod, silicone rubber (SiR) housing, and metal end-fittings, as shown in Fig. 1. The FRP rod plays an essential role as it provides the required mechanical strength and internal insulation, and it is made of glass fiber with the epoxy resin matrix. SiR housing protects the FRP rod against various stresses (electrical, environmental, etc.) and limits the surface leakage current by providing the required creepage distance [2]. SiR is made of polydimethylsiloxane (PDMS, known as silicone rubber) with added inorganic fillers such as fumed silica and alumina trihydrate to improve electrical, thermal, and mechanical properties [3]. Because of the organic nature of SiR housing, composite insulators are more susceptible to aging and degradation due to various stresses (dry and wet corona discharges, dry band discharges, leakage currents, ultraviolet radiation, high temperature, etc.).

However, it is felt researchers/manufacturers should give greater attention to the FRP rod, as a new type of fracture of the FRP rod has been reported recently in many countries [4–6], known as the decay-like fracture. This type of fracture is found to be different from normal or brittle fractures, as shown in Fig. 2. Normal fracture of FRP rods is due to the application of higher mechanical strength than the rated value. In contrast, brittle fracture is due to stress corrosion cracking (SCC) in the FRP rod. Stress corrosion cracking (SCC) in FRP rods may result from the penetration of water due to sealing failure followed by the formation of acids due to electrical discharge activities, according to research by Kuhl et al. [7] on the brittle fracture phenomena. When hardeners are present during the production of FRP rods, Tourreil et al. [8] noticed the formation of acids, and these acids can lead to brittle fracture. The failure of the end fitting sealant and water ingress were found to be the most frequent causes of brittle fracture, according to an IEEE Taskforce report [9] that investigated many brittle fracture incidents around the world.

Fig. 1 Main components of composite outdoor insulator



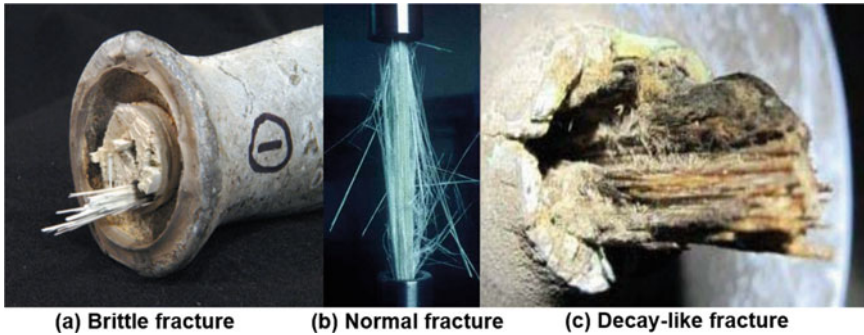


Fig. 2 Three different types of FRP rod fractures

The decay-like fracture was totally different from the brittle or normal fracture. A significant deterioration degree is observed in the epoxy resin matrix near the fracture spot, and the FRP rod looks like decayed wood. Also, the degradation of the FRP rod was found to be more in the external region than in the internal region [4]. The authors identified the failure reason because of poor interface adhesion between SiR housing and FRP rod and liquid penetration coupled with damp conditions and electrical discharges and currents resulting in stress corrosion and degradation of the epoxy resin matrix and, ultimately, FRP rod fracture [10]. The authors in [11] found that many air holes and bubbles are formed in the epoxy resin of decay-like fracture because of pyrolysis gas's generation, expansion, and release. Also, it is responsible for the initial formation of holes in the SiR housing during the degradation process. However, FRP rod fractures represent a significant concern for power utilities as they could lead to a complete line shutdown. Authors [12–14] conducted detailed experimental investigations on the failure analysis of the field FRP rods; further authors reported their findings using chaotic analysis and estimated the long-term degradation on the FRP rods. It is felt that additional experimental studies need to be done on FRP rod degradation to avoid their detrimental electrical consequences on power systems.

In this study, experimental investigations are conducted using rotating wheel and dip facility for 1046 h, primarily to evaluate the performance of two types of FRP rods (normal and developed tracking-resistant rods). This investigation is conducted to understand the degradation and aging process of the FRP rod under the action of surface discharges with damp conditions. The leakage current was continuously monitored during the experiment period. The physio-chemical analysis, including scanning electron microscope (SEM), energy dispersive X-ray spectroscopy (EDS), and Fourier transform infrared spectroscopy (FTIR), are carried out to evaluate the degradation level after the experiment ends for the samples.

2 Experiment Details

In this study, the performance of two different types of FRP rod (normal FRP rod (A) and developed tracking-resistant FRP rod (B)) are investigated using the rotating wheel and dip experiment as per IEC 62,217 and IEC/TR 62730 [15, 16]. The circuit diagram of the experiment setup with all details is shown in Fig. 3. However, this experiment is designed to test the tracking resistance of actual composite insulators with SiR housing. In this investigation, the bare FRP rods are used without SiR housing to evaluate the performance of the FRP rods under the worst-case scenario where the FRP rod will be directly exposed to the conducting water film. The tank was filled with tap water and NaCl to obtain the required conductivity (2.2 mS/cm was adopted in this study). Each cycle comprises four different stages: dipping, dripping, energization, and cooling. Initially, the sample is dipped in the electrolyte for 40 s. Then, the sample is shifted by 90° for 40 s to drip the water. Then, it is shifted again by 90° to apply the AC high voltage for another 40 s. Finally, the sample is shifted by 90 degrees for cooling. The transition time between the two stages is 8 s. Therefore, the total cycle time was 192 s. The experiment was conducted for up to 1046 h as per standards.

The leakage current was monitored during the HV energization stage. The dry arc distance of both FRP rods (A and B) is 250 mm. Therefore, as per standards, the applied voltage was calculated by dividing the dry arc distance (250 mm) by 27.6 mm/kV, i.e., 9.05 kV. The content of the epoxy resin matrix can be determined using the TGA, as the glass fiber is thermally stable up to 1600 °C compared to the epoxy resin with a lower thermal decomposition temperature of about 500 °C. Figure 4 shows TGA curves of the fresh and aged FRP rods (A and B). Normal FRP rod (A) shows higher content of epoxy resin matrix of 19.4% compared with tracking-resistant FRP rod (B) of 16.25%. The glass fiber content in A and B samples is 80.6% and 83.75%, respectively. In this study, the FRP rod was exposed directly to

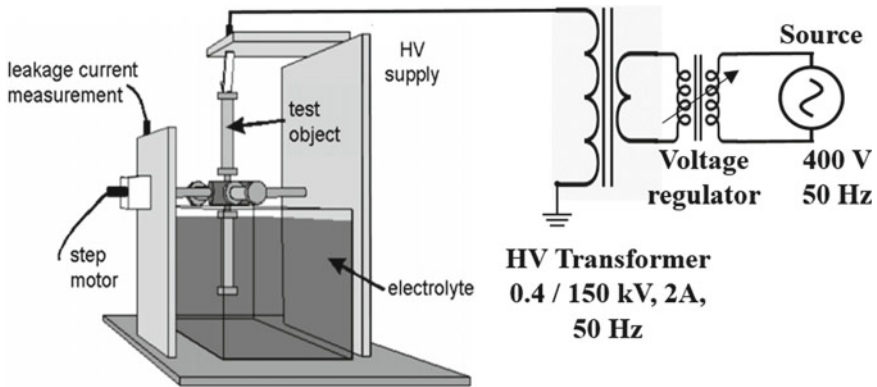
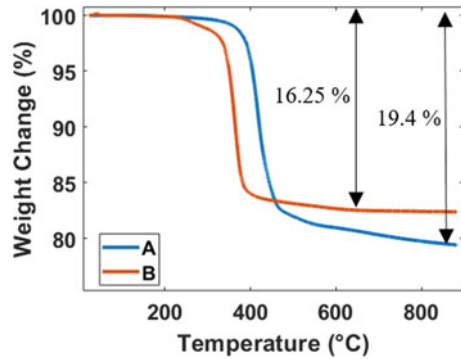


Fig. 3 Circuit diagram of rotating wheel and dip experimental facility

Fig. 4 TGA curves of both types of FRP rod (normal FRP rod (A) and tracking-resistant FRP rod (B))



water (damp conditions) coupled with leakage current flow, arcs, and scintillations, which are the main reasons for decay-like and brittle fractures.

3 Visual Observation and Leakage Current Analysis

Figure 5 shows the FRP rod samples (normal FRP rod (A) and tracking-resistant FRP rod (B)) before, during, and at the end of the experiment. Electrical discharges at the triple junction (near the HV end) are observed in sample A at 48 h. Also, dry band arcs are observed in sample A at 192 h. At 246 h, discharges at the HV triple junction of sample B are seen with less severity than sample A, as shown in Fig. 5. At 312 h, initialization of tracking was observed for sample A near the ground end. At 648 h, dry band arcs are observed in sample B. However, no tracking signs were observed on sample B till the experiment ended. The leakage current is monitored during the experiments at regular intervals for each sample. A shunt resistor (200 ohms) with RIGOL-made digital oscilloscope DS1102E (100MHz, 1 G Sample/s) is used to measure the leakage current at the ground side. Figure 6 shows the leakage current plot for FRP rod samples during the experiment (1046 h). Scintillations were observed in leakage current waveforms for both samples. The leakage current at the start of the experiments for samples A and B was 3.62 mA and 6.95 mA, respectively. However, the leakage current at the end of the investigation for samples A and B was 17.65 mA and 21.15 mA, respectively. That is mainly because of the decrement of the surface resistance due to FRP rod degradation under the action of electrical discharges. The leakage current of sample B increased by more than 204.3%. On the other hand, the leakage current of sample A increased by more than 387.56%, indicating severe degradation of the FRP rod.

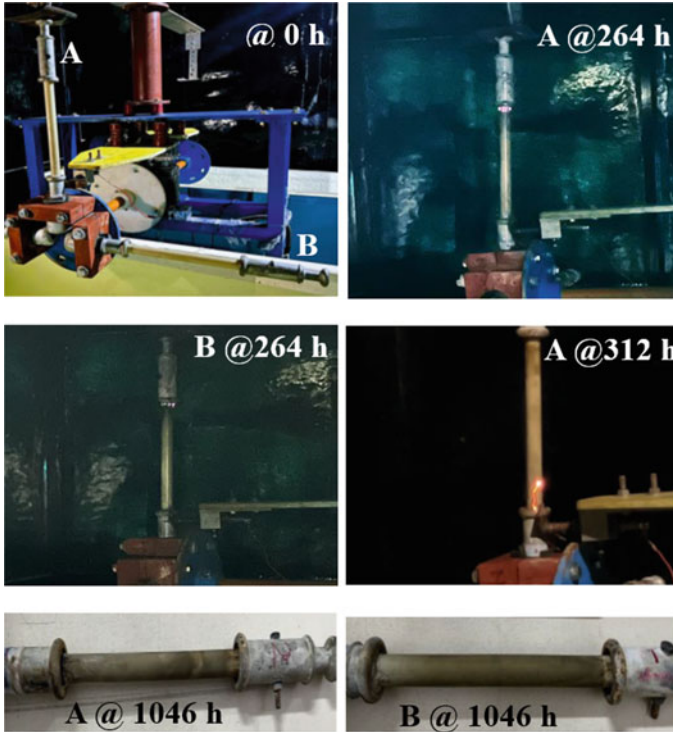


Fig. 5 FRP rod samples (normal FRP rod (A) and tracking-resistant FRP rod (B)) before, during, and at the end of the experiment

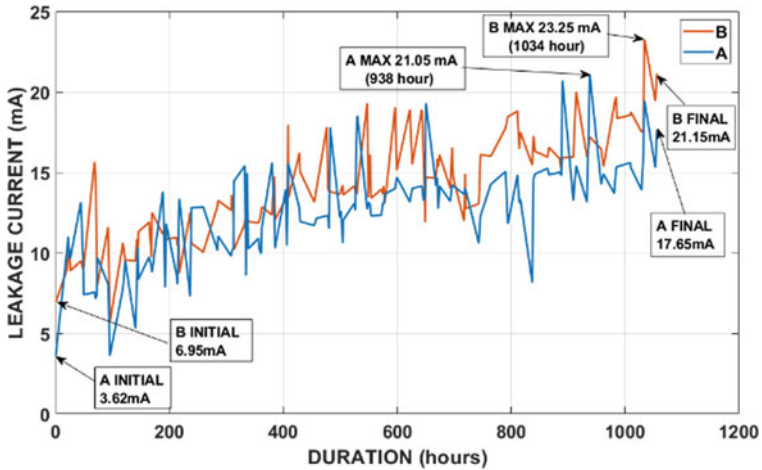


Fig. 6 Leakage current plot of FRP rod samples (normal FRP rod (A) and tracking-resistant FRP rod (B)) during the experiment

4 Physio-chemical Analysis

4.1 Scanning Electron Microscopy (SEM) and Energy Dispersive X-ray Spectroscopy (EDS)

Samples are collected after 1046 h to investigate the physio-chemical properties of the aged FRP rod samples (normal FRP rod (A) and tracking-resistant FRP rod (B)) and distinguish the degradation degree between these two samples. JEOL scanning electron microscopy (SEM IT 300) equipped with an energy dispersive X-ray detector is used to evaluate the morphological changes and the elemental composition on the surface of the FRP rod samples. Samples from the most degraded parts (near the ground end) were collected after visual observations, and a gold layer of 10 nm thickness was coated on the sample surfaces using the sputtering technique to avoid surface charging and better visualization.

Figure 7 shows the SEM images of the fresh and aged samples after 1046 h for both types of FRP rods (A) and (B). Table 1 shows the atomic elemental composition of fresh and aged FRP rod samples. As seen in Fig. 7, glass fiber is totally covered by the epoxy resin matrix in fresh samples. In addition, the interface between glass fiber and epoxy resin matrix is intact and well bonded. After 1046 h in the rotating wheel and dip experiment, the normal FRP rod (A) shows a significant degradation level compared with the new tracking-resistant FRP rod (B). Many broken glass fibers are seen and separated from the epoxy resin matrix. In addition, the remaining epoxy resin matrix in the normal FRP rod is found to be clastic and uneven with many holes, which could be formed as a result of the gas generated due to epoxy resin degradation under the effect of arcs, discharges, and leakage current. However, glass fiber in the tracking-resistant FRP rod is still well covered with epoxy resin matrix with a slightly more rough texture compared with the fresh sample.

A substantial decrement of carbon content in the normal FRP rod sample by more than 65% and an increment of oxygen content by more than 40% are noticed in Table 1. This is because of the decomposition of the epoxy resin matrix into water and carbon dioxide under the effect of electrical discharges and current [4]. However, carbon content decreased from 46.27 to 28.15% in the tracking-resistant FRP rod, with an improvement of about 26% compared to the normal FRP rod. Nitrogen is detected in both aged samples (A) and (B), which is evidence of nitric acid and amide formation. Nitric acid is formed under the impact of electrical discharges, where NO_2 is formed from oxygen and nitrogen from the medium. Then, NO_2 dissolved in water leading to nitric acid (HNO_3) formation. Also, amide is normally formed under the impact of electrical activities with oxygen and nitrogen presence [2]. The higher nitrogen content in the normal FRP rod (4.33%) compared to (0.72%) in the tracking-resistant FRP rod indicates that electrical discharge activities were more severe in the normal FRP rod. In addition, a significant increment in sodium content in aged samples is due to the deposit of Na on the FRP rod surface from the contamination solution used in the experiment (made of water and NaCl). The content of silicon is found to be higher for both aged samples, which could be due to the breakage of

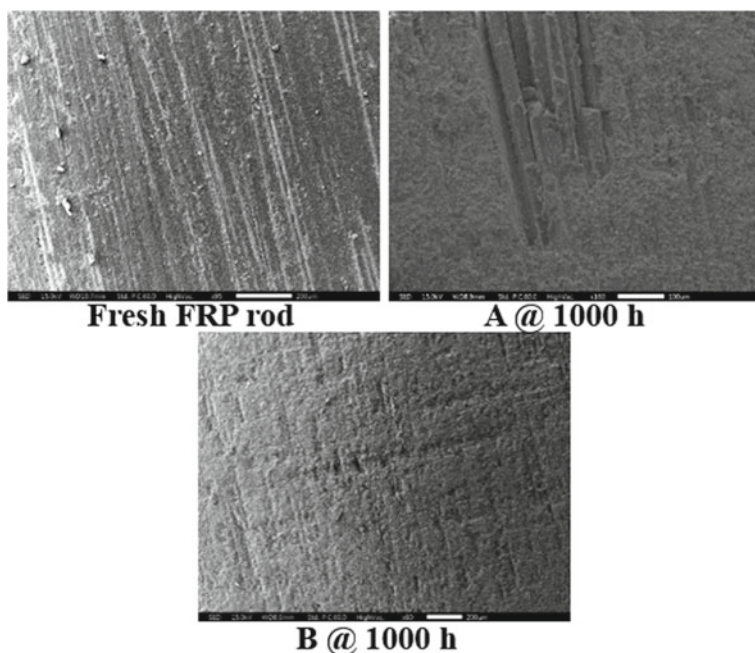


Fig. 7 SEM images of fresh and aged FRP rods (normal (A) and tracking-resistant (B))

Table 1 Atomic elemental composition (%) of fresh and aged FRP rods (normal (A) and tracking-resistant (B))

Sample	C	O	Na	Mg	Al	Si	N	Ca
A Fresh	64.01	29.37	0.72	0.67	0.87	3.56	–	0.80
A @ 1046 h	22.32	41.16	17.91	1.32	2.28	7.25	4.33	3.43
B Fresh	46.27	38.25	0.58	0.66	2.91	7.73	–	3.60
B @ 1046 h	28.15	37.79	17.43	1.51	2.34	8.81	0.72	3.25

the Si–O bond in glass fiber due to discharge activities. Also, the content of metallic elements such as Mg and Ca is increased due to ionic exchange with free H^+ when the FRP rod is subjected to water or nitric acid. For example, Ca^{2+} is replaced with H^+ , resulting in more Ca content.

4.2 Fourier Transform Infrared Spectroscopy (FTIR)

FTIR analysis is conducted on fresh and aged FRP rod samples (A and B) using PerkinElmer FTIR/FIR Spectrometer with the attenuated total reflection (ATR) accessory over the infrared frequency range between 400 and 4000 cm^{-1} . FTIR

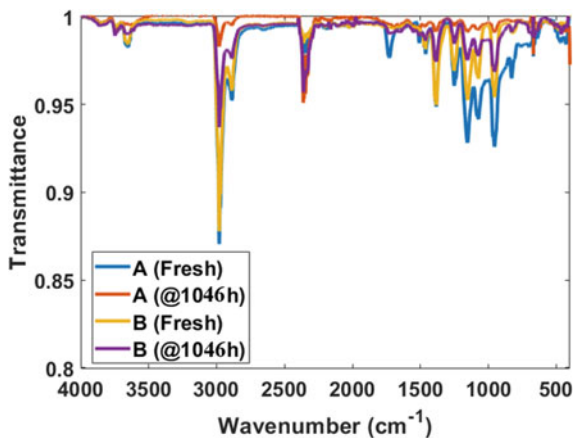
is used to identify the molecular bonding over the sample surface and determine the degradation level. Table 2 shows important functional groups of the FRP rod. Figure 8 shows the FTIR transmittance spectrum of the fresh and aged FRP rods (A and B).

By comparing the transmittance spectrum peaks of both types of FRP rods (A and B) used in this study, there is no significant difference. However, the absence of curing and coupling agent is noticed in the tracking-resistant FRP rod as the functional group at 1736 cm^{-1} is not seen in sample B. In addition, aromatic structure, methyl, and aliphatic groups in the epoxy resin matrix were found to be more in sample A as their corresponding transmittance peaks were higher. Significant degradation is observed in the normal FRP rod compared to the tracking-resistant FRP rod, where a significant decrement in the transmittance peaks corresponding to C-H in methyl (at 2920 and 2970 cm^{-1}), aromatic structure (at 831 , 1508 , and 1608 cm^{-1}), and aliphatic groups (at 1038 , 1182 , and 1451 cm^{-1}). This indicates that the epoxy resin matrix is severely degraded under the effect of electrical arcs and discharges. Consequently, it is decomposed into H_2O and CO_2 under the action of electrical discharges. However, the decrement in transmittance peaks corresponding to C-H

Table 2 Important functional groups of the FRP rod

Chemical bond	Wavenumber (cm^{-1})
Si-O	480
Aromatic structure	831, 1508, 1608
C-O-C in aliphatic	1038
C-C in aliphatic	1182
C-H in aliphatic	1451
C=O in ester	1736
C-H in methyl	2962
Hydroxyl OH	3200-3700

Fig. 8 FTIR transmittance spectrum of the fresh and aged FRP rods (normal (A) and tracking-resistant (B))



in methyl, aromatic structure, and aliphatic groups is less in the tracking-resistant FRP rod sample, indicating less degradation of the epoxy resin matrix. Moreover, an obvious reduction in the transmittance spectrum peak at 480 cm^{-1} corresponding to Si–O is seen to be more in sample A. This indicates more hydrolysis and ion exchange of glass fiber in sample A during the experiment. Where hydrogen ions replace the aluminum, sodium, and calcium ions in glass fiber, destroying the three-dimensional network structure of glass fiber [17]. The hydroxyl OH group between 3500 and 3700 cm^{-1} is found to be more in aged samples, which is due to hydrolysis and ion exchange of the Si–O–Si group in glass fiber.

5 Conclusion

This study investigates the performance of FRP rods used in high voltage composite insulators using the rotating wheel and dip experiment. The experiment was conducted for 1046 h on two different types of FRP rods (normal and tracking-resistant). Visual observation and leakage current analysis show that the new tracking-resistant FRP rod offers superior performance compared with the normal FRP rod, where tracking is observed near the ground terminal of the normal FRP rod after about 312 h with higher leakage current and discharge activities. To evaluate the degradation and aging degree, a detailed microscopic analysis is conducted on both FRP rods (fresh and aged samples), including scanning electron microscope (SEM), energy dispersive X-ray spectroscopy (EDS), and Fourier transform infrared spectroscopy (FTIR). SEM images and EDS analysis show a significant deterioration in the epoxy resin matrix in the normal FRP rod with many fractured glass fiber. Also, hydrolysis and ion exchange are more severe in the normal FRP rod. FTIR analysis shows a significant decrement in the normal FRP rod in various transmittance peaks corresponding to methyl, aromatic structure, and aliphatic groups. The tracking-resistant FRP rod showed superior performance over the normal FRP rod.

Acknowledgements The authors are grateful to M/s. Bana Engineering Pvt Ltd, Hyderabad, and M/s Ashrita Energy Intellects Pvt Ltd, Bangalore, for the support and help.

References

1. Halloum MR, Reddy BS, Reddy GN (2022) Stress control for polymeric outdoor insulators using nonlinear resistive field grading materials operating under different conditions. *IEEE Trans Dielectr Electr Insul* 29(3):1175–1182
2. Papailiou K, Schmuck F (2013) *Silicone composite insulators: materials, design, applications, power systems series*, 2nd edn. Springer, Berlin, Heidelberg, Germany
3. Meyer LH, Cherney EA, Jayaram SH (2004) The role of inorganic fillers in silicone rubber for outdoor insulation alumina tri-hydrate or silica. *Electr Insul Mag* 20(4):13–21

4. Liang X, Bao W, Gao Y (2018) Decay-like fracture mechanism of silicone rubber composite insulator. *IEEE Trans Dielectr Electr Insul* 25(1):110–119
5. Ramirez-Vazquez I, Hernandez-Corona R (2016) Tracking and erosion requirements for high voltage silicone rubber insulators. In: *IEEE international conference in dielectric*. Montpellier, France, pp 740–743
6. Wang J, Liang X, Gao Y (2014) Failure analysis of decay-like fracture of composite insulator. *IEEE Trans Dielectr Electr Insul* 21(6):2503–2511
7. Kuhl M (2001) FRP rods for brittle fracture-resistant composite insulators. *IEEE Trans Dielectr Electr Insul* 8(2):182–190
8. de Tourreil C et al (2005) Determination of the brittle fracture process of field failed HV insulators. In: *Proceedings of the XIVth international symposium on high voltage engineering*. Beijing, China
9. Burnhamet JT et al (2002) IEEE task force report: brittle fracture in nonceramic insulators. *IEEE Trans Power Deliv* 17(1):848–856
10. Yuan Z, Tu Y, Zhao Y, Jiang H, Wang C (2019) Degradation behavior and aging mechanism of decay-like fractured GRP rod in composite insulator. *IEEE Trans Dielectr Electr Insul* 26(3):1027–1034
11. Gao Y, Liang X, Bao W, Wu C, Li S (2019) Degradation characteristics of epoxy resin of GFRP rod in the decay-like fracture of composite insulator. *IEEE Trans Dielectr Electr Insul* 26(1):107–114
12. Nandi S, Reddy BS (2020) Understanding field failures of composite transmission insulators. *Eng Fail Anal* 116(1):104758
13. Negi M, Reddy BS (2021) Experimental study on the failures of FRP rods used in composite insulators. In: *Proceedings of symposium on power electronic and renewable energy systems control*. Singapore, pp 489–499
14. Nandi S, Reddy BS (2022) Chaotic analysis as a tool for sensing long-term degradation of high voltage fibre reinforced plastic core used in composite insulators. *IEEE Trans Power Deliv* 37(5):4103–4111
15. IEC Polymeric HV Insulators for Indoor and Outdoor Use-General Definitions (2012) Test methods and acceptance criteria. IEC 62217
16. IEC HV polymeric insulators for indoor and outdoor use tracking and erosion testing by wheel test and 5000h Test, IEC/TR 62730, 2012–03
17. Montesinos J, Gorur RS, Mobasher B, Kingsbury D (2002) Mechanism of brittle fracture in nonceramic insulators. *IEEE Trans Dielectr Electr Insul* 9(2):236–243

MFO Algorithm-Based Profit Maximization of Distribution Companies by Optimal Placement and Sizing of DGs Under Deregulated Environment



Ch. S. K. B. Pradeep Kumar, G. Balamurugan, Y. Butchi Raju,
and Pasupulati Baburao

Abstract Distributed generations (DGs) have persuaded researchers for more than ten years that they are the most cost-effective and environmentally friendly alternative that can be combined with centralized generations. The right location and size are necessary for distributed generation planning to achieve the best possible technical, economic, commercial, and regulatory outcomes. Most of these goals are incompatible, hence multi-objective solutions are needed. Only a few articles have examined this issue from the standpoint of the DG owners among the several studies that have been suggested to address the DG placement problem, and none of them have taken into account the crucial role that the gearbox expenses allotted to DGs play. As a result, the multi-objective optimization voltage stability problem in a deregulated environment is solved in this work. By placing and sizing the distributed generation (DG) units in the radial distributed system (RDS) optimally, the problem's main goal is to increase the profit of distribution companies (DISCOs). To get the best results, a straightforward and parameter light moth-flame optimization (MFO) approach is obtainable. Additionally, improvements to voltage stability and a decrease in reactive and actual power loss ensure been made. The cost of operating and maintaining a single DG unit, as well as the best location for installation, are calculated. The IEEE 33-bus test system now uses the recommended method. Simulation is used to evaluate how well the strategy performs, and the results are contrasted with those

Ch. S. K. B. Pradeep Kumar (✉)

Department of EEE, Ramachandra College of Engineering, Eluru, Andhra Pradesh 534007, India
e-mail: skbpradeepkumarch@gmail.com

G. Balamurugan

Department of EEE, Annamalai University, Chidhambaram, Tamil Nadu 500007, India

Y. Butchi Raju

Department of EEE, Sir C.R. Reddy College of Engineering, Eluru, Andhra Pradesh 534006, India

P. Baburao

Department of EEE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu 600062, India
e-mail: drbaburao@veltech.edu.in

of alternative approaches already in use. The comparison shows that the suggested strategy boosts DISCO's revenue in a radial distribution system.

Keywords Distribution company · Radial distributed system · Distributed generation · Moth–flame optimization · Differential evolution · Deregulated power system

1 Introduction

Distributed generation (DG) has been more popular in recent years for usage in power grids. There are a number of ways in which DG units may affect system operating conditions, including the voltage profile, voltage stability, dependability, and the safety of the power market [1, 2]. Electricity is always a need, and distribution companies (DISCOs) work tirelessly to meet the needs of its consumers while keeping costs low. The design, operation, and maintenance of the DISCOs are based on the principle of maximizing return on investment at the lowest possible cost. Improving the voltage profile and minimizing power loss are crucial responsibilities for DISCOs to achieve maximum profit [3, 4]. To enhance the functioning of the system, the DISCOs are developing a great number of inventive technological concepts and programmes [5].

To maximize the DG owner's return while minimizing the distribution company's expenditure, this framework analyses and places DG's units in the best possible location of the RDS [6, 7]. To maximize the best placement and appropriate value of DG units, a new parameter-free MFO approach is suggested [8–11]. The two searching operators of MFO, such as moths and flames, have the greatest capacity to find the ideal global solution with the least amount of computing time. Consequently, it effectively optimizes the size and ideal position of DG with conventional operational restrictions in order to boost the voltage profile and decrease DISCO network loss [12–15]. This paper considers the unpredictability of load demand, power production, electricity pricing, and dependability. The improved performance of the MFO algorithm is shown by testing the method on standard IEEE 33 node test system. The simulation result of voltage profile, power loss, DG location and size, and the cost–benefit analysis of DISCOs and DG owners are presented cutting-edge this research both numerically and graphically. Research based on comparisons was also carried out to prove the method's efficacy.

2 Problem Formulation

2.1 Objective Functions of DISCOs

This study's primary purpose is to maximize the return of DG owners and decrease the costs of DISCOs. The profit of DG owners is expressed numerically as,

$$\max(F) = \text{MPF} \quad (1)$$

Anywhere F remains the system's impartial purpose and MPF remains the maximum profit of DG owners.

2.1.1 Maximum Profit Function

The ten-year change amid the benefits understood then the costs expended aimed at DG deployment is used to compute the maximum savings cost (Eq. 2).

$$\text{MPF} = \sum \text{Benefit} - \sum \text{Expenses} \quad (2)$$

2.1.2 Net Present Value

Equation (3) demonstrates how to calculate the remaining contemporary worth issue (γ) using a 9% rise amount and a 12.5% attention amount to assess the worth of current costs over a planning period (N). The current cost value is calculated by multiplying this factor by each expense heading.

$$\text{NPV Factor, } \gamma = \sum_{t=1}^N \frac{1 + \text{IF}}{1 + \text{IR}} \quad (3)$$

NPV stands for remaining contemporary worth, IF for inflation, and IR for interest rate when $t = 1, 2, 3, \dots, N$.

2.1.3 Benefit Evaluation

The expected price of DG-generated power is US\$ 300 per kWh, and the estimated cost of reducing energy loss is US\$ 0.05 per kWh, for a total cost of US\$ 300 per kWh.

2.1.4 Cost of Energy Loss Reduction

To simulate actual power losses, the load movement explanation for the test system is carried out deprived of DG, and then the procedure is repeated. The discrepancy in losses, which is indicated in Eq. (4), demonstrates a reduced net loss.

$$\text{NLR} = P_{\text{loss}} - P_{\text{loss,DG}}, \quad (4)$$

where NLR = net loss reduction, P_{loss} = system power loss without DG, and $P_{\text{loss,DG}}$ = system power loss with DG. The DG loss reduction is transformed to a cost value using Eq. (5).

$$C_{(\text{NLR})} = \text{NLR} \times (\text{cost of energy saving/kWh}) \times 8760 \quad (5)$$

CNLR's net present cost value is derived using Eq. (6).

$$\text{NPV}(C_{\text{NLR}}) = C_{\text{NLR}} \sum_{t=1}^N \gamma^t \quad (6)$$

wherever $t = 1, 2, 3, \dots, N$; CNLR = cost of net loss reduction.

2.1.5 Cost of DG Power Generation

This inquiry has examined a PV system. The cost information aimed at this PV system is retrieved after [6] and computed using Eq. (7).

$$C_{(\text{DG,GEN})} = (\text{DG Size}) \times \left(\frac{\text{DG Generation cost}}{\text{kW}} \right) - \text{yr} \quad (7)$$

The present-day charge worth of CDG, Gen is computed by Eq. (8).

$$\text{NPV}(C_{\text{DG,Gen}}) = C_{\text{DG,Gen}} \sum_{t=1}^N \gamma^t \quad (8)$$

wherever $t = 1, 2, 3, \dots, N$; CDG, Gen = cost of DG power generation.

2.1.6 Expenses Cost

Amount comprises initial DG asset plus increasing O&M expenses throughout the forecasted time frame.

2.1.7 Operation and Maintenance Cost

This price takes into consideration the O&M expenses of grid-connected DG and is calculated as follows in Eq. (9). O&M data for DG implementation was gathered from

$$C_{(DG,O\&M)} = (DG \text{ Size}) \times (DGO\&M\text{cost}/kW - yr) \quad (9)$$

The present-day cost value of CDG, O&M is calculated using Eq. (10)

$$NPV (C_{DG,O\&M}) = C_{DG,O\&M} \sum_{t=1}^N \gamma^t \quad (10)$$

wherever $t = 1, 2, 3, \dots, N$; CDG, O&M = Cost of DG maintenance and operation.

2.1.8 Investment Cost

Investment cost of a solar PV-type DG is calculated at the best location using Eq. (11).

$$C_{(DG,Inv)} = (DG \text{ Size}) \times (DG \text{ Investment cost}/kW) \quad (11)$$

The economic justification of the aforementioned goal function is contingent upon the proper placement and rating of DG.

2.2 Limitations and Restrictions of DISCOs

The following restrictions apply to this optimization problem.

Bus voltages and branch currents limits during the planning process, locations and dimensions of DGs should be chosen so that branch currents and bus voltages remain within predetermined ranges. These restrictions are outlined below:

$$I_{b,t,j} \leq I_b^{\max} \quad (12)$$

$$V^{\min} \leq V_{n,t,j} \leq V^{\max} \quad (13)$$

where the minimum and maximum voltage levels for each bus are and indicate the maximum allowable current in a given line due to thermal constraints.

2.2.1 DG Capacity Limit

Assume the following period defines the limits on the active and reactive capacity of each DG.

$$P_{DG,i}^{\min} \leq P_{DG,i} \leq P_{DG,i}^{\max} \quad (14)$$

$$Q_{DG,i}^{\min} \leq Q_{DG,i} \leq Q_{DG,i}^{\max} \quad (15)$$

In these inequalities, $P_{DG,i}^{\min}$, $P_{DG,i}^{\max}$, $Q_{DG,i}^{\min}$, $Q_{DG,i}^{\max}$ are the lowest and highest amounts of active and reactive power that the DG unit is capable of producing, respectively.

2.2.2 Contract Price Limit

Constraints imposed by the electricity market on the contract price between the DG owner and the DISCO may be represented as follows:

$$C P_{DG}^{\min} \leq C P_{DG} \leq C P_{DG}^{\max}, \quad (16)$$

where $C P_{DG}^{\min}$ and $C P_{DG}^{\max}$ are the minimum and maximum contract price amounts that may be decided based on the market price of energy and other economic factors.

2.2.3 Power-Flow Constraints

Active and reactive power both need to be introduced into the system in order to satisfy the power-flow equations.

$$P_n = V_n \sum_{m \in N} V_m (g_{mn} \cos(\theta_{mn}) + b_{mn} \sin(\theta_{mn})) \quad (17)$$

$$Q_n = V_n \sum_{m \in N} V_m (g_{mn} \sin(\theta_{mn}) + b_{mn} \cos(\theta_{mn})) \quad (18)$$

where $C_{\text{investment}}^{\max}$ reflects the maximum amount of funding the DG owner can afford Fig. 1.

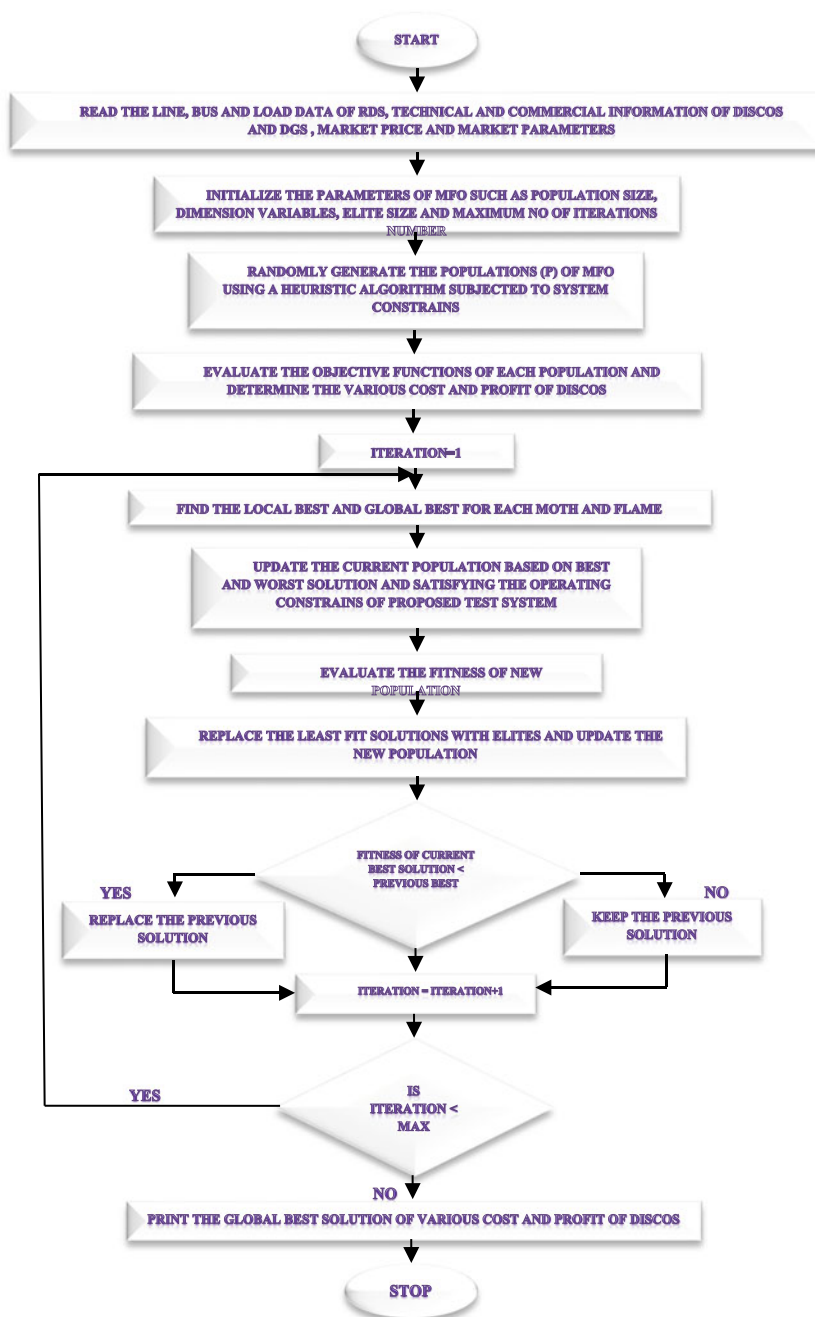


Fig. 1 Flowchart of DISCOs PROFIT maximization problem using proposed MFO algorithm

3 Result and Discussion

In order to determine the location of the DG, its size, and conduct economic analysis in a competitive open market environment, this section shows how to apply the MFO algorithm to the DG planning problem. IEEE 33 node test system [16] is used with the MFO algorithm to examine performance. The test case takes into account the DG system that is based on solar PV [17]. The investment cost and operation and maintenance costs for DGs in the 100–1000 kW range are US\$ 2493/kW and US\$ 19/kW, respectively while they are US\$ 2025/kW and US\$ 16/kW for the 1000–2000 kW range. Table 1 lists the market price and duration details for DISCOs. Additionally, it is anticipated that the contract price will range between \$35 and \$50 per MWh. As far as actual power goes, DG has a value between 0.2 and 1 MW. The technical and financial details of the DGs are displayed in Table 2. Table 3 displays the variables that were used in this study. An Intel Core i3 computer running at 2.10 GHz and equipped with 4 GB of RAM was used to carry out the optimization method. MATLAB version R2014a was used. Table 1 documents and summarizes the test case's findings for the IEEE 33 node RDS.

3.1 33 Node Test System

Table 4 and Fig. 2 may be found in the supplementary materials [18, 19], and they provide line data, bus data, and a single-line schematic of the 33-bus test system, respectively. It has a reactive and real load of 3755 kW and 2330 kVAR, correspondingly, then operates at a radial voltage of 1266 kV with 33 buses and 32 branches.

Table 1 Technical and commercial information of DISCOs

Level	Percentage of peak load	Network situation	Time duration (h/year)	Market Price (\$/MWh)
1	0–80	Light load	4380	35
2	80–95	Medium load	3285	45
3	95–100	Peak load	1095	50

Table 2 Commercial information of DGs

Parameter	Unit	Value
DG investment cost	\$/MW	318,000
DG operating cost	\$/MWh	29
DG maintenance cost	\$/MWh	7
Interest rate	%	12.5
Inflation rate	%	9

Table 3 Details of the technical parameters

Parameters	Values
Annual growth rate of loads	2%
λ_b (f/km year)	0.12
CF	1
Δt_{Fault} (h)	8
C_{int} (\$/kw)	20
N_{DG}	1
$Max(CP_{DG})$ (\$/MWh)	50
$Min(CP_{DG})$ (\$/MWh)	35
$Max(P_{DG})$ (MW)	1
$Min(P_{DG})$ (MW)	0.2
PF	0.9 lag
T_h	8760
T_d	365
INT_R (%)	12.5
INF_R (%)	9

By applying the MFO algorithm to the recommended RDS, the voltage contour and voltage stability directory are enhanced. The best sitting and sizing of DG is 29 as shown in Fig. 2 and 1.335924 MW each, respectively. The status of voltage profile and VSI for base case analysis with DG location is numerically reported in Tables 4 and 5. It is also graphically displayed in Figs. 3 and 4, respectively.

Table 6 shows the system’s minimum VSI, actual power loss, and reactive power loss with and without DG installation. The value of the indices is nearby zero (except VSI), and it is obvious that the performance of the system is enhanced. Since the optimal result of voltage stability index is nearby one, it indicates that the system voltage stability has been improved.

The simulated costs of various DISCO outlays over the course of a single year are shown in Table 7. The whole cost of DG ownership consists of the initial investment, as well as the running and maintenance expenses. The benefit, expenses and total profit of DISCOs for 10 year of planning period are clearly reported in Table 8 [20, 21]. The benefit, expenses and total profit of DISCOs are 21.59 M \$/year, 3.81 M \$/year and 17.78 M \$/year, respectively.

Table 7 then Table 8 compare the consequences of the proposed multi-objective MFO approach to those of the fuzzy-DE algorithm to show which is more effective. From Table 8, it is established that the MFO algorithm provides higher benefit, total profits with less computational time by proper sitting and sizing of DG in RDS. The total profits of DISCOs are increased by 13.72% when compared with fuzzy-DE algorithm. Figure 5 depicts a comparison between the benefit, expenses and total profit of DISCOs and the present system. Figure 6 shows the convergence curve for

Table 4 Base case voltage profile by installing DG in 33-bus node test system

Bus no.	Base case voltage	MFO voltage (proposed)
1	1.0000	1.0000
2	0.9970	0.9978
3	0.9829	0.9888
4	0.9754	0.9845
5	0.9680	0.9803
6	0.9495	0.9688
7	0.9460	0.9649
8	0.9323	0.9496
9	0.9260	0.9423
10	0.9201	0.9366
11	0.9192	0.9357
12	0.9177	0.9342
13	0.9115	0.9282
14	0.9092	0.926
15	0.9078	0.9246
16	0.9064	0.9233
17	0.9043	0.9213
18	0.9037	0.9207
19	0.9965	0.9973
20	0.9929	0.9936
21	0.9922	0.9929
22	0.9916	0.9922
23	0.9793	0.9855
24	0.9726	0.9793
25	0.9693	0.976
26	0.9475	0.9688
27	0.9450	0.9691
28	0.9335	0.9695
29	0.9253	0.9722
30	0.9217	0.9695
31	0.9176	0.9637
32	0.9167	0.9627
33	0.9164	0.9622

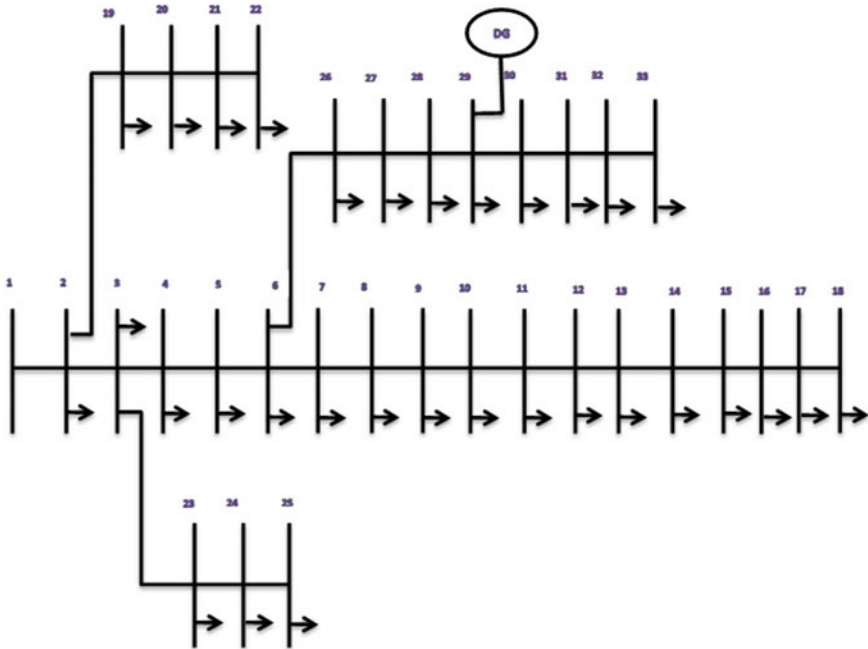


Fig. 2 Structure of 33 node test system with optimal location of DG

the 33-bus test system. The data shows that it can be extrapolated that just 25 to 30 iterations are required for the strategy to provide the optimal global answer.

The distribution load flow method may be used to freely calculate the base case voltage and VSI and to carry out the base case power flow. To enhance the voltage profile and increase DISCOs profits, the MFO algorithm is then utilized to find the sweet spot for the DG's placement and size. It has more ability and well suitable for solving the problems of large system and to obtain the accurate solution by using it's searching operators of moth and flame. The two operators are effectively tuning the optimal value of RDS for maximizing the profit of DISCOs. In this work, a single unit of DG is considered and rating of DG is 0 to 2 MW. In this algorithm, only limited number of parameters is used to obtain the solutions. The feasible control parameters obtained by the training process of MFO algorithm is as follows. Population size = 50; number of variables = 7; maximum number of iterations = 100.

Table 5 Base case values of VSI with the placement of DG in 33-bus test system

Bus no	Base case VSI	MFO VSI (proposed)
1	1.0000	1.0000
2	0.9978	0.9978
3	0.9888	0.9888
4	0.9845	0.9845
5	0.9803	0.9803
6	0.9688	0.9688
7	0.9649	0.9649
8	0.9496	0.9496
9	0.9423	0.9423
10	0.9366	0.9366
11	0.9357	0.9357
12	0.9342	0.9342
13	0.9282	0.9282
14	0.926	0.926
15	0.9246	0.9246
16	0.9233	0.9233
17	0.9213	0.9213
18	0.9207	0.9207
19	0.9973	0.9973
20	0.9936	0.9936
21	0.9929	0.9929
22	0.9922	0.9922
23	0.9855	0.9855
24	0.9793	0.9793
25	0.976	0.976
26	0.9688	0.9688
27	0.9691	0.9691
28	0.9695	0.9695
29	0.9722	0.9722
30	0.9695	0.9695
31	0.9637	0.9637
32	0.9627	0.9627
33	0.9622	0.9622

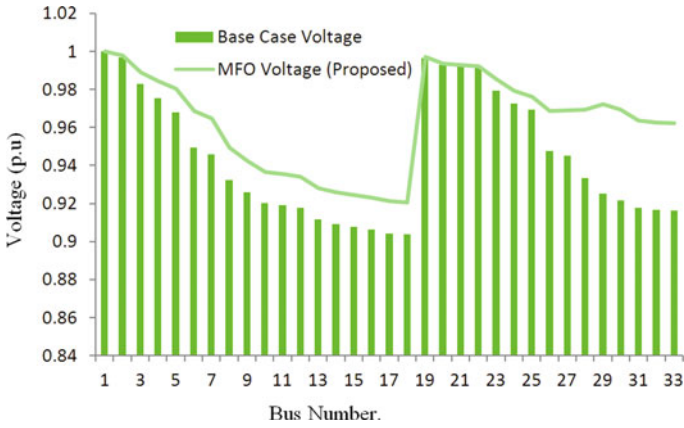


Fig. 3 Base case voltage profile by the placement of DG in 33-bus test system

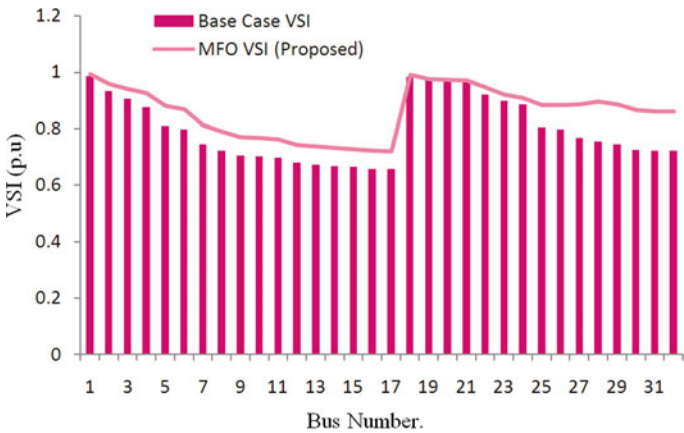


Fig. 4 Base case VSI and DG placement in 33-bus test system

Table 6 Comparison of real and reactive power loss for 33 node system

Parameters	Without DG	With DG
Real power loss (kW)	210.679	129.31
Reactive power loss (kVar)	130.121	86.3121
Minimum VSI bus no	18	27
Minimum VSI (p.u.)	0.718439	0.881879
Optimal location of the DG	29	
Optimal size of the DG	1.35924 MW	

Table 7 Simulation results of 33-bus system after 1 year

Various costs of DISCOs	Fuzzy with DE algorithm	MFO (proposed algorithm)
Cost of energy loss reduction (K\$/year)	39.89	36.43
Cost of DG power generation (K\$/year)	341.16	407.77
Operation and maintenance cost of DG (K\$/year)	18.19	21.748
Investment cost of DG (K\$/year)	2302.86	2752.46

Table 8 Simulation results of 33-bus system after 10 years of planning period

Various costs and profits of DISCOs	Fuzzy with DE algorithm	MFO (proposed algorithm)
Cost of energy loss reduction (M\$/year)	1.94	1.77
Cost of DG power generation (M\$/year)	16.58	19.82
Operation and maintenance cost of DG (M\$/year)	0.88	1.06
Investment cost of DG (M\$/year)	2.30	2.75
Benefit of DISCOs (M\$/year)	18.53	21.59
Expenses of DISCOs (M\$/year)	3.19	3.81
Total profit of DISCOs (M\$/year)	15.34	17.78



Fig. 5 Comparison of cost, benefit and profit of DISCOs for 33-bus test system

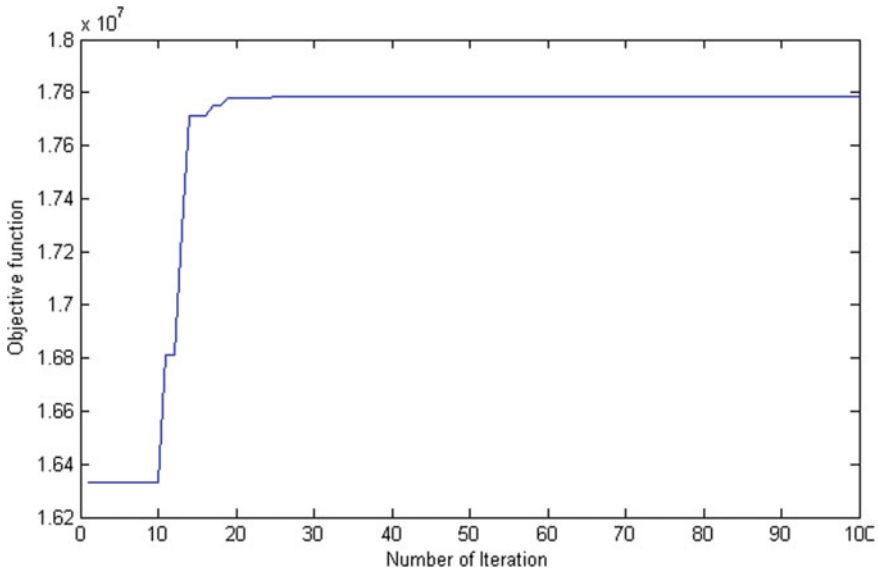


Fig. 6 Convergence curve of 33-bus test system

4 Conclusion

The various voltage stability issues in both regulated and deregulated power systems are solved using the intelligent computational optimization algorithm moth-flame optimization (MFO) in this research work. The MFO meta-heuristic algorithm has been successfully used to increase the voltage stability of radial distribution systems, and it is regarded as one of the most promising of the meta-heuristic algorithms. Moths and flames are its two primary constituents. A solution has also been considered using both moths and flames. Through the proper placement and sizing of various capacitors, DG units, and network reconfiguration processes, the two searching operators of MFO have successfully improved the voltage profile, maximized net savings, and decreased network losses of RDS.

To enhance the advantages of DISCOs and DG owners in a deregulated environment, a devised MFO algorithm has also been used. The costs and income of the different DG owners and DISCOs are calculated here. It includes the costs of purchasing capacitors and DGs, as well as their costs of operation and maintenance. It also includes the costs of substations and customer interruptions, as well as the revenue and profits of DG owners. To demonstrate the effectiveness of the MFO, a numerical example using IEEE standard RDS test systems has been taken into consideration. The outcome demonstrates that, in comparison with other competing algorithms, the suggested algorithm offers an increase in profit with less computational time. Therefore, it can be said that the suggested MFO approach paves the best path for resolving power system optimization issues in a deregulated environment.

References

1. Ackermann T, Andersson G, Söder L (2001) Distributed generation: a definition. *Electr Power Syst Res* 57(3):195–204. [https://doi.org/10.1016/s0378-7796\(01\)00101-8](https://doi.org/10.1016/s0378-7796(01)00101-8)
2. Acharya N, Mahat P, Mithulananthan N (2006) An analytical approach for DG allocation in primary distribution network. *Int J Electr Power Energy Syst* 28(10):669–678. <https://doi.org/10.1016/j.ijepes.2006.02.013>
3. Wang C, Nehrir M (2004) Analytical approaches for optimal placement of distributed generation sources in power systems. *IEEE Trans Power Syst* 19(4):2068–2076. <https://doi.org/10.1109/TPWRS.2004.836189>
4. Abapour S, Nojavan S, Abapour M (2018) Multi-objective short-term scheduling of active distribution networks for benefit maximization of DisCos and DG owners considering demand response programs and energy storage system. *J Mod Power Syst Clean Energy* 6(1):95–106
5. Kim KH, Lee YJ, Rhee SB, Lee SK, You SK (2002) Dispersed generator placement using fuzzy-ga in distribution systems. In: 2002 IEEE power engineering society summer meeting, vol 3, pp 1148–1153. <https://doi.org/10.1109/pess.2002.1043458>
6. Singh K, Yadav VK, Padhy NP, Sharma J (2014) Congestion management considering optimal placement of distributed generator in deregulated power system networks. *Electr Power Compon Syst* 1(42):13–22. <https://doi.org/10.1080/15325008.2013.843218>
7. Shukla TN, Singh SP, Srinivasarao V, Naik KB (2010) Optimal sizing of distributed generation placed on radial distribution systems. *Electr Power Compon Syst* 38(3):260–274. <https://doi.org/10.1080/15325000903273403>
8. Kansal S, Tyagi B, Kumar V (2017) Cost benefit analysis for optimal distributed generation placement in distribution systems. *Int J Ambient Energy*. <https://doi.org/10.1080/01430750.2015.1031407>
9. NREL Distributed generation renewable energy estimate of costs. <http://www.nrel.gov=analysis=techlcoercoctest.html>
10. Devi AL, Subramanyam B (2007) Optimal DG unit placement for loss reduction in radial distribution system—a case study. *ARPN J Eng Appl Sci* 2(6):57–61
11. Ghatak SR, Sannigrasi S, Acharjee P (2018) Comparative performance analysis of DG and DSTATCOM using improved PSO based on success rate for deregulated environment. *IEEE Syst J* 12(3):2791–2802
12. Ameli A, Farrokhifard M, Ahmadifar A, Haghifam MR (2015) Distributed generation planning based on the distribution company's and the DG owner's profit maximization. *Int Trans Electr Energy Syst* 25(2):216–232
13. Saboori H, Hemmati R (2017) Maximizing DISCO profit in active distribution networks by optimal planning of energy storage systems and distributed generators. *Renew Sustain Energy Rev* 71:365–372
14. Ameli A, Bahrami S, Khazaeli F, Haghifam MR (2014) A multiobjective particle swarm optimization for sizing and placement of DGs from DG owner's and distribution company's viewpoints. *IEEE Trans Power Deliv* 29(4):1831–1840
15. Ameli A, Farrokhifard M, Ahmadifar A, Haghifam M-R (2013) Distributed generation planning based on the distribution company's and the DG owner's profit maximization. *Int Trans Electr Energy Syst*
16. Ghaemi S, Zare K (2017) A new method of distribution marginal price calculation in distribution networks by considering the effect of distributed generations location on network loss. *J Oper Autom Power Eng* 5(2):171–180
17. Salehi J, Jannati Oskuee MR (2017) Optimal planning of distributed generation from the DisCo and the DGO viewpoints considering the uncertainties in future demand and electricity price. *Int J Ambient Energy* 39(8)
18. Kansal S, Tyagi B, Kumar V (2017) Cost–benefit analysis for optimal distributed generation placement in distribution systems. *Int J Ambient Energy* 38(1):45–54

19. Kianmehr E, Nikkhah S, Rabiee A (2019) Multi-objective stochastic model for joint optimal allocation of DG units and network reconfiguration from DG owner's and DISCO's perspectives. *Renew Energy* 132:471–485
20. Ouali S, Cherkaoui A (2020) Optimal allocation of combined renewable distributed generation and capacitor units for interconnection cost reduction. *J Electr Comput Eng* 2020:1–11
21. Roy K, Dixit S (2020) Optimal placement and sizing of distributed generation and capacitors: a review. *Int J New Innov Eng Technol* 2(4):101–106

Peer-To-Peer Transaction Model Among Prosumers Considering Franchise Rights of Distribution Companies



Zhemin Lin, Bochun Zhan, Zihao Li, Changsen Feng, and Fushuan Wen 

Abstract In developing a peer-to-peer (P2P) electricity trading mechanism for the distribution system level, it is necessary to consider the franchise rights owned by the concerned distribution company for designing an appropriate network charging model, so as to effectively compensate the distribution system investment and operating costs. Given this background, this paper proposes a prosumer P2P transaction model that takes into account the distribution company franchise and distribution network operation constraints. First, based on the leader–follower interaction relationship between the distribution company and prosumers, a two-layer game model of determining network tariffs based on electrical distance is established. Then, based on the obtained network tariffs, the Alternating Direction Method of Multipliers (ADMM) is used to determine the P2P real-time transaction power and price of the prosumers with the data privacy protected. Finally, the IEEE 33-node distribution system is employed to demonstrate the proposed method, and it is found by simulation results that interests of all prosumers can be fairly and reasonably protected, on the basis of ensuring the secure operation of the distribution network, and the distribution company can be reasonably compensated for the resulting loss of revenue due to giving up some franchise rights.

Z. Lin

Anhui Power Exchange Center Company Ltd, Hefei, China

B. Zhan · Z. Li · F. Wen (✉)

College of Electrical Engineering, Zhejiang University, Hangzhou, China

e-mail: fushuan.wen@gmail.com

B. Zhan

e-mail: zhanbc@zju.edu.cn

Z. Li

e-mail: zihao_li_ee@zju.edu.cn

C. Feng

College of Information Engineering, Zhejiang University of Technology, Hangzhou, China

e-mail: fcs@zjut.edu.cn

Keywords Prosumer · Peer-to-peer (P2P) transaction · Distribution system · Franchise · Leader–follower game · Network tariff · KKT condition · Alternating direction method of multipliers (ADMM)

1 Introduction

Under the premise of following market rules and ensuring the secure and stable operation of the power system, it is necessary to design a secure, transparent and fair market operation mode and transaction mechanism according to the characteristics of prosumers P2P transactions.

In order to coordinate the emerging prosumers in the electricity market, the academia circle has conducted extensive research on P2P transactions in recent years. References [1–3] proposed a multi-prosumer P2P trading market structure. References [4, 5] considered the output uncertainty of distributed renewable energy, and established a robust optimization model for P2P transactions of prosumers. The main advantage of the above P2P transaction model is that it can maximize social benefits, but centralized optimization methods cannot protect the privacy of prosumers. Reference [6] established a multi-virtual power plants electricity-carbon-backup P2P transaction model, and distributed clearing through the adaptive step size ADMM algorithm. Reference [7] established a microgrid power trading mechanism based on blockchain. The above studies realize the clearing of P2P transactions through distributed algorithms and blockchain platforms to protect the privacy of prosumers, but do not consider network operation constraints in P2P transactions. Reference [8] proposed a P2P transaction mechanism considering the operation constraints of the distribution system and the clearing method, but did not consider the network tariffs in the P2P transaction.

The existing models of determining network tariffs mainly include: pricing models based on stamp method and contract path method [9, 10], based on pricing model of network tariffs for electrical distance. Reference [11] proposed a DLMP model based on the optimal power flow model. Reference [12] established a DLMP model based on second-order cone relaxation, but its computational complexity is large and there is a certain degree of computing error. References [13–15] established a pricing model based on electrical distance, and conducted in-depth analysis on issues of grid connection tariffs between generators and loads. The above research does not consider the full recovery of the distribution company's operating costs and network loss costs, and it is difficult to achieve a reasonable distribution of benefits among multiple subjects.

In the above background, this paper will introduce the network tariffs pricing model based on electrical distance into the optimization decision-making of distribution companies. At the same time, considering the difference in time scale between the power distribution company's network tariffs price decision and the prosumer's P2P transaction decision, ADMM is used to realize the distributed clearing of P2P transactions between producers and consumers at a given network tariffs price. Finally,

an example is used to verify the rationality and effectiveness of the network tariffs pricing mechanism designed in this paper.

2 Pricing Model of Network Tariffs Considering Franchise Loss

2.1 Franchise of Distribution Company

The franchise loss is caused by the fact that electricity transactions between generators and consumers do not go through distribution companies. The franchise loss function of distribution companies can be expressed as:

$$f_{\text{loss}} = R_s - R_d \quad (1)$$

$$R_s = \sum_{i \in A} \sum_{t=1}^T [(\varepsilon_c^t - \varepsilon_{bd}^t) P_{i,b}^t - C_{\text{loss}}^t] \quad (2)$$

$$R_d = \sum_{i \in A} \sum_{t=1}^T \left[P_{i,m}^t (\varepsilon_c^t - \varepsilon_{bd}^t) + \sum_{j \in A} |P_{ij}^t| \delta_{ij,\text{net}} - C_{\text{loss}}^t \right] \quad (3)$$

where A represents the set of prosumers; T represents the optimization cycle of the network tariffs price; f_{loss} represents the franchise loss of the electricity distribution company; R_s represents electricity revenue when the franchise of the electricity distribution company is not damaged; R_d represents the income of the power distribution company during the energy transaction; ε_c^t represents the electricity price of the distribution company in time period t ; ε_{bd}^t is the electricity purchase price of the distribution company from the generation company in time period t ; $P_{i,b}^t$ represents power purchased from the distribution company in time period t when prosumer i does not participate in P2P transactions; $P_{i,m}^t$ represents power purchased from the distribution company in time period t when prosumer i participates in P2P transactions; P_{ij}^t represents the transaction power between prosumer i and j ; $\delta_{ij,\text{net}}$ represents the network connection fee price of the transaction between prosumer i and j ; C_{loss}^t is the network loss cost of time period t .

2.2 Network Tariffs Pricing Model

Assume that there are two distributed prosumers i and j for P2P transactions, and the corresponding network tariffs can be expressed as:

$$\delta_{ij,\text{net}} = \mathcal{V}_{\text{net}} C_{ij,\text{MWkm}} \quad (4)$$

where \mathcal{V}_{net} represents the network tariffs decided by the power distribution company to compensate for the loss of its franchise rights, and the unit is yuan $\text{MW}^{-1} \text{km}^{-1}$; $C_{ij,\text{MWkm}}$ represents the megawatt kilometer of P2P energy transaction for node i and node j , the unit is km.

The marginal MW-kilometer of the electricity trade between node i and node j is related to the basic MW-kilometer of node i and node j [11], which can be calculated by Eqs. (5) and (6), respectively.

$$C_{i,\text{MWkm}} = \sum_{(\alpha,\beta) \in L} f_{i,\alpha\beta} l_{\alpha\beta} \quad (5)$$

$$C_{j,\text{MWkm}} = \sum_{(\alpha,\beta) \in L} f_{j,\alpha\beta} l_{\alpha\beta} \quad (6)$$

where $C_{i,\text{MWkm}}$ and $C_{j,\text{MWkm}}$ are the basic megawatt kilometers of node i and node j , respectively, and the unit is km. $f_{i,\alpha\beta}$ and $f_{j,\alpha\beta}$ are the power flow transfer factors, which are the injection of increased unit power generation at node i and node j and the increase at any selected reference node. The power flow of the line at unit load power is calculated using the AC power flow model; $l_{\alpha\beta}$ is the reference length of the line (α, β) , which is obtained by converting the line voltage level and line type into the reference line according to the cost ratio; L is the set of regional lines.

In order to make the basic MW-km of each node irrelevant to the position of the selected reference node, and to make P2P trading parties i and j each share half of the network tariffs, this paper uses the correction coefficient C_{ij} to correct the basic MW-km of each node. On the premise that the basic MW-kilometer $C_{i,\text{MWkm}}$ and $C_{j,\text{MWkm}}$ of nodes i and j is known, the correction coefficient can be calculated by Eq. (7), and then the marginal MW-km $C_{ij,\text{MWkm}}$ of P2P power trading between node i and j can be obtained, which reflects the degree to which distribution lines are used by P2P transactions between prosumers.

$$C_{ij,\text{MWkm}} = \frac{C_{i,\text{MWkm}} \pm C_{ij}}{2} = \frac{C_{j,\text{MWkm}} \mp C_{ij}}{2} \quad (7)$$

3 Two-Layer Model Framework Between Distribution Companies and Distributed Prosumers

3.1 Distribution Company Optimization Decision Model

The source of revenue of the power distribution company in the P2P transaction of the prosumers includes the collection of network tariffs and electricity sales fees, and the objective is to maximize the total revenue in the network tariffs price optimization cycle:

$$\max R_d = \sum_{i \in A} \sum_{t=1}^T \left[P_{i,c}^t (\varepsilon_c^t - \varepsilon_{bd}^t) + \sum_{j \in A} |P_{ij}^t| \delta_{ij,\text{net}} \right] - \varphi_{\text{loss}} \sum_{(\alpha,\beta) \in L} \sum_{t=1}^T I_{\alpha\beta}^{t2} r_{\alpha\beta} \quad (8)$$

$$\delta_{\text{net}}^{\min} \leq \delta_{ij,\text{net}} \leq \delta_{\text{net}}^{\max} \quad (9)$$

$$R_d \leq R_s = \sum_{i \in A} \sum_{t=1}^T (\varepsilon_c^t - \varepsilon_{bd}^t) P_{i,b}^t - \varphi_{\text{loss}} \sum_{(\alpha,\beta) \in L} \sum_{t=1}^T I_{\alpha\beta}^{t2} r_{\alpha\beta} \quad (10)$$

$$\text{s.t. (4) - (7)} \quad (11)$$

where φ_{loss} is the network loss cost coefficient; $I_{\alpha\beta}^t$ is the current of the line (α, β) in time period t and r_{ab} is the resistance of the line (α, β) ; $\delta_{\text{net}}^{\max}$ and $\delta_{\text{net}}^{\min}$ are the upper and lower limits of the network tariffs price set by the regulatory department, respectively.

Equation (10) ensures that the distribution company's income when the franchise is damaged does not exceed the income when its franchise is not damaged, so as to protect the interests of distributed prosumers and prevent the distribution company from obtaining excess compensation.

In addition, the power distribution company should be responsible for checking the security constraints of P2P transactions to ensure that the results of P2P transactions meet the network constraints:

$$P_{\alpha\beta,F}^t = P_{\beta,z}^t + \sum_{\gamma \neq \alpha: (\beta,\gamma) \in L} P_{\beta\gamma,F}^t + r_{\alpha\beta} I_{\alpha\beta}^{t2} \quad (12)$$

$$Q_{\alpha\beta,F}^t = Q_{\beta,z}^t + \sum_{\gamma \neq \alpha: (\beta,\gamma) \in L} Q_{\beta\gamma,F}^t + \chi_{\alpha\beta} I_{\alpha\beta}^{t2} \quad (13)$$

$$V_{\beta}^{t2} = V_{\alpha}^{t2} - 2(r_{\alpha\beta} P_{\alpha\beta,F}^t + \chi_{\alpha\beta} Q_{\alpha\beta,F}^t) + (r_{\alpha\beta}^2 + \chi_{\alpha\beta}^2) I_{\alpha\beta}^{t2} \quad (14)$$

$$I_{\alpha\beta}^{t2} V_{\alpha}^{t2} = P_{\alpha\beta,F}^{t2} + Q_{\alpha\beta,F}^{t2} \quad (15)$$

$$V_{\alpha,\text{down}}^2 \leq V_{\alpha}^{t2} \leq V_{\alpha,\text{up}}^2 \quad (16)$$

$$0 \leq I_{\alpha\beta}^{t2} \leq I_{\alpha\beta,\text{up}}^2 \quad (17)$$

where $P_{\alpha\beta,F}^t$ and $Q_{\alpha\beta,F}^t$ are the active power flow and reactive power flow of the line (α, β) during period t , respectively; $P_{j,z}^t$ and $Q_{j,z}^t$ are the active power and reactive power injected into the node β during period t , respectively; $\chi_{\alpha\beta}$ is the reactance of the line (α, β) ; $I_{\alpha\beta,\text{up}}$ is the current upper limit of the line (α, β) in period t ; V_{α}^t and V_{β}^t are the voltage modulus value of node α and node β in period t , respectively; $V_{\alpha,\text{up}}$ and $V_{\alpha,\text{down}}$ are the voltage upper limit and lower limit of node α , respectively.

3.2 Optimal Decision-Making Model for Prosumers Considering Network Tariffs

The objective is to minimize total energy cost of distributed prosumers:

$$\min C_{\text{total}} = \sum_{i \in A} \sum_{t=1}^T \left(C_{i,\text{dg}}^t + C_{i,\text{bess}}^t + C_{i,\text{re}}^t + C_{i,\text{net}}^t + C_{i,\text{com}}^t \right) \quad (18)$$

where C_{total} represents the total cost of distributed prosumers.

(1) Distributed Generator Cost:

$$C_{i,\text{dg}}^t = C_{i,\text{ope}}^t + C_{i,\text{eqi}}^t = \varphi_{i,\text{dg}}^{\text{ope}} P_{i,\text{dg}}^t + \varphi_{i,\text{dg}}^{\text{eqi}} P_{i,\text{dg}}^t = \varphi_{i,\text{dg}} P_{i,\text{dg}}^t \quad (19)$$

where the distributed generator cost of prosumer i includes power generation cost $C_{i,\text{ope}}^t$ and equipment maintenance cost $C_{i,\text{eqi}}^t$; $P_{i,\text{dg}}^t$ is the active power of distributed generators in time period t . $\varphi_{i,\text{dg}}^{\text{ope}}$, $\varphi_{i,\text{dg}}^{\text{eqi}}$, and $\varphi_{i,\text{dg}}$ are respectively power generation cost coefficient, aging loss coefficient, and total cost coefficient.

(2) Battery Energy Storage Device Cost:

$$C_{i,\text{bess}}^t = \varphi_{i,\text{bess}} (P_{i,\text{c}}^t + P_{i,\text{dc}}^t) \quad (20)$$

where $P_{i,\text{c}}^t$ and $P_{i,\text{dc}}^t$ are respectively the charging power and discharging power of the prosumer i battery energy storage device; $\varphi_{i,\text{bess}}$ is the cost coefficient of the operation.

(3) Renewable Energy Generation Costs:

$$C_{i, \text{re}}^t = \varphi_{i, \text{w}} P_{i, \text{w}}^t + \varphi_{i, \text{s}} P_{i, \text{s}}^t \quad (21)$$

where $C_{i, \text{re}}^t$ is the power generation cost of renewable energy unit equipment for prosumer i in period t ; $P_{i, \text{w}}^t$ and $P_{i, \text{s}}^t$ are the output of wind turbine and photovoltaic equipment in period t ; $\varphi_{i, \text{w}}$ and $\varphi_{i, \text{s}}$ are the kWh cost of wind power and photovoltaic equipment, respectively.

- (4) Network Tariffs Cost for P2P Transactions:

$$C_{i, \text{net}}^t = \sum_{j \in A} |P_{ij}^t| \delta_{ij, \text{net}} \quad (22)$$

where $C_{i, \text{net}}^t$ represents network tariffs cost of the P2P transaction of the prosumer i in the time period t .

- (5) Power Purchase Fee from Power Distribution Company:

$$C_{i, \text{com}}^t = P_{i, \text{m}}^t \varepsilon_c^t \quad (23)$$

where $C_{i, \text{com}}^t$ represents the electricity purchase fee of prosumer i from distribution company.

Constraints include:

- (1) Power Balance Constraints:

$$\sum_{j \in A} P_{ij}^t + P_{i, \text{m}}^t = P_{i, \text{dg}}^t + P_{i, \text{dc}}^t - P_{i, \text{c}}^t + P_{i, \text{w}}^t + P_{i, \text{s}}^t - P_{i, \text{L}}^t : \lambda_{i, e}^t \in R \quad (24)$$

$$P_{ij}^t = -P_{ji}^t : \lambda_{ij}^t \in R \quad (25)$$

where $P_{i, \text{L}}^t$ represents the load of prosumer i in period t ; $\lambda_{i, e}^t$ and λ_{ij}^t are the multiplier variables corresponding to the constraints.

- (2) Distributed Generator Output Constraints:

$$P_{i, \text{dg}}^{\min} \leq P_{i, \text{dg}}^t \leq P_{i, \text{dg}}^{\max} : \bar{\mu}_{i, \text{dg}}^t, \underline{\mu}_{i, \text{dg}}^t \geq 0 \quad (26)$$

where $P_{i, \text{dg}}^{\max}$ and $P_{i, \text{dg}}^{\min}$ represent the upper and lower limits of distributed generator output respectively; $\bar{\mu}_{i, \text{dg}}^t$ and $\underline{\mu}_{i, \text{dg}}^t$ represent the multiplier variables corresponding to the constraints.

- (3) Power Constraints of Battery Energy Storage Devices:

$$S_{i, \text{bess}}^t = S_{i, \text{bess}}^{t-1} + \left(\eta_{i, \text{c}} P_{i, \text{c}}^t - \frac{1}{\eta_{i, \text{dc}}} P_{i, \text{dc}}^t \right) \frac{\Delta t}{Q_{i, \text{bess}}} : \lambda_{i, \text{bess}}^t \in R \quad (27)$$

$$S_{i, \text{bess}}^{\min} \leq S_{i, \text{bess}}^t \leq S_{i, \text{bess}}^{\max} : \bar{\mu}_{i, \text{bess}}^t, \underline{\mu}_{i, \text{bess}}^t \geq 0 \quad (28)$$

$$0 \leq P_{i,c}^t \leq P_{i,c}^{\max} : \bar{\mu}_{i,c}^t, \underline{\mu}_{i,c}^t \geq 0 \quad (29)$$

$$0 \leq P_{i,dc}^t \leq P_{i,dc}^{\max} : \bar{\mu}_{i,dc}^t, \underline{\mu}_{i,dc}^t \geq 0 \quad (30)$$

where $S_{i,\text{bess}}^t$ is the state of charge of the battery energy storage device for time period t . $S_{i,\text{bess}}^{\min}$ and $S_{i,\text{bess}}^{\max}$ are the lower limit and upper limit of the state of charge of the battery energy storage device; $\eta_{i,dc}$ and $\eta_{i,c}$ are the discharge and charge efficiency of the battery energy storage device, respectively; $Q_{i,\text{bess}}$ is the rated capacity of the battery energy storage device; $P_{i,dc}^{\max}$ and $P_{i,c}^{\max}$ are respectively the maximum discharge power and maximum charge power of the battery energy storage device; $\lambda_{i,\text{bess}}^t$, $\bar{\mu}_{i,\text{bess}}^t$, $\underline{\mu}_{i,\text{bess}}^t$, $\bar{\mu}_{i,c}^t$, $\underline{\mu}_{i,c}^t$, $\bar{\mu}_{i,dc}^t$, and $\underline{\mu}_{i,dc}^t$ are the multiplier variables corresponding to the constraints.

(4) Renewable Energy Generating Unit Output Constraints:

$$(1 - h_w)p_{i,w}^{\max} \leq p_{i,w}^t \leq p_{i,w}^{\max} : \bar{\mu}_{i,w}^t, \underline{\mu}_{i,w}^t \geq 0 \quad (31)$$

$$(1 - h_s)p_{i,s}^{\max} \leq p_{i,s}^t \leq p_{i,s}^{\max} : \bar{\mu}_{i,s}^t, \underline{\mu}_{i,s}^t \geq 0 \quad (32)$$

where $p_{i,w}^{\max}$ and $p_{i,s}^{\max}$ are the output upper limits of wind turbines and photovoltaic equipment, respectively. h_w and h_s are the maximum curtailment rates of wind turbines and photovoltaic equipment that promote renewable energy consumption; $\bar{\mu}_{i,w}^t$, $\underline{\mu}_{i,w}^t$, $\bar{\mu}_{i,s}^t$, and $\underline{\mu}_{i,s}^t$ are multiplier variables corresponding to constraints.

(5) Transaction Power Constraints:

$$-p_{ij}^{\max} \leq p_{ij}^t \leq p_{ij}^{\max} : \bar{\mu}_{ij}^t, \underline{\mu}_{ij}^t \geq 0 \quad (33)$$

$$0 \leq p_{i,m}^t \leq p_{i,m}^{\max} : \bar{\mu}_{i,m}^t, \underline{\mu}_{i,m}^t \geq 0 \quad (34)$$

where p_{ij}^{\max} and $p_{i,m}^{\max}$ are respectively the upper limit of the P2P transaction power of the prosumer i and j and the upper limit of the power purchased by the prosumer i from the power distribution company; $\bar{\mu}_{ij}^t$, $\underline{\mu}_{ij}^t$, $\bar{\mu}_{i,m}^t$, and $\underline{\mu}_{i,m}^t$ are the multiplier variables corresponding to the constraint conditions.

3.3 Game Interaction Mechanism Between Power Distribution Companies and Distributed Prosumers

As the leader of the leader–follower game, the power distribution company determines the price coefficient of the network tariffs as the game space according to the P2P transaction needs of the distributed prosumers and the power purchase

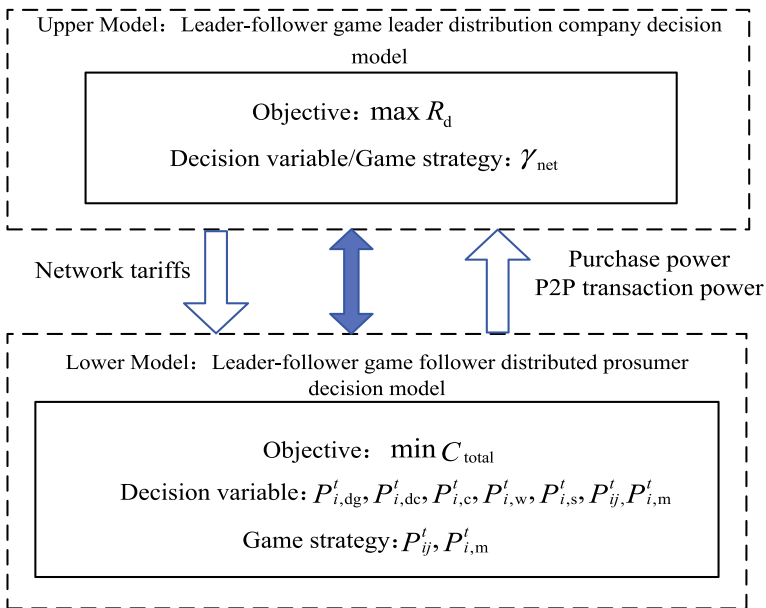


Fig. 1 Schematic diagram of leader–follower interactions in the gaming framework

demand from the power distribution company. The game strategy set of distribution company $v_d = \{\gamma_{net}\}$. Distributed prosumers, as followers of the leader–follower game, have accepted the price of network tariffs. The direction of the game is to meet the load demand with the minimum total cost or to maximize the total income on the basis of meeting the load demand. The game strategy set of distributed prosumers $v_i = \{P_{ij}^t, P_{i,m}^t\}$ (Fig. 1).

4 Prosumer P2P Energy Transaction Distributed Clearing

Based on ADMM, the prosumer P2P transaction model including the network tariffs is solved to obtain the actual P2P transaction volume and corresponding transaction price in the transaction cycle. Global auxiliary variables $\bar{P}_{i,z}^t$ can be introduced to reconstruct the original problem as:

$$\min C_{total} = \sum_{i \in A} \sum_{t=1}^{T_{opt}} C_i \left(\mathbf{P}_{i,opt}^t, \mathbf{P}_{i,P2P}^t, \mathbf{P}_{i,z}^t, \mathbf{X}_i \right) \quad (35)$$

$$h_i \left(\mathbf{P}_{i,opt}^t, \mathbf{P}_{i,P2P}^t, \mathbf{P}_{i,z}^t, \mathbf{X}_i \right) \leq 0 \quad (36)$$

$$\mathbf{P}_{i,\text{opt}}^t = \left[P_{i,\text{dg}}^t, P_{i,\text{dc}}^t, P_{i,\text{c}}^t, P_{i,\text{w}}^t, P_{i,\text{s}}^t, P_{i,\text{m}}^t \right]^T \quad (37)$$

$$\mathbf{P}_{i,\text{P2P}}^t = \left[P_{ij}^t, \forall j \in A \right] \quad (38)$$

$$P_{i,z}^t = P_{i,\text{dg}}^t + P_{i,\text{dc}}^t - P_{i,\text{c}}^t + P_{i,\text{w}}^t + P_{i,\text{s}}^t - P_{i,\text{L}}^t \quad (39)$$

$$g\left(\bar{P}_{i,z}^t, \mathbf{Y}\right) \leq 0 \quad (40)$$

$$\frac{P_{ij}^t - P_{ji}^t}{2} = P_{ij}^t \quad (41)$$

$$\bar{P}_{i,z}^t = P_{i,z}^t \quad (42)$$

where T_{opt} represents the P2P transaction optimization period; $P_{i,z}^t$ is the power injected into the distribution network node for the prosumer i ; $\mathbf{P}_{i,\text{opt}}^t$ is the decision variable set for the prosumer i equipment output and power purchase from the power distribution company; $\mathbf{P}_{i,\text{P2P}}^t$ is the P2P electric energy transaction volume decision for the prosumer i variable set; $h_i\left(\mathbf{P}_{i,\text{opt}}^t, \mathbf{P}_{i,\text{P2P}}^t, P_{i,z}^t, \mathbf{X}_i\right)$ is the decision variable related constraints of prosumer i (24), (26)–(34), \mathbf{X}_i is the relevant parameters of prosumer i ; $g\left(\bar{P}_{i,z}^t, \mathbf{Y}\right) \leq 0$ is the network security constraints (12)–(17). \mathbf{Y} is the distribution network related parameters. Equation (41) is rewritten from Eq. (25).

The specific solution process of the ADMM-based P2P power transaction optimization model for prosumer considering network tariffs is as follows:

Step 1: Set the initial value of iteration parameters $P_{ij}^{t,0}$, $\bar{P}_{i,z}^{t,0}$, $\lambda_{ij}^{t,0}$, $\pi_i^{t,0}$ and the allowable range of residual error ε_{pri} , $\varepsilon_{\text{dual}}$, $k = 1$;

Step 2: In the k th iteration, based on the optimization result of the network constraint problem of the $k - 1$ th iteration $\bar{P}_{i,z}^{t,k-1}$, the optimization result of the prosumer j $P_{ji}^{t,k-1}$, the sum of multiplier variables $\lambda_{ij}^{t,k-1}$ and $\pi_i^{t,k-1}$, the prosumer i is optimized according to Eqs. (43)–(45) to solve $\mathbf{P}_{i,\text{opt}}^{t,k}$, $\mathbf{P}_{i,\text{P2P}}^{t,k}$ and $P_{i,z}^{t,k}$, and send the value of the k th round of P2P selectric energy transaction to all the prosumers j who trade with it, and accept all the prosumers j who trade with it;

$$\begin{pmatrix} \mathbf{P}_{i,\text{opt}}^t \\ \mathbf{P}_{i,\text{P2P}}^t \\ P_{i,z}^t \end{pmatrix}^k = \arg \min_{P_{i,\text{opt}}^t, P_{i,\text{P2P}}^t, P_{i,z}^t}$$

$$= \sum_{t \in T} \left\{ C_i(\mathbf{P}_{i,\text{opt}}^t) + \pi_i^{t,k-1} (P_{i,z}^t - \bar{P}_{i,z}^{t,k-1}) + \frac{\rho}{2} (P_{i,z}^t - \bar{P}_{i,z}^{t,k-1})^2 \right. \\ \left. + \sum_{j \in A} \left[|P_{ij}^t| \delta_{ij,\text{net}} + \lambda_{ij}^{t,k-1} \left(\frac{P_{ij}^{t,k-1} - P_{ji}^{t,k-1}}{2} - P_{ij}^t \right) \right] \right. \\ \left. + \frac{\rho}{2} \left(\frac{P_{ij}^{t,k-1} - P_{ji}^{t,k-1}}{2} - P_{ij}^t \right)^2 \right\} \quad (43)$$

$$= \sum_{t \in T} \left\{ C_i(\mathbf{P}_{i,\text{opt}}^t) + \frac{\rho}{2} \left(P_{i,z}^t - \bar{P}_{i,z}^{t,k-1} + \frac{\pi_i^{t,k-1}}{\rho} \right)^2 \right. \\ \left. + \sum_{j \in A} \left[|P_{ij}^t| \delta_{ij,\text{net}} + \frac{\rho}{2} \left(\frac{P_{ij}^{t,k-1} - P_{ji}^{t,k-1}}{2} - P_{ij}^t + \frac{\lambda_{ij}^{t,k-1}}{\rho} \right)^2 \right] \right\} \\ C_i(\mathbf{P}_{i,\text{opt}}^t) = C_{i,\text{dg}}^t + C_{i,\text{bess}}^t + C_{i,\text{re}}^t + C_{i,\text{com}}^t \quad (44)$$

Step 3: Based on the optimization results of all prosumers $P_{i,z}^{t,k}$ in the k th iteration and the multiplier variables in the $k - 1$ iteration, the network constraint problem is solved according to Eqs. (46) and (47), and get $\bar{P}_{i,z}^{t,k}$;

$$\bar{P}_{i,z}^{t,k} = \arg \min_{\bar{P}_{i,z}^t} \sum_{i \in A} \sum_{t \in T} \frac{\rho}{2} \left(P_{i,z}^{t,k} - \bar{P}_{i,z}^t + \frac{\pi_i^{t,k-1}}{\rho} \right)^2 \quad (45)$$

$$\text{s.t. (12) - (17)} \quad (46)$$

Step 4: Based on the results of the prosumer optimization problem and the network constraint problem of the k th iteration, the multiplier variables are updated according to Eqs. (48) and (49);

$$\pi_i^{t,k} = \pi_i^{t,k-1} + \rho (P_{i,z}^{t,k} - \bar{P}_{i,z}^{t,k}) \quad (47)$$

$$\lambda_{ij}^{t,k} = \lambda_{ij}^{t,k-1} + \rho \left(\frac{P_{ij}^{t,k} - P_{ji}^{t,k}}{2} - P_{ij}^t \right) = \lambda_{ij}^{t,k-1} - \frac{\rho (P_{ij}^{t,k} + P_{ji}^{t,k})}{2} \quad (48)$$

Step 5: According to Eqs. (50), calculate the primary residual and dual residual of the k th iteration, and judge whether it is globally converged. If converged, go to Step 6; otherwise $k = k + 1$, and go to Step 2 again;

$$r_i^k = \sum_{i \in A} \sum_{t \in T} \left[\left(P_{i,z}^{t,k} - \bar{P}_{i,z}^{t,k} \right)^2 + \sum_{j \in A} \left(P_{ij}^{t,k} + P_{ji}^{t,k} \right)^2 \right] \leq \varepsilon_{\text{pri}} \quad (49)$$

$$s_i^k = \sum_{i \in A} \sum_{t \in T} \left[\left(P_{i,z}^{t,k} - P_{i,z}^{t,k-1} \right)^2 + \sum_{j \in A} \left(P_{ij}^{t,k} - P_{ij}^{t,k-1} \right)^2 \right] \leq \varepsilon_{\text{dual}} \quad (50)$$

Step 6: Output the prosumer P2P electric energy transaction plan including network tariffs, and the transaction plan should include the P2P electric energy transaction volume, transaction price, and the power purchase power from the power distribution company.

5 Case Study

5.1 Parameter Setting

The improved IEEE 33-node power distribution system is used for case study. It is assumed that there are 12 prosumers participating in P2P transactions, corresponding to prosumers $a \sim l$. The load data and output of renewable energy power generation equipment are taken from reference [11]. To simplify the analysis process, this paper sets the network tariffs price optimization cycle to be the same as the P2P transaction optimization cycle, $T_{\text{opt}} = T = 24$ h. The P2P transaction interval is 1 h. The electricity price parameters are taken from Ref. [15].

5.2 Game Strategy Analysis of Power Distribution Company

Figure 2 shows the P2P transaction volume and the distribution company's revenue under different network tariffs prices. In the process of increasing the network tariffs price coefficient from zero to the equilibrium point, the profit of the distribution company gradually increases, and the P2P transaction volume of the distributed prosumers gradually decreases, which reflects the leading position of the distribution company in the game. After that, continuing to increase the network tariffs price coefficient will lead to a decline in the revenue of the power distribution company, because the excessively high network tariffs price inhibits the demand of prosumers to participate in P2P transactions. Under the optimal strategy, the power distribution company takes into account the income from the network tariffs and the revenue from electricity sales, so as to maximize the total income.

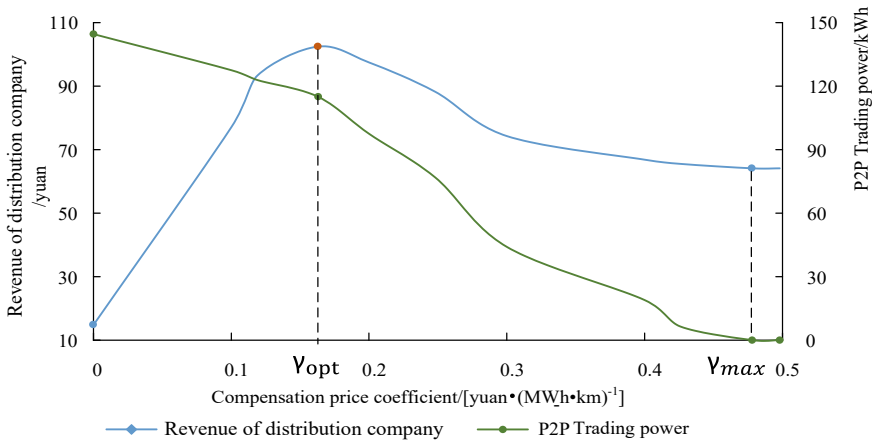


Fig. 2 Revenue of distribution company and P2P transaction amount at different prices

5.3 Distributed Prosumer P2P Transaction Analysis

Figure 3 shows the electricity consumption structure of 12 prosumers in the transaction cycle from 0:00 to 24:00, including P2P electricity transaction volume, electricity purchased from power distribution companies, and self-produced and consumed electricity. It can be seen that from 0:00 to 8:00, the proportion of P2P electric energy transaction volume in the energy consumption structure of prosumers is relatively low, and prosumers mainly meet the electric energy demand by dispatching distributed generators and energy storage devices. In 10:00–20:00, the power generation of renewable energy power generation equipment increased, and the proportion of P2P transaction volume in the energy consumption structure of prosumers increased significantly, indicating that P2P transactions promoted the consumption of distributed renewable energy.

Figure 4 shows the average price of prosumer P2P transactions in each time period. During the low electricity consumption period of 0:00–8:00, the average price of P2P transactions is close to the average network tariffs, while during the peak electricity consumption period of 11:00–14:00 and 18:00–21:00. The average price of P2P transactions is closer to the electricity sales price of power distribution companies. This is because the economic dispatchability of equipment such as distributed generators and energy storage devices of prosumers during the low power consumption period is far greater than the demand for electric energy. The power distribution company purchases very little electricity, and the power supply and demand relationship between prosumers is tense during the peak period of electricity consumption. In addition to self-scheduling, self-production, self-consumption, and P2P transactions, it is also necessary to purchase electricity from the power distribution company to meet the demand.

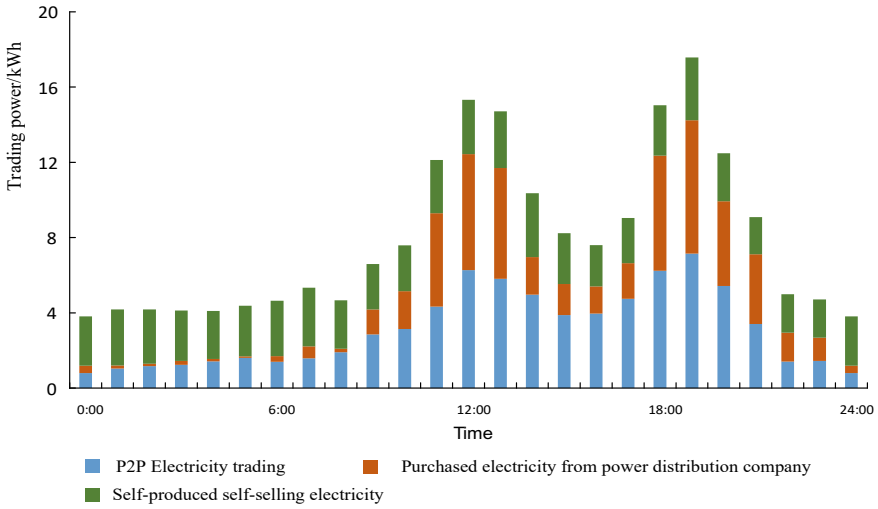


Fig. 3 Energy consumption structure of prosumers for each period

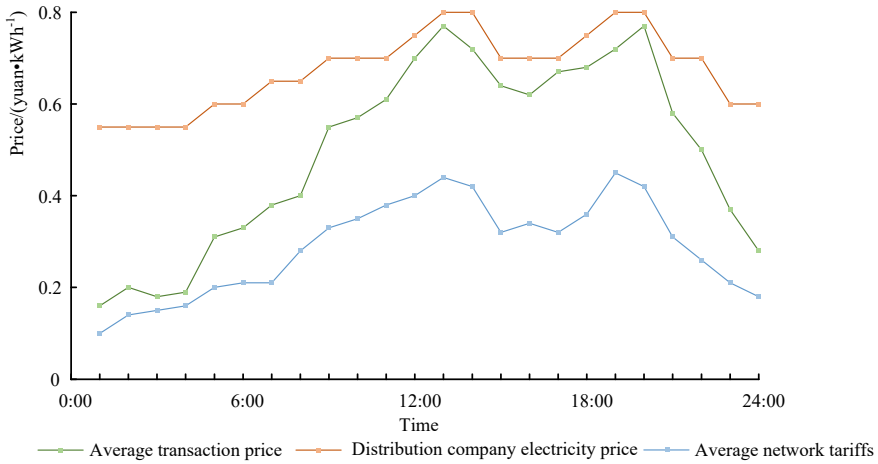


Fig. 4 Average P2P transaction price for each period

5.4 Comparison and Analysis

This paper compares the following three transaction models to analyze the impact of the proposed model on distributed prosumers and power distribution companies, and then illustrates the rationality and feasibility of the proposed model.

Mode 1: The P2P transaction model proposed in this paper.

Mode 2: Considering the franchise rights of power distribution companies, distributed prosumers cannot conduct electricity transactions directly, and can only use electricity themselves or purchase electricity from power distribution companies to meet demand.

Mode 3: P2P transactions can be carried out between prosumers, and the network connection tariffs can only be charged according to the cost of balancing network loss. The power distribution company provides power supply services.

Table 1 shows the results of the comparison mode. In Mode 2, after the self-optimization of 12 prosumers in the transaction cycle from 0:00 to 24:00, the total power purchase demand from the power distribution company is 203.8 kW, and the P2P transaction cost is 132.53 yuan. The total cost of electric energy is 193.05 yuan. Under this model, the total income of the power distribution company in this transaction cycle is 121.28 yuan. In Mode 3, prosumers only need to pay the network connection tariffs corresponding to the network loss cost, so the total cost is reduced, and the income of power distribution companies is greatly reduced.

In Mode 1, 12 prosumers in the trading cycle from 0:00 to 24:00 can successfully trade a total of 114.89 kW of electric energy through distributed distribution, and the total cost of distributed prosumers is 134.86 yuan, which is 30.1% less than Mode 2. In this transaction cycle, the power distribution company's network connection tariffs income is 66.31 yuan and electricity sales income is 48.76 yuan. The network loss cost is 6.33 yuan, and the total profit is 108.74 yuan, which is 1.89 times more than the total net income of Mode 3. Its franchise loss is 85.3% less.

The above results show that the power distribution company can minimize the damage to franchise rights caused by P2P transactions. At the same time, for distributed prosumers, compared with when the power distribution company is fully monopolized, it can reduce electricity costs through distributed P2P transactions. Under the optimal network tariffs price, both power distribution companies and prosumers interests are taken into account.

Table 1 Comparisons of results attained under three modes

Number	Prosumer Cost/ yuan	Distribution Company revenue/yuan			Franchise damages of power distribution companies / yuan
		Sales revenue/ yuan	Network tariffs/ yuan	Loss cost/yuan	
Mode 1	134.86	48.76	66.31	6.33	12.54
Mode 2	193.05	132.53	0	11.25	0
Mode 3	80.77	37.62	9.84	9.84	85.66

6 Conclusion

This paper establishes a prosumer peer-to-peer (P2P) transaction model that takes into account the franchise rights of distribution companies and network constraints. After analysis and comparison of numerical cases, the following conclusions are drawn.

This paper proposes a quantitative assessment method for distribution companies' franchise losses. Modeling the leader–follower game relationship between the power distribution company and the prosumer, and introducing the optimization decision of the power distribution company into the network tariffs pricing model, effectively reduces the franchise loss of the power distribution company. Based on ADMM, the distributed solution of the P2P transaction model including network tariffs can effectively protect the privacy of prosumer. Simulation analysis shows that the proposed model can achieve a win–win situation for both prosumers and power distribution companies.

Acknowledgements This work is supported by National Key Research and Development Program (2022YFB2403104).

References

1. Wu Y, Yang S, Pan Z et al (2023) Complete peer-to-peer transaction mechanism for multiple prosumers without coordination entity. *Autom Electr Power Syst* 47(3):96–103
2. Sorin E, Bobo L, Pinson P (2018) Consensus-based approach to peer-to-peer electricity markets with product differentiation. *IEEE Trans Power System* 34(2):994–1004
3. Zhang F, Gao H, Wu Z et al (2022) Design of P2P trading framework for multiple prosumers in local energy market. *Electr Power Autom Equipment* 42(12):17–25
4. Mohan V, Bu S, Jisma M et al (2021) Realistic energy commitments in peer-to-peer transactive market with risk adjusted prosumer welfare maximization. *Int J Electr Power Energy Syst* 124:106377
5. Wang S, Sun G, Wu C et al (2022) Two-stage robust optimization model of multiple prosumers based on centralized-decentralized trading mechanism. *Electr Power Autom Equipment* 42(5):175–182
6. Shen S, Han H, Zhou Y et al (2022) Electricity-carbon-reserve peer-to-peer trading model for multiple virtual power plants based on conditional value-at-risk. *Autom Electr Power Syst* 46(18):147–157
7. Deng M, Tang Z, Huang Da et al (2021) Power transaction matching mechanism of microgrid based on block chain. *Electr Power Autom Equipment* 41(12):95–101
8. Feng C, Li Z, Shahidepour M et al (2018) Decentralized short-term voltage control in active power distribution systems. *IEEE Trans Smart Grid* 9(5):4566–4576
9. Ilic D (1997) Toward regional transmission provision and its pricing in New England. *Utilities Policy* 6(3):245–256
10. Zolezzi J, Rudnick H (2001) Review of usage-based transmission cost allocation methods under open access. *IEEE Trans Power System* 16(4):933–934
11. Nema S, Mashhadi H (2017) Distribution locational marginal price analysis considering technical constraints. In: 2017 Iranian conference on electrical engineering (ICEE). IEEE, Tehran, Iran, pp 1021–1025

12. Kim J, Dvorkin Y (2019) A p2p-dominant distribution system architecture. *IEEE Trans Power Syst* 35(4):2716–2725
13. Wen A, Huang W, Zhang H et al (2015) Analysis on transmission use of system charging methodology in UK. *South Power System Technol* 9(8):3–8
14. Iria J, Scott P, Attarha A (2020) Network-constrained bidding optimization strategy for aggregators of prosumers. *Energy* 207:118266
15. Baroche T, Moret F, Pinson P (2019) Prosumer markets: a unified formulation. Milan power technology conference. IEEE, Milan, Italy, pp 1–6

Comparative Study of Different Grid Connected Wind Generator



Vaishnavi Pachkawade and Rutuja Hiware

Abstract The primary energy needs of today's day to day life are met by conventional sources, with coal-based thermal power production making up a significant portion of those sources. It is critical to deploy alternative energy technologies that are more dependable of given rate where conventional energy source are use and that gives negative effects on the environment. A global environment to hazards the development of renewable energy sources is now under way. Among the several hydroelectricity sources, wind energy system is more popular economically viable energy sources and the ability to provide all of our energy demands. Wind energy for the production of electricity, has grown recently among the several renewable energy sources utilised to provide power, wind energy is one of the most affordable options with the ability to satisfy our demand for energy. Taking into account the difficulties of connecting big wind farms with various generation types. The researchers have tried multiple times to come up with a way to use wind energy effectively. As a result of thorough research and analysis on the subject, wind energy is widely used hydroelectricity sources of energy.

Keywords DFIG · SCIG · PMSG · Wind turbine system

1 Introduction

All across the world, wind energy is the upcoming energy source. Due to the drawbacks of non-renewable sources, such as the greenhouse effect, which damages the climate and our planet, the generation of energy from renewable sources has expanded significantly globally. Due to this the production of power from wind is the energy technique that is expanding the quickest worldwide. The amount of energy is greatly

V. Pachkawade (✉) · R. Hiware

Electrical Engineering Department, G. H. Raisoni College of Engineering, Nagpur, Maharashtra, India

e-mail: vaishnavi.pachkawade.mtechcps@ghrce.raisoni.net

R. Hiware

e-mail: rutuja.hiware@raisoni.net

influenced by the wind speed [1]. Winds transport enormous amounts of energy and are mostly brought on by solar heating of the atmosphere. A wind turbine is powered by this wind energy and is connected to an electrical generator. The wind machine's blades must revolve in order to use the wind's kinetic energy. When converting wind energy into other forms of energy, the blades are crucial [2]. Generators are used to transform mechanical energy into electrical energy. A wind turbine generator and electrical grid are directly connected. There are downsides, such as reactive power, which make it impossible to manage the level of grid voltage when a wind turbine respect to induction generator is linked directly to the grid [3]. Utilising wind turbines with variable speeds prevents these issues. These turbines lessen noise at low wind speeds and enhance the turbine dynamic behaviour. Power electronic machinery is use to regulate the generator in case a wind turbine with different speed. Wind turbines can be used for acoustic noise reduction, active and reactive power regulation, and mechanical structure stresses. There are several different types of generators, including permanent magnet generators, doubly fed induction generator, and square cage induction generators [4]. Due to a number of benefits, including the most widespread DFIG that is double fed induction generator is increasingly preferred. Due to its better performance, capacity to harness the most wind energy, and affordable induction machine, variable speed double fed induction generator wind turbine (DFIG) is now the most common variant. In this study, we investigate the various system characteristics, including the active power and reactive power, rotor speed, power factor, stator speed, etc.

2 Wind Power Generator's

A wind turbine (WT) and nacelle make up the conventional WECS, i.e. is wind energy conversion system, as depicted. A speedometer is frequently used on the nacelle cover to measure wind speed. The nacelle houses the electrical generator, bearings, gear box, rotor shafts, yaw, control system, and control system. The number of blade options, rotor rotation orientation (downwind), constant of variable rotor speed, gearbox and direct drive generator, and other characteristics are now the most prevalent design criteria for wind turbines. The type of turbine tower (Lattice towers and tube tower, as shown in Fig. 1.5), the axis rotation as either horizontal or vertical, its position by aligning actions (free yaw) or direct control (active yaw) the synchronous or induction generator, the hub design (rigid, hinged), the control of power by aerodynamic stall control of variable pitch blades. The rotor, which is made up of blades coupled to a hub and used to gather wind energy, is controlled by a pitch control system to either maximise wind collection or slow the rotor's spin. Modern wind turbines may operate at variable speeds for cost savings, enhancing power quality and increased efficiency. Reduced mechanical stresses in the turbine system and islands capabilities are additional benefits of different speed wind turbines.

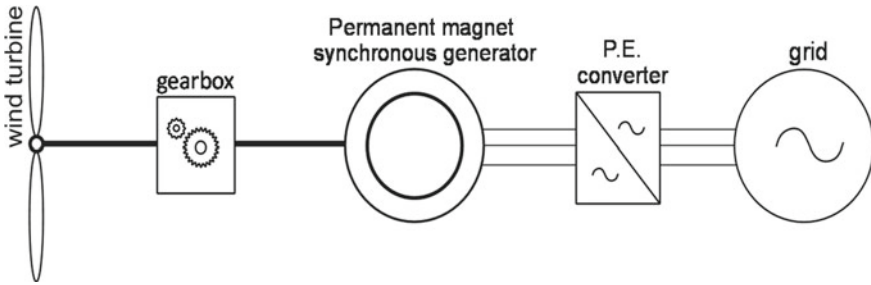


Fig. 1 Permanent magnet synchronous generator

2.1 PMSG

Permanent magnet generators that are synchronous are utilised in varying-speed wind turbines together with axles and completely rated converters, as illustrated in Fig. 1. Better power quality because of their great power density and low bulk, permanent magnet synchronous generators have become more popular recently. Additional benefits of PMSGs include enhanced efficiency and reliability owing to the lack of field winding losses and slide rings and brushes.

Though many manufacturers employ permanent magnet synchronous generators, which are thought of as the best generator for small wind turbine applications, the expense of permanent magnets can occasionally be prohibitive for bigger systems. Synchronous generators commonly have inadequate damping, which prevents unexpected gusts of wind from being adequately absorbed electrically and causes unintended oscillations in the machines, lowering the quality of the electricity. To increase the dependability of wind turbines, DD systems often forgo gearboxes and employ PMSGs with fully rated converters. The requirement for low-speed generator, which are heavier and less effective than high speed generators, arises from the lack of gearboxes.

2.2 SCIG

The resilience, simplicity, great dependability, and affordability of squirrel cage induction devices are well known. In order to produce electricity, squirrel cage induction generators, or SCIGs, revolve faster than synchronous speed. Compared to synchronous machines, squirrel cage induction machines have a higher amount of damping, which improves the rotor speed fluctuation and drive transient absorption. Squirrel cage induction machines are run by fully rated converters and gearboxes in variable speed wind turbine, as seen in Fig. 2. Due to the absence of voltage regulation, SCIGs require reactive power during operation from the utility grid. Voltage instability is another issue with SCIGs.

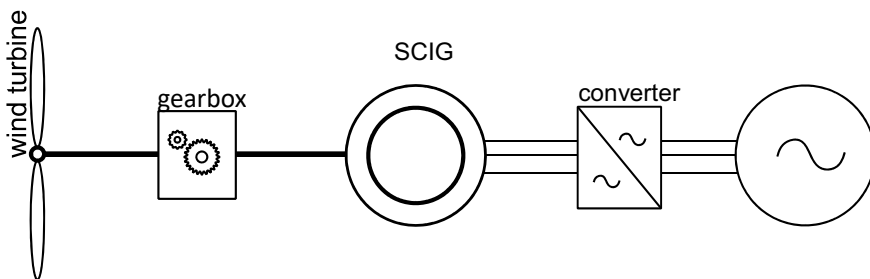


Fig. 2 Squirrel cage induction generator

2.3 DFIG

DFIGs currently outnumber their equivalents in terms of usage frequency. Simply said, DFIGs are wound rotor induction generators that typically link their stator windings to the grid directly while connecting their rotor windings through fractionally rated wires. Figure 3 shows how converters and gearboxes are used in these vehicles' drivetrains. The rotor frequency, current magnitude phase angle may be adjusted thanks to the access of an rotor windings provided by slip rings and carbon brushes [5]. They may therefore be used at a variety of speeds (usually between around 30% of the synchronous speed). Wide operating speed range, simple power factor management, reduced mechanical strains, and power fluctuations are all benefits of DFIGs [6].

Additionally, DFIGs employ partially rated power converter rather than fully loaded converter used with PMSGs and SCIGs, which have reduced converter losses. Early DFIGs had poor grid-fault ride-through capabilities, but more recent research has made these capabilities better. DFIGs provide voltage constant magnitudes and frequency varying mechanical speed by supplying voltage at slip frequency of an rotor contacts to create currents adding the variations in mechanical speed [3]. All of

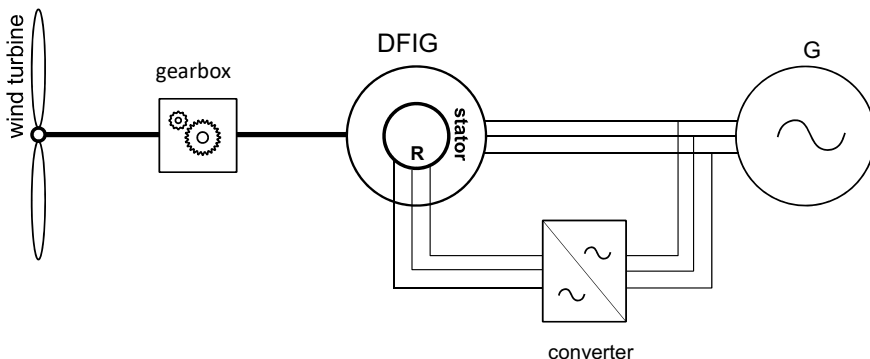


Fig. 3 Doubly fed induction generator

this is realized using DFIGs and a reliable control system that also handles jobs like tracking of maximum power points, pf management, and harmonic filter.

Without respect to power factor specifications, DFIG run in sub-synchronous or positive slip, synchronised, or zero slip and super-synchronous, or negative slip, modes. The rotor speed is slower than the synchronous rate in the sub-synchronous area than it is in the super-synchronous area. While it is in the sub-synchronous mode, the DFIG submit power of an grid throughout the rotor connections at slip frequency. The super-synchronous mode of the grid is powered by the rotor terminals. If the DFIG rotates at synchronous speed, there will be no net power supply or withdrawal and only DC current flowing through the rotor terminals. The stator terminals in all 3 instances provide electricity to the grid when the DFIGs' generating quadrants are taken into consideration.

The converter size determines the speed range that a DFIG can handle. The converter ratings are typically between a quarter and a third of the rated machine power, and they provide or draw power from the stator power at fractional levels. This equals a slip range of around 0.3.

3 Mathematical Modling of Different Wind Turbines

3.1 Modling of SCIG

Here, the 120 kV generator is grounded, and the A, B, and C points are linked to the reactor's secondary half, which is connected to the bus's primary side, according to the block diagram. Buses are used to link several gearbox systems. It serves as the connecting element that links the computer components, then provide the necessary data to them. A bus system is then linked to combine the various lines and transmit data after the secondary part of bus attached with an transformer (step down transformer) that is connected to the ground transformer. To assess the active and passive power of the scope's junction with gearbox line, the secondary side of the bus is connected to it which is linked to the bus, which is linked to the resistive load that supplies electricity to the turbine system, and the bus. We connect a scope to the system to monitor the internal measurements, which include the rotor speed, current stator current, and stator speed. We use the turbine system and the feedback loop as our reference points for measuring the rotor current, stator current, and speed (Fig. 4 and Table 1).

The synchronous reference frame, which is determined by the induction machine equations, is

$$\theta_e(t) = \sum_{-}(t = 0)^z 2\pi f d$$

whereas

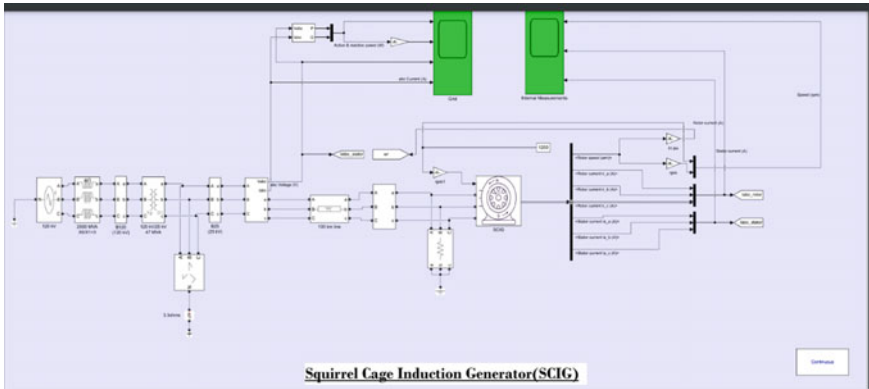


Fig. 4 Block diagram of squirrel cage induction generator

Table 1 The parameter of wind generator

Parameters	Rating
Generator	120 kV
Reactor	2500 MVA
Step down transformer	120 kV/25 kV
Transmission line	100 km
Transmission line	30 km

f the maximum electrical frequency.
 θ electrical angle.

3.2 Modling of DFIG

The block diagram of DFIG is as shown in Fig. 5 and the parameters of the generating is similar to the table.

The formula for the connection of the stator angular frequencies and rotor is follow:

$$(w_r + w_m = w_e)$$

where

W_m Angular frequency.
 W_r rotor angular frequency.

The mathematical form shows the differential arrangement with respect to time:

$$\rho Q_s = (3\iota(mW_e\lambda_{ds}\rho\lambda_{dr}))/ (2\sigma\iota_s\iota_r)$$

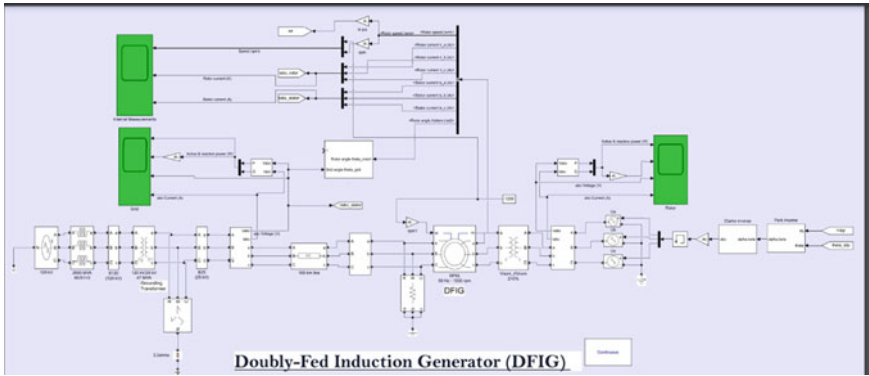


Fig. 5 Block diagram of doubly fed induction generator

3.3 Modling of PMSG

Figure 6 shows the block diagram of permanent magnet synchronous generator and the parameters are in table.

The relationship between EMF and the rotor speed is

$$f = \frac{pn_s}{120}$$

where

p no of poles in generator

f frequency

N_s rotor speed

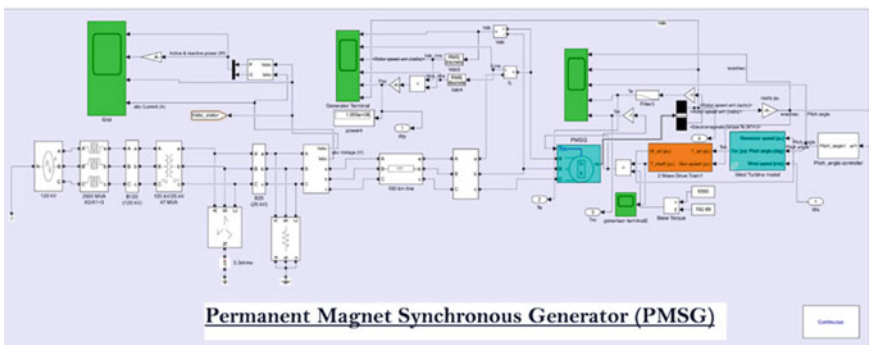


Fig. 6 Block diagram of the permanent magnet synchronous generator

4 Simulation Result

4.1 SCIG

The wave shape of a SCIG is seen in the following Figs. 5 and 7 of the SCIG block diagram illustrates. In order to test the parameters, we use MATLAB. The waveform displays the active power and reactive power over time; at first, the active power is large, but after some fluctuation, the waveform stabilises at a point with 0.9 active power and 0.5 reactive power demonstrates the abc voltage with respect to time at the initial period in two scales of [10]7 and [10]4. As we can see from the abc current measured in amper, the reactive current is strong in the upward direction while the active current is high in the downward direction. abc voltage is high up to 1.7 before becoming stable in SCIG following a 0.05-s fluctuation.

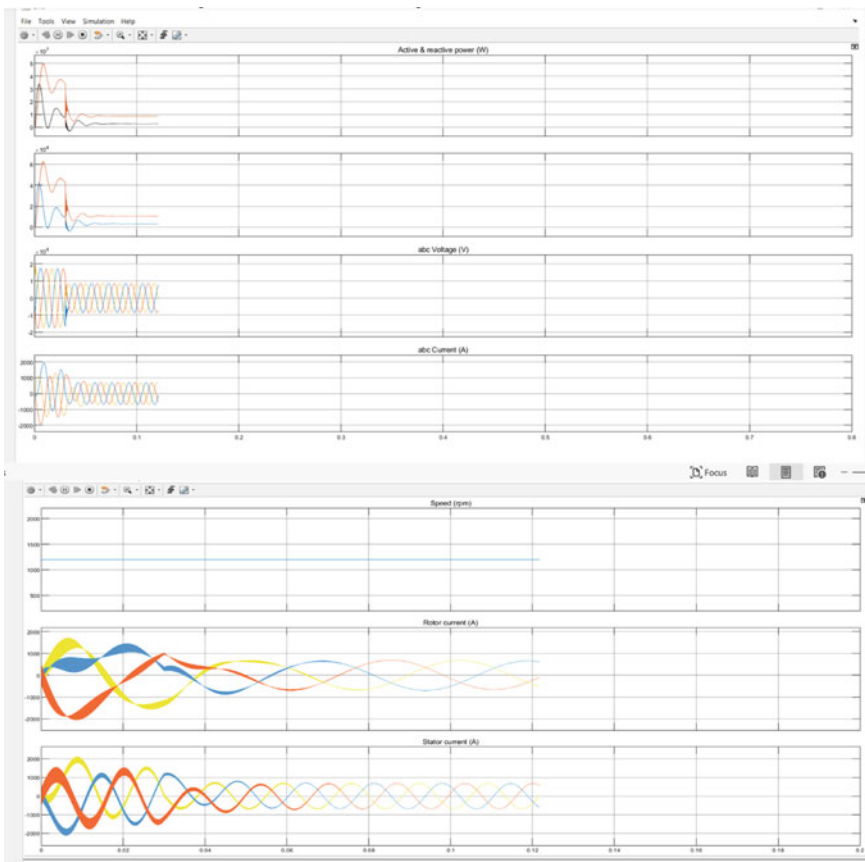


Fig. 7 The wave form of squirrel cage induction generator (with a transmission line of 100 km)

4.2 PMSG

The waveform of a PMSG parameter is shown in Fig. 8 as a function of time. The active power is high, up to 5, and the reactive power is 3.5. The active and reactive power rapidly decline to 0.3 and -0.1 , stabilise for a moment, and then quickly increase to 4 and 2, respectively. After some time, the fluctuation stabilises at 0.2 s. In the case of the abc voltage starting point, the voltage is stable at the span of 0.04; nevertheless, the magnitude decreases until 0.13, at which point it stabilises. Similar to how current at the starting point of a fluctuation happens up to 0.05, becomes unstable up to 0.13, its amplitude drops, and then becomes stable. The highest level of variation in the PMSG system's voltage at startup is up to -2 to 2 with regard to time. The turbine's rotor speed is similarly variable at startup up to 0.34 before becoming stable, although at the point of 0.04 it climbs abruptly up to 0.15. The highest level of variation in the PMSG system's voltage at startup is up to -2 to 2 with regard to time. The turbine's rotor speed is similarly variable at startup up to 0.34 before becoming stable, although at the point of 0.04 it climbs abruptly up to 0.15. The highest level of variation in the PMSG system's voltage at startup is up to -2 to 2 with regard to time. The turbine's rotor speed is similarly variable at startup up to 0.34 before becoming stable, although at the point of 0.04 it climbs abruptly up to 0.15. The highest level of variation in the PMSG system's voltage at startup is up to -2 to 2 with regard to time. The turbine's rotor speed is similarly variable at startup up to 0.34 before becoming stable, although at the point of 0.04 it climbs abruptly up to 0.15.

4.3 DFIG

The stator power, as well as the reactive recording with respect of equation set points shown in Fig. 9. The results highlight how serious the system simulation was during the drop in voltage phase, when both the reactive and active power fluctuated and lost control. The active power decreases virtually to nothing as the machine absorbs the reactive power. Once the voltage dip starts, the rotor-side converter's vector control is momentarily lost. Figures show that the fast magnetic detachment of the DFIG, which is followed by strong, fluctuating reactive and active energies, is the cause of the temporary rise in power production [4]. After the problem is fixed, it takes the power controller 20 ms to normalise the output, which is quite erratic. The DFIG begin to re-magnetise after a fault clearance. This impact, though, only lasts for a fraction of a system cycle.

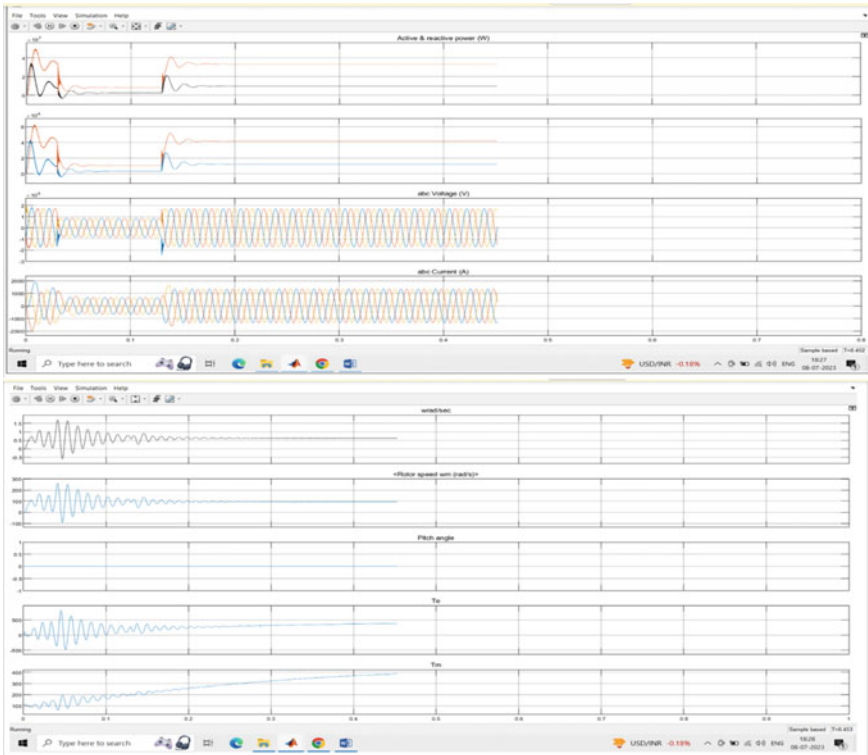


Fig. 8 The wave form of permanent magnet synchronous generator (with a transmission line of 100 km)

4.4 Result

In this paper analyte study with the different types generator of the wind turbine system, i.e. the SCIG, PMSG, and the DFIG. As we compare this three we get to know that, the superiority of the fixed-speed wind generator study is simple, dependable, steady, and well-tested, and the price of its electrical parts is not too high. Mechanical force, unpredictable passive power consumption, and limited control over power quality are some of its drawbacks. As a result of ongoing any differences in wind speed are further conveyed as changes in mechanical torque and electrical power on the grid during speed operation. By using variable wind energy systems, the majority of the problems of constant wind energy system are avoided. Power electronics converters are needed in wind energy systems with varying transmission line and speed. In essence, a wind energy system probably fitted with any three-phase generator, i.e. synchronous generators. Out of them, the DFIG is more favoured due to its many benefits. By optimising the turbine speed, the DFIG technology uses low wind speeds to capture the most energy possible from the wind, as we decrease

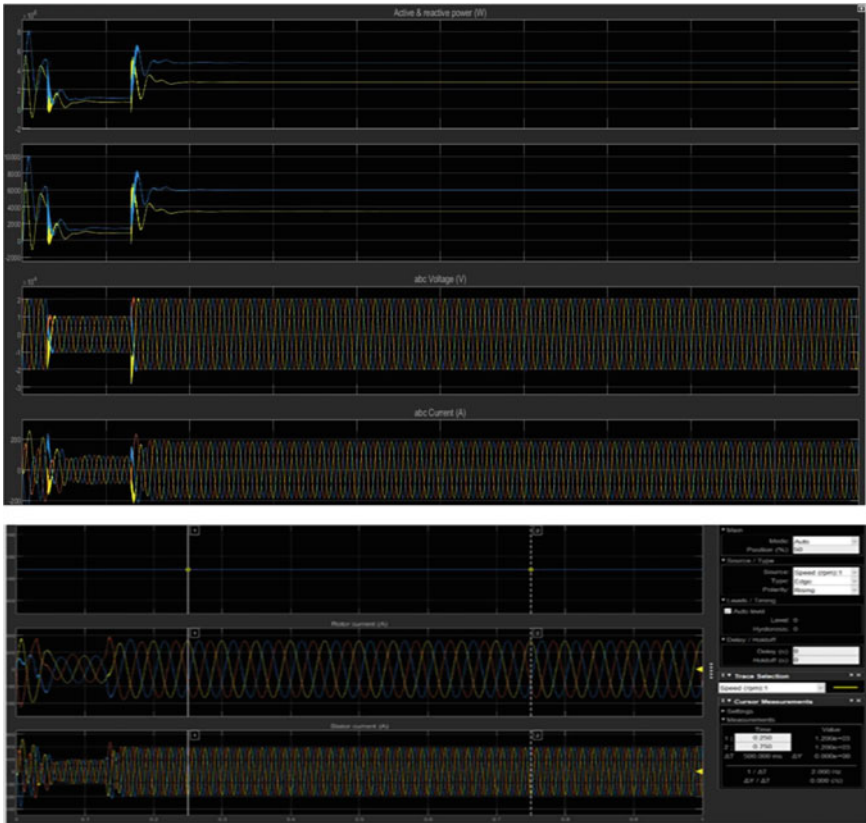


Fig. 9 The wave form of doubly fed induction generator (with a transmission line of 100 km)

mechanical loads attach turbine during a wind burst. Power electronic converters are utilised in DFIG technology to produce or absorb reactive power, negating the need to construct capacitor banks as in SCIG and PMSG. SCIG has two drawbacks: the capacitance value must be adjusted with the generator speed, which reduces efficiency since the generator for the stable of voltage because saturation of iron core. The advantages of PMSG are its compact size, light weight, low losses, excellent efficiency, and lack of a need for a gearbox or external excitation current. The downsides include the need to raise the wind's speed and the permanent magnets' demagnetization as a result of meteorological conditions.

5 Conclusion

This three generator types and their associated characteristics, such as active power, reactive power, power factor, employed by the wind turbine system were discussed in this paper. We compare these three generators of the wind turbine system using MATLAB. The primary goal of comparing these three generators is to determine which one enhances the generator's performance when the transmission line changes and speed as well. Based on our research, the DFIG performs better than the SCIG and PMSG.

References

1. IRENA (2018) Global energy transformation: a roadmap to 2050, 2018th edn. IRENA, Abu Dhabi
2. Fernández LM, García CA, Saenz JR, Jurado F (2009) Equivalent models of wind farms by using aggregated wind turbines and equivalent winds. *Energy Convers Manage* 50:691–704
3. Abad G, Lopez J, Miguel AR, Luis M, Iwanski G (2011) Doubly fed induction machine: modeling and control for wind energy generation, 1st edn. Wiley, INC., Publication
4. Precup R, Kamal T, Hassan SZ (2019) Advanced control and optimization paradigms for wind energy systems. Springer, Singapore. ISBN 978-981-13-5994-1
5. Manwell JF, McGowan JG, Rogers AL (2009) Wind energy explained theory, design, and application, 2nd edn. Wiley
6. Wu B, Lang Y, Zargari N, Kouro S, Institute of Electrical and Electronics Engineers (2011) Power conversion and control of wind energy systems. Wiley Blackwell, Oxford
7. Pleßmann G, Erdmann M, Hlusiak M, Breyer C (2014) Global energy storage demand for a 100% renewable electricity supply. *Energy Procedia* 46:22–31
8. Bussar C, Moos M, Alvarez R, Wolf P, Thien T, Chen H, Cai Z, Leuthold M, Sauer DU, Moser A (2014) Optimal allocation and capacity of energy storage systems in a future European power system with 100% renewable energy generation. *Energy Procedia* 46:40–47
9. IEC White Paper Energy Challenge (2010) Coping with the energy challenge the IEC's role from 2010 to 2030. IEC, Geneva, Switzerland
10. Sun, Z, Wang H, Li Y (2012) Modelling and simulation of doubly-fed induction wind power system based on Matlab/Simulink. *IET Conf. Publ.* 2012
11. Jami H (2014) World wind resource assessment report, 2014th edn. World Wind Energy Association, Bonn, Germany
12. Rolán A, Pedra J, Córcoles F (2014) Detailed study of DFIG-based wind turbines to overcome the most severe grid faults. *Int J Electr Power Energy Syst* 62:868–878
13. Krim Y, Abbas D, Krim S, Mimouni MF (2018) Intelligent droop control and power management of active generator for ancillary services under grid instability using fuzzy logic technology. *Control Eng Pract* 81:215–230
14. Fox B, Flynn D, Bryans L, Jenkins N, Milborrow D, Malley MO, Weston R, Olimpo Anaya-lara (2007) Wind power integration connection and system operational aspects, vol 50. *IET Power and Energy Series*
15. Lynn PA (2012) Onshore and offshore wind energy. Wiley
16. Amin AZ (2013) Renewable power generation costs in 2012: an overview. International Renewable Energy Agency
17. Blaabjerg F, Chen Z (2006) Power electronics for modern wind turbines. Morgan & Claypool, San Rafael

18. Zhang Z, Matveev A, Øvrebø S, Nilssen R, Nysveen A (2011) Review of modeling methods in electromagnetic and thermal design of permanent magnet generators for wind turbines. *IEEE ICCEP*, pp 377–382
19. Yang X, Patterson D, Hudgins J (2012) Permanent magnet generator design and control for large wind turbines. *IEEE PEMWA*, pp 001–005
20. Baroudi JA, Dinavahi V, Knight AM (2005) A review of power converter topologies for wind generators. *IEEE IEMDC*, pp 458–465
21. Bhukya J (2013) Modeling and analysis of double fed induction generator for variable speed wind turbine. *IEEE ICEETS*, pp 1324–1329
22. Polinder H, Ferreira JA, Jensen BB, Abrahamsen AB, Atallah K, McMahon RA (2013) Trends in wind turbine generator systems. *IEEE J Emerg Sel Top Power Electron* 1(3):174–185
23. Verma N, Pachori A (2015) Theoretical approach for comparison of various types of wind generator systems. *Int J Recent Res Electr Electron Eng (IJRREEE)* 2(2):(29–35)
24. Kapse SSS, Daigavane MB, Daigavane PM (2020) Optimal localization and sizing of UPFC to solve the reactive power dispatch problem under unbalanced conditions. *IETE J Res* 66(3):396–413
25. Sarkar M, Subudhi B, Daigavane P (2019) Robust PI damping controller design using wide-area signal for inter-area oscillation in power system. In: 2019 8th international conference on power systems (ICPS). *IEEE*, pp 1–6
26. Chaudhari M, Babu K, Khubalkar SW, Daigavane P (2019) Off-grid hybrid solar power conditioning unit for critical and non-critical loads. In: 2019 international conference on intelligent computing and control systems (ICCS). *IEEE*, pp 969–974
27. Katole DN, Daigavane MB, Gawande SP, Daigavane PM (2018) Improved single phase instantaneous pq theory for DVR compensating nonlinear load. In: 2018 IEEE international conference on power electronics, drives and energy systems (PEDES). *IEEE*, pp 1–6
28. Katole DN, Daigavane MB, Gawande SP, Daigavane PM (2018) Vector based analysis for design of single phase SRF controller in Dynamic Voltage Restorer. In: 2018 IEEE international conference on power electronics, drives and energy systems (PEDES). *IEEE*, pp 1–6
29. Lanjewar A, Khubalkar SW, Junghare AS (2018) Comparative analysis of two loop integer and fractional order PID controller for inverted pendulum. *ICSEDPS 2018, GHRCE, Nagpur*
30. Bose S, Khubalkar S (2022) Power quality analysis of textile industry—findings and recommendations. In: 2022 2nd Asian conference on innovation in technology (ASIANCON). *IEEE*, pp 1–6
31. Shende AG, Khubalkar SW, Vaidya P (2021) Hardware implementation of automatic power factor correction unit for industry. *J Phys Conf Ser* 2089(1):012032 (IOP Publishing)
32. Chaudhari M, Babu K, Khubalkar SW, Talokar S (2019) Off-grid hybrid online solar power conditioning unit for domestic purposes. In: 2019 international conference on computing, power and communication technologies (GUCON). *IEEE*, pp 121–126
33. Shende D, Jagtap P, Hiware R (2021) Enhanced power quality using unified power flow controller systems. *J Phys Conf Ser* 2089(1):012035 (IOP Publishing)
34. Shrawane Kapse SS, Daigavane MB, Daigavane PM (2018) Improvement of ORPD algorithm for transmission loss minimization and voltage control using UPFC by HGAPSO approach. *J Inst Eng (India) Ser B* 99:575–585
35. Shende D, Jagtap P, Hiware R (2021) Review of enhanced power quality using unified power flow control system in electrical network. *J Phys Conf Ser* 2089(1):012034. <https://doi.org/10.1088/1742-6596/2089/1/012034>
36. Balaganur M, Parimala V, Arunsrinivas T, Muthuraj B, Hiware R, Umathe S (2022) DFIG powered yaw function for wind energy network. In: 2022 international conference on automation, computing and renewable systems (ICACRS). <https://doi.org/10.1109/ICACRS55517.2022.10029128>

37. Hiware RS, Chaudhari JG (2012) Indirect field oriented control for induction motor. In: 2011 fourth international conference on emerging trends in engineering & technology. IEEE. <https://doi.org/10.1109/ICETET.2011.56>
38. Adware R, Chandrakar V (2023) Power quality enhancement in a wind farm connected grid with a fuzzy-based STATCOM. Eng Technol Appl Sci Res 13(1):10021–10026

Photovoltaic with Battery and Supercapacitor Energy Storage System for Better Performance Devices and Modelling



Somesh Harinkhede and Prajakta Vaidya

Abstract This paper's objective is to show how battery and supercapacitor devices are superior. When compared with traditional battery energy storage systems (BEES), the proposed different energy storage system by battery and supercapacitor has advantages that it can store surplus energy and use it again when necessary. This paper discusses several energy storage systems that can be utilized with renewable energy sources like solar energy and as remote or backup energy storage systems when there is no functioning electrical grid. In order to maximize this system's efficiency, supercapacitors will be employed in parallel with the battery and load pulsed. In addition to the foregoing, this paper presents the modelling of battery and supercapacitor-based different energy storage systems using MATLAB/Simulink software.

Keywords Photovoltaic (PV) · Energy storage system (ESS) · Battery energy storage system (BESS) · Supercapacitor (SC) · Supercapacitor energy storage system (SCESS)

1 Introduction

An energy storage system, which is the process of converting excess electricity into other kinds of energy, can then be used to produce electrical energy (ESS). We concentrate on battery and supercapacitor energy storage systems among others, but energy storage systems (ESS) can be applied to both traditional and renewable energy sources, storing energy in the form of mechanical, electrostatic, electrochemical, thermal energy, etc., that can be used whenever necessary. When demand exceeds supply, there is an imbalance between the two, which leads to various issues with the power grid, including decreased power quality, decreased efficiency, decreased

S. Harinkhede (✉) · P. Vaidya
Department of Electrical Engineering, G. H. Raison College of Engineering, Nagpur, India
e-mail: somesh.1415@gmail.com

P. Vaidya
e-mail: prajakta.vaidya@raisoni.net

dependability and stability of the system, and introduces many losses that are avoided by ESS systems. Energy storage devices are commonly utilized in both permanent and transient activities, making them one of the most prominent and effective instruments for the proper operation of smart grids and micro grids. Electric energy storage systems (EESS) are frequently utilized for frequency and voltage control (stability enhancement), as well as dynamic compensation of energy with high renewable energy penetration, in the latter instance. Batteries are a common technique for peak shaving in the former [1–3].

The following are four more major advantages of EESS:

- (a) There is no need to convert electrical energy into chemical or mechanical energy.
- (b) Bidirectional functionality.
- (c) A deep discharge capacity and high-power density.
- (d) Less useable life deterioration per charge/discharge cycle. EESS frequently includes flywheel energy storage (FWES), superconducting magnetic energy storage (SMES), and supercapacitor energy storage (SCES) technologies.

In order to preserve system stability and prevent the negative effects of power transients on battery life, the battery/supercapacitor hybrid energy storage system (HESS) concept was developed. In this hybrid system, batteries and supercapacitors are used combined. Batteries, with their enormous storage capacity, offer a steady source of energy, while SC, with their high power efficiency and rapid adaptability, supply unexpected peak powers during transients. As a result, the battery experiences less strain, lives a longer time, avoids expensive battery replacement, and the system reliability is improved.

In this study, a solar power system that operates independently with an operational battery and SC HESS is investigated. A method is developed to control power sharing between the PV and ESS sections in order to assure energy balance and preserve voltage stability at the DC-bus. The battery makes up for consistent power drifts while the SC is forced to handle quick transients by the applied ESS control. In addition, the PV control has been changed to optimize charging a battery according to its maximum state of charge (SOC), preventing degradation and lengthening battery life. To test the effectiveness of the system under various load and irradiance profiles, simulation tests are conducted using MATLAB/Simulink. The latter is reassured by a new test of system performance using a commercial load and real-time weather information [2–4].

2 Energy Storage Systems Devices

2.1 Energy Storage System (ESS)

Wind and photovoltaic (PV) energy are two examples of renewable energy sources that are widely employed as independent power systems to support a variety of electrical demands in remote and rural areas. Because of the sporadic nature of these sources, storage battery banks should be included in stand-alone power systems. When extra energy is stored in storage battery banks and sent to the load bank when solar or wind energy is unavailable or insufficient, this increases system reliability. As storage battery banks, mostly lithium-ion and lead-acid batteries are employed. After hundreds of charging-discharging cycles, battery energy storage systems (BESS) require routine battery replacement because cycle efficiency is low. Because of its high-energy density, efficiency, long battery life, low cost, and environmental friendliness, lead acid has advantages in the energy storage system (ESS) market. Because lead-acid batteries have a cheap cost per energy, they are appropriate for large-scale energy storage. Both pulsing power loads and continuous power loads can be handled by lead-acid batteries [3].

2.2 Lead-Acid Battery

The energy storage system (ESS) industry benefits from lead-acid batteries' high-energy density, efficiency, long battery life, low cost, and environmental friendliness. Lead acid batteries are useful for large-scale energy storage because of their relatively low cost per energy. Both steady and pulsing power loads can be handled by lead-acid batteries.

2.3 Lithium-Ion Cell

Lithium-ion batteries have a greater energy density than lead-acid batteries, a longer lifespan, higher efficiency, less weight, and are more environmentally friendly, but they are also more expensive. Mobile and automotive applications, among others, frequently use lithium-ion batteries.

2.4 Different Energy Storage Methods

A method for storing energy is something that differs from two or more other energy storage systems. In this study, we employed both a supercapacitor battery energy

storage system and an energy storage system. The photovoltaic system benefits from the various systems for storing energy in batteries (BESS) and supercapacitors (SCESS), such as the ability to meet peak power demands temporarily, stabilize system voltage, enhance system capabilities, etc. This means battery supercapacitor-based energy storage systems (BSESS) increase the efficiency of the system. Peak load demand and load demand are consistent because diverse storage methods based on battery supercapacitors are too expensive for large-scale deployment. A hybrid storage system can provide more specific power than a battery storage system when a supercapacitor is added. This happens when the system efficiency increases. As the supercapacitor has a higher power density and can thus deliver more power for a shorter time or peak power for a shorter time, we may conclude that the hybrid storage system's charging capacity has increased. Supercapacitors (SC) play a key function in raising the buffer level in various energy storage systems. The battery provides the lower constant power requirement, while the supercapacitor provides reduced size of the battery pack for large storage while providing the load's peak power needs. We may claim that the battery improves the system's storage capacity and the peak power requirements of the load, resulting in a reduction in the size of the battery pack for big storage. We may say the battery enhances the storage system capacity and reduces the system's discharge capacity because it can store more energy and release it over a longer period of time, giving it a much higher density [3–5].

2.5 Benefits of a Battery Supercapacitor Energy Storage System

Long cycle life, energy buffering, increased reliability, and high cycle efficiency are all being pushed as high-energy density and power density, and low rate of self-discharge. Enhance uniformity and efficiency, boost the effectiveness of the electricity system, and use low-cost, lightweight technology for widespread deployment [5–8].

3 Methodology

Consider the electrical connections and control mechanisms between the battery and the supercapacitor (Fig. 1).

Electrical connections and control mechanisms in a photovoltaic (PV) system with a battery and supercapacitor hybrid storage system mechanisms are crucial for proper integration and optimal performance. Here are the key aspects of the electrical connections and control mechanisms between the PV system, battery, and supercapacitor.

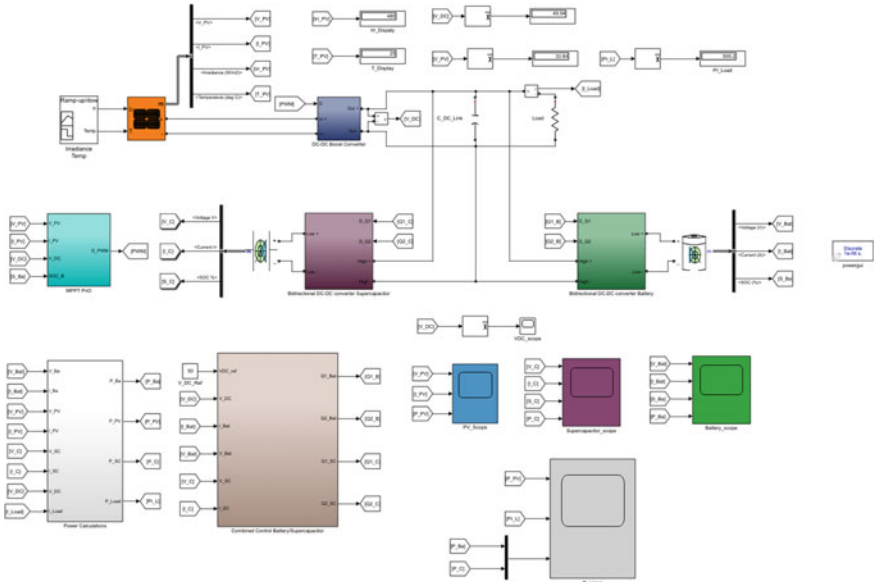


Fig. 1 Block diagram

3.1 Electrical Connections

PV Array to Battery: The photovoltaic array is connected through a charge controller and the battery. The charge controller regulates the charging process, ensuring that the battery receives the appropriate current as well as voltage from the PV array.

Battery to Supercapacitor: The battery and supercapacitor are interconnected through a DC–DC converter. This converter enables the transfer of energy between the two storage devices, allowing efficient energy management and balancing between them.

Supercapacitor and Battery to Load: Both the battery and supercapacitor are connected to the load, which represents the electrical devices or systems that consume energy. The load receives power from either the battery, supercapacitor or a combination of both, depending on the system’s control mechanism and energy requirements.

3.2 Control Mechanisms

Energy Management System: The photovoltaic (PV) system, battery, with supercapacitor are all under the supervision of the energy management system (EMS). The EMS monitors the state of charge (SOC) and state of health (SOH) of the battery and supercapacitor, as well as the PV array’s output. Based on this information and

predefined algorithms, the EMS determines the optimal distribution and utilization of energy between the battery, supercapacitor, and load.

Charger Controller: The controller that controls the charge is necessary for controlling how the battery is charged. In order to avoid overcharging or undercharging, which can reduce battery performance and longevity, it makes sure that the battery receives the proper voltage and current from the PV array.

DC–DC Converter: The converter DC–DC between the battery and supercapacitor manages energy flow and ensures efficient transfer between the two storage devices. It controls the charging and discharging processes, allowing energy to be transferred based on the system's needs and performance requirements.

These electrical connections and control mechanisms enable the seamless integration of the PV system, battery, and supercapacitor. The coordination between these components ensures efficient energy harvesting from the PV array, optimal storage, and discharge of energy using the battery and supercapacitor, and effective management of energy distribution to the load. The control mechanisms play a vital role in maintaining system stability, maximizing energy utilization, and extending the lifespan of the storage devices.

3.3 Elements Are Interconnected as Follows

The DC-DC converter, which controls the voltage and current from the PV array, is linked to the PV array and provides the appropriate input to charge the battery and supercapacitor. Supercapacitor and battery are linked in parallel through the DC–DC converter, allowing energy transfer between the two storage devices based on the system's control mechanism. The load is coupled with the supercapacitor and battery, drawing power from either or both, depending on the energy management system's control strategy. The equivalent circuit model captures the key electrical connections and components in the PV system with a battery and supercapacitor. It provides a simplified representation that helps in analyzing and understanding the energy flow and interactions between the different components in the system [4–15].

3.4 Power Flow and Control

In this study, a power flow control approach is presented to maintain the necessary balance between the production and consumption of energy, maintaining steady DC load voltage. This entails managing the bidirectional converters in the ESS elements to coordinate the distribution of power between the battery and the supercapacitor. It also comes into contact with PV boost converter control in order to accomplish PV MPPT. When the battery is full, the MPPT should switch to DC voltage control mode to prevent battery overcharging.

Power Flow

PV Power Generation: The PV array converts solar energy into electrical power. The generated power depends on factors such as solar irradiation, temperature, and shading. The energy storage system and the load are both powered by the PV array.

Charging and Discharging of Batteries: Battery is charged during periods of excess PV power generation or low load demand. Transferring electrical energy from the PV array to the battery is the process of charging. When demand for load is high or low PV power generation, the battery discharges stored energy to supply power to the load.

Supercapacitor Charging and Discharging: The supercapacitor is primarily used for short-term power requirements and fast response times. It rapidly charges and discharges electrical energy to compensate for sudden load fluctuations and provides power during transient events.

Power Distribution to the Load: The load receives power from both the battery and supercapacitor, depending on the system control strategy. The control mechanism determines the optimal power distribution to the load based on factors such as load demand, available energy from the PV array, battery state of charge (SOC), and supercapacitor voltage [8–12].

Control Strategies

Energy Management System (EMS): The power flow between the PV system, battery, and supercapacitor is monitored and managed by an EMS. It optimizes energy utilization and ensures the efficient operation of the system.

Charge Control: The charge controller regulates the charging process of the battery, ensuring optimal charging voltage and current levels. It prevents overcharging or undercharging, which can affect battery performance and lifespan.

Power Electronics Control: The DC–DC converter between the battery and supercapacitor manages the energy flow and facilitates efficient transfer between the two storage devices. The converter's control mechanism adjusts the voltage conversion ratio and controls the direction of power flow based on system conditions and requirements.

Load Management: The control system manages the distribution of power to the load, optimizing the utilization of available energy from the PV array, battery, and supercapacitor. It considers factors such as load demand, priority settings, and battery and supercapacitor states to determine the plan for distributing the power.

State of Charge (SOC) Voltage Control: The control system monitors the SOC of the battery and voltage of the supercapacitor to ensure they remain within safe and optimal operating ranges. It may implement algorithms to balance the charging and discharging rates to extend the lifetime of the storage devices.

By effectively controlling the power flow and managing energy storage, the system can ensure stable power supply, minimize reliance on the grid, optimize energy utilization, and enhance overall system performance and reliability. It is important for one to understand that the system design might affect the specific power flow and

Table 1 Array module data

Parallel strings	4
Series connect module per string	2
Module Waaree energies	WV-120
Maximum power	120.7 W
Cell per module	72Ncell
Open circuit voltage (<i>Voc</i>)	21 V
Short circuit (<i>Isc</i>)	8 A
Voltage at maximum power point (<i>Vmp</i>)	17 V

control tactics, requirements, and the desired application. Advanced control algorithms and monitoring techniques are often employed to achieve optimal performance and efficiency in PV systems with supercapacitor and battery energy storage.

In summary, the hybrid storage system using batteries and supercapacitors employs chemical reactions in batteries and electrostatic charge separation in supercapacitors to store and release electrical energy. By combining the strengths of these mechanisms, the hybrid system achieves a balance between high-energy capacity and rapid power delivery, enhancing overall performance and ensuring a reliable and efficient energy storage solution [5, 6, 8–23].

4 Calculations and Tables

At various irradiance measurements from the PV array, the characteristics between current and voltage explain the voltage/current and power/voltage characteristics. The irradiance values on the model validate the array's operation and are calibrated in KW/m^2 . The highest power is at 1000 irradiance when the current reading is 8 A and the voltage is 21 V. According to the results, the voltage at the 0 V reading is 21 V, the open circuit voltage of the PV module and the current at the 0 V reading is 8 A, which represents the short circuit current. The next is a representation of the shift in voltage and power. The highest output power is listed as 12 W with a voltage of 21 V at 1000 irradiance, although this fluctuates (Tables 1 and 2).

5 Result

In Fig. 2 supercapacitor waveform shows the initial voltage (32 V), current (0–10 A), SOC (98.9–99%), and power (0–500).

Simulations may show the outcomes and the system's effectiveness in fulfilling the load's energy requirements and coordinating. The real output voltage's reaction is simulated in the simulation, current, SOC, power of supercapacitor.

Table 2 System component data

Data/quantities	Supercapacitor	Battery
Rated capacitance/capacity	29 F	14 Ah
Operating temperature	25 °C	25 °C
Rated voltages	32 V	24 V
Fully charged voltage	32 V	27.9357 V
Nominal discharge current	10 A	6.087 A
Charge current	10 A	5 Ah
Initial state of charge	98.9%	50%

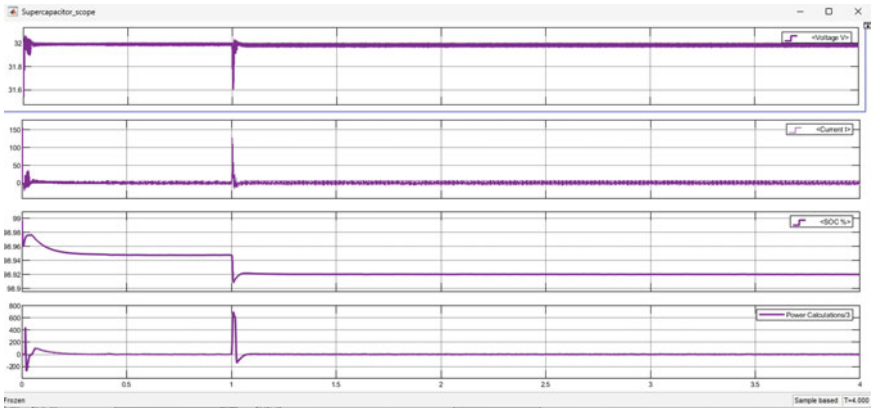


Fig. 2 Behaviour of voltage and current of supercapacitor

For supercapacitor

- X axis = time in second ($t = 01-04$ s).
- Y axis = voltage (32 V).
- Y axis = current (10 A).

With a starting voltage of 32 V and a current of 10 A, figure depicts the simulation findings and system performance in fulfilling the energy demands of the load and coordination.

The above graph depicts the responsiveness of the real output voltage and current. As illustrated in the simulation, the system reacts by continuing from its origin point and attaining steady state response at time $t = 1$ s and constant voltage as voltage and current are gradually altered from 32 V at time $t = 0$, 31.6 V at time $t = 1$ s, 32 V at time $t = 2$ s, and 32 V at time $t = 3$ s.

Figure 3 shows the battery response with voltage (25–26.5), current (0–10 A), and SOC (50%).

For battery

- X axis = time in second ($t = 01-04$ s).

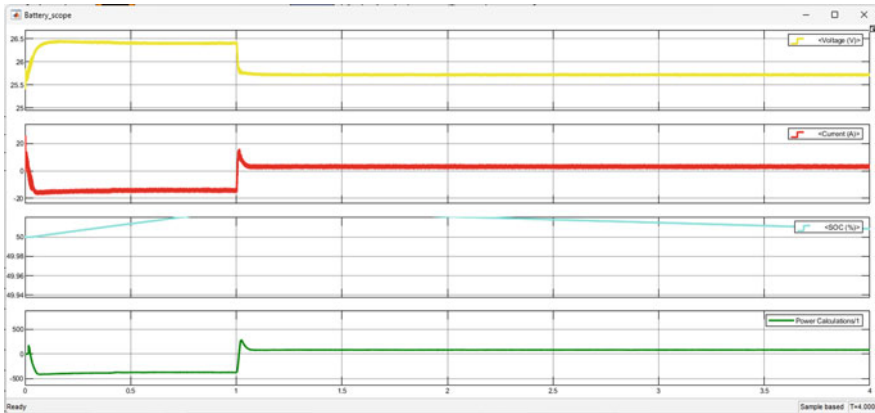


Fig. 3 Response of voltage and current battery

Y axis = voltage (24 V).

Y axis = current (0–10 A).

With a starting voltage of 25.5 V and a current of (0–10A), the figure depicts the simulation findings and system performance in fulfilling the energy demands of the load and coordination. The graph depicts the responsiveness of the real output voltage and current. As illustrated in the simulation, voltage and current are progressively changed.

6 Conclusion

The outcomes show that the overall control technique effectively regulated the power flows between system components in the model. As a result, it established the required balance, met the variable load demand, and preserved DC voltage stability. Supercapacitor significance in enhancing performance and offering the necessary buffer during transients was also clear and readily apparent.

It can be determined from the analysis and discussion findings from the design and modelling of PV integration using the simulation tools MATLAB/Simulink, this combined ESS modelling approach not only enhances system performance but also promotes the utilization of renewable energy sources and reduces reliance on conventional grid infrastructure. By efficiently storing and utilizing excess solar energy, the dependence on non-renewable energy sources is minimized, contributing to a more sustainable and environmentally friendly energy landscape. The project is accelerating the transition to a cleaner and more resilient energy future and the potential of renewable energy.

References

1. Wang M, Yu L, Wang H, Xu W (2019) Optimal sizing and control strategy of PV-battery-supercapacitor energy storage system for residential applications. *Appl Sci* 9(23):5151
2. Feng Z, Hu X, Qi Z, Liu Y, Shao Z (2019) An energy management system for a photovoltaic-battery-supercapacitor hybrid power system. *Energies* 12(19):3721
3. Zhang Y, Zhang W (2021) Energy management and capacity optimization of photovoltaic-battery-supercapacitor hybrid system with fuzzy control. *Appl Energy* 289:116645
4. Baghmare K, Daigavane PM (2021) Supercapacitor based controller for electric vehicle to grid power. In: 2021 second international conference on electronics and sustainable communication systems (ICESC). IEEE, pp 1–8
5. Lindberg DV, Lee HKH (2015) Optimization under constraints by applying an asymmetric entropy measure. *J Comput Graph Statist* 24(2):379–393. <https://doi.org/10.1080/10618600.2014.901225>
6. Rieder B (2020) *Engines of order: a mechanology of algorithmic techniques*. Amsterdam University Press, Amsterdam
7. Boglaev I (2016) A numerical method for solving nonlinear integrodifferential equations of Fredholm type. *J Comput Math* 34(3):262–284. <https://doi.org/10.4208/jcm.1512-m2015-0241>
8. Logeris P-O, Riou O, Camera MA, Durastanti J-F (2013) Study of photovoltaic energy storage by super capacitors through both experimental and modeling approaches (Hindawi Publishing Corporation). *J Solar Energy* 2013(659014):9. <https://doi.org/10.1155/2013/659014>
9. Musolino V (2012) Supercapacitor storage systems: modeling, control strategies, and applications and sizing criteria. Ph.D. thesis in Electrical Engineering XXIV Cycle 2009-2011 (Matriculation number: 738557)
10. Vaidya P, Chandrakar VK (2022) Exploring the enhanced performance of a static synchronous compensator with a super-capacitor in power networks. *Eng Technol Appl Sci Res* 12(6):9703–9708
11. Rajderkar VP, Chandrakar VK (2023) Design coordination of a fuzzy-based unified power flow controller with hybrid energy storage for enriching power system dynamics. *Eng Technol Appl Sci Res* 13(1):10027–10032
12. Mahale P, Joshi KD, Chandrakar VK (2009) Static synchronous compensator (STATCOM) with energy storage. In: 2009 second international conference on emerging trends in engineering & technology, Nagpur, India, 2009, pp 560–563. <https://doi.org/10.1109/ICETET.2009.213>
13. Padwad D, Naidu HK (2022) A review of electrical BoatVehicle solar generation for energy sustainability and development. In: 2022 10th international conference on emerging trends in engineering and technology-signal and information processing (ICETET-SIP-22). IEEE, pp 1–6
14. Edke K, Adware RH (2017) Improvement in dynamic performance of an three phase induction motor by using ultra capacitor. In: 2017 third international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB), Chennai, India, 2017, pp 515–519. <https://doi.org/10.1109/AEEICB.2017.7972367>
15. Boglaev (2016) A numerical method for solving nonlinear integrodifferential equations of Fredholm type. *J Comput Math* 34(3):262–284. <https://doi.org/10.4208/jcm.1512-m2015-0241>
16. Sun M, Zhang S, Mei S (2020) Optimal sizing and energy management strategy for a photovoltaic-battery-supercapacitor hybrid system. *Appl Energy* 276:115408
17. Li R, Lin X, Cheng K, Gu W, Jiang Z (2021) Optimal sizing and scheduling of a stand-alone photovoltaic-battery-supercapacitor system considering load uncertainty. *IEEE Trans Smart Grid* 12(4):3121–3132
18. Deshmukh SR, NS BK, Saharabudhe S, Khubalkar S, Parameswaran AP (2019) Designing of control strategy for high voltage battery isolation in an electric vehicles. In: 2019 IEEE 5th international conference for convergence in technology (I2CT), Bombay, India, pp 1–4. <https://doi.org/10.1109/I2CT45611.2019.9033891>

19. Kakde AA, Tarnekar SG, Daigavane PM (2018) MatLab simulation of reconfigured ultracapacitors for co-working with battery for faster energy exchange. In: 2018 international conference on smart electric drives and power system (ICSEDPS), Nagpur, India, pp 1–6. <https://doi.org/10.1109/ICSEDPS.2018.8536005>
20. Gudadhe AR, Daigavane PM, Tutkane DR (2018) Reversible battery operated vehicle drive with regenerative braking & multitasking application. In: 2018 second international conference on electronics, communication and aerospace technology (ICECA), Coimbatore, India, pp 679–684. <https://doi.org/10.1109/ICECA.2018.8474654>
21. Patil M, Vadirajacharya K, Khubalkar SW (2020) Design of fractional order controllers using constrained optimization and reference tracking method. *Int J Power Electron Drive Syst* 11(1)
22. Ramtekkar P, Harikumar N, Suraj D (2022) An innovative safe electrical vehicle using fuzzy logic based voice control device. In: 2022 10th international conference on emerging trends in engineering and technology-signal and information processing (ICETET-SIP-22). IEEE, pp 01–04
23. Raut S, Daigavane P, Shaikh MB. Test bench of automotive component of an electric vehicle for electrical parameter measurement. *Innovations in Electrical and Electronic Engineering*
24. Mane JJ, Aware MV, Khubalkar SW (2020) Multicarrier multi-modulation techniques in multi-level neutral point clamped inverter setup for advanced power electronics and drives lab. *Int J Elect Eng Edu*

Optimal Planning of Hydrogen Refueling Stations Considering Balanced Utilization of Resources



Xuyao Meng, Fushuan Wen, Kun Ding, Changhai Yang, and Yalu Sun

Abstract Compared with electric vehicles, hydrogen fuel cell vehicles offer several advantages, including faster speed and longer range. Additionally, hydrogen fuel cell vehicles are powered by hydrogen energy, which is considered the cleanest energy source available. Therefore, the widespread adoption of hydrogen fuel cell vehicles would contribute to enhancing clean energy usage and reducing dependence on fossil fuels. Optimal planning of hydrogen refueling stations is an important factor restricting the development and promotion of hydrogen fuel cell vehicles, and is also the subject of this work. Specifically, the optimal siting and sizing for hydrogen refueling stations is addressed, with the disparities in resource utilization rates among station taken into account, so as to achieve balanced utilization of refueling resources across all stations. Finally, a sample example is served for demonstrating the feasibility and efficiency of the proposed method.

Keywords Hydrogen fuel cell vehicle · Hydrogen refueling station · Optimal planning · Utilization rate

1 Introduction

The promotion and widespread application of hydrogen fuel cell vehicles (HFCVs) hold great potential in reducing exhaust particle emissions and lowering fossil fuel consumption, thereby contributing to effective reduction of greenhouse gas emissions. Moreover, HFCVs outperform traditional gasoline vehicles in terms of cleanliness, while also surpassing electric vehicles (EVs) in terms of refueling speed and mileage. This inherent flexibility and cleanliness make HFCVs an attractive option for urban transportation, warranting a gradual increase in their market share.

X. Meng (✉) · F. Wen
College of Electrical Engineering, Zhejiang University, Hangzhou 310058, China
e-mail: eemxy@zju.edu.cn

K. Ding · C. Yang · Y. Sun
Economic and Technological Research Institute of State Grid Gansu Electric Power Company,
Lanzhou 730030, China

However, the development of HFCVs faces a major hurdle—the limited availability of hydrogen refueling stations (HRSs). To address this challenge and promote the expansion of HFCVs, it is crucial to develop effective planning methods for HRSs that address the need for increased infrastructure and eliminate the inconvenience of hydrogen refueling [1, 2].

Some publications are available on the optimal planning of HRSs from various perspectives [3]. The determination of an HRS planning scheme relies on aligning it with the hydrogen refueling demand from HFCVs [4]. Additionally, the selection of HRS locations and the configuration of in-station facilities are heavily influenced by the hydrogen supply network [5]. The complexity of HRS planning arises from the consideration of numerous factors, which further complicates the development of an optimal model. Consequently, there has been significant attention dedicated to searching efficient solution algorithms [6, 7].

In different stages of HFCV development, there are varying factors and approaches to consider when planning HRSs [8, 9]. As shown in Fig. 1, the market penetration of HFCVs varies across different development stages. During the market promotion phase of HFCVs, it is crucial to prioritize the convenience of hydrogen refills and the service quality for HFCV users. In the growth phase, with the significant rise in HFCV numbers, focus turns to assessing power grid impact and promoting sustainable HRS construction. In the development plateau phase of HFCVs, greater emphasis is placed on the economic operation and commercial model of HRSs. Currently, as HFCVs are in the market promotion period, the focus lies in obtaining the optimal siting and sizing of HRSs that effectively meet the driving demands of HFCVs within the planning area. This approach aligns with the goal of providing convenient and reliable refueling services for HFCV users, thus fostering the adoption and market penetration of HFCVs.

In this paper, a set of candidate HRS locations is first determined based on engineering practices and municipal requirements. Subsequently, an optimal planning model for HRSs is developed, considering constraints such as service radius, facility utilization rate, land area, and queuing time limitations. The objective functions aim

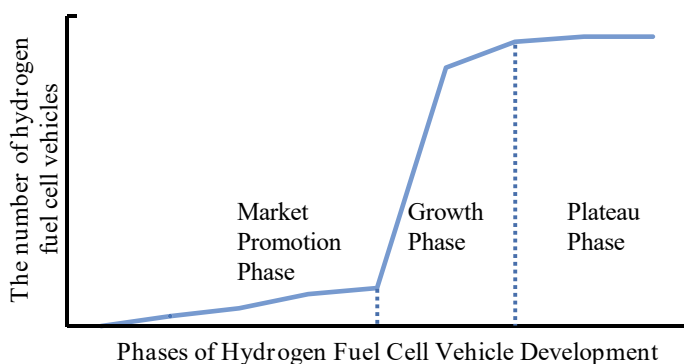


Fig. 1 Phases of hydrogen fuel cell vehicle development

to minimize hydrogen and time costs for HFCV users and optimize the total number of facilities in HRSs. Furthermore, an improved optimal planning model is proposed by incorporating an additional objective function that minimizes resource utilization rate differences among HRSs, which enhances the service balance between HRSs while achieving the optimal location and facility configuration planning.

2 Research Framework

As illustrated in Fig. 2, under the research background that the hydrogen refueling demands of HFCVs and candidate construction sites are determined, the planning problem of HRSs could be described as deciding the suitable number and location of HRSs and the reasonable number of in-station hydrogen refueling facilities.

To ensure the service quality for HFCV users, it is essential to determine the correlation between the number of in-station facilities and users' queue time. This paper applies queuing theory to depict the hydrogen refueling service process for HFCVs, enabling the calculation of users' queuing time and facility utilization rate, which should be restrained to ensure satisfactory service quality at the station. Detailed discussions on this part are presented in Sect. 3.1.

During the promotion phase, the primary factors influencing the HFCV users' selection of HRS are the time and hydrogen consumption involved in finding an HRS. Hence, the first objective function of the optimal planning model for HRSs established in this paper is to minimize the time and hydrogen costs associated with searching for HRSs. Detailed discussions on this part are presented in Sects. 3.2 and 3.3.

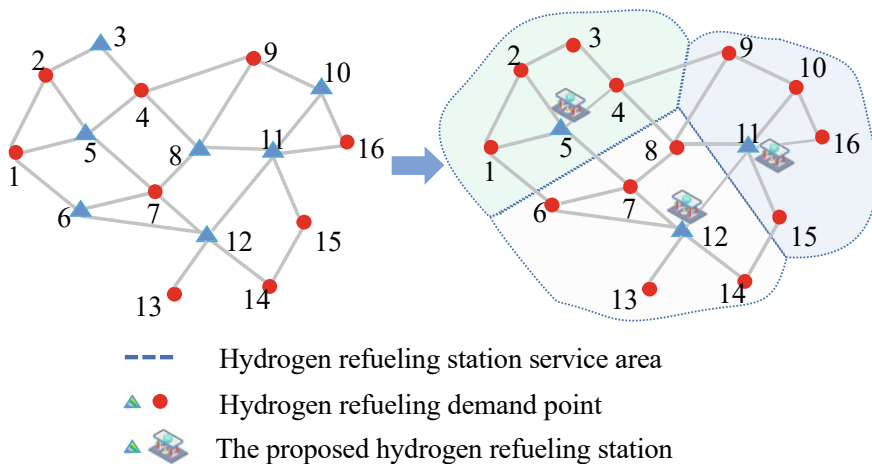


Fig. 2 Diagram illustrating hydrogen refueling station selection and service area division

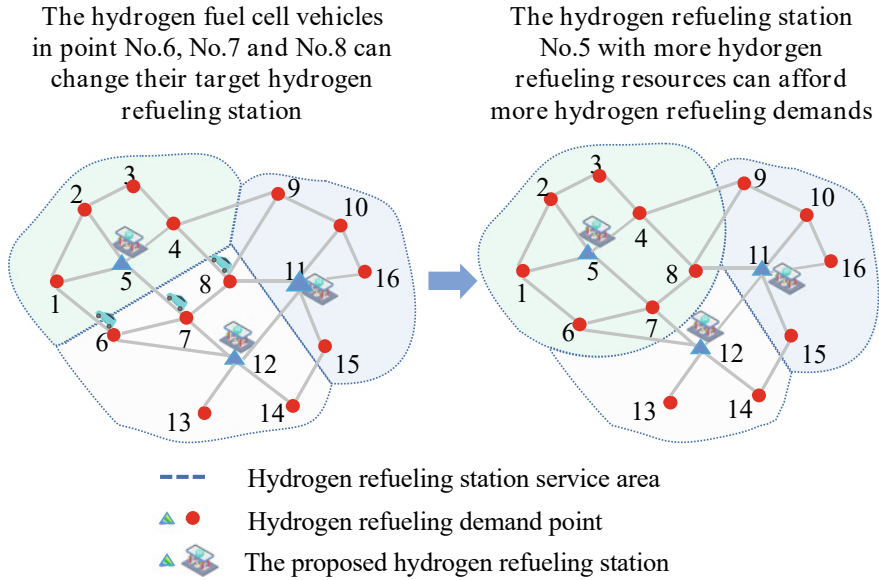


Fig. 3 Diagram illustrating hydrogen refueling station selection and service area division after considering the balance of resource utilization among stations

In some cases, not all candidate sites have sufficient space to accommodate an adequate number of hydrogen refueling facilities to meet the refueling demands of HFCVs arriving via the shortest route. Consequently, a portion of HFCVs should be directed to another station with more refueling facilities. As depicted in Fig. 3, HRS No. 5 possesses greater construction resources than HRS No. 12, such as ample space or a higher hydrogen supply. However, the service area division based solely on distance would result in HRS No. 12 having to handle more hydrogen refueling demands than HRS No. 5. This outcome is evidently unreasonable. To address this issue, an improved optimal planning scheme is proposed, which takes into account the balance of resource utilization among stations. Specifically, the service area of HRS No. 5 is adjusted to cover a greater number of hydrogen refueling demand points. Further details will be discussed in Sect. 3.4.

3 Research Framework

It is assumed that HFCV users within the planning area have access to real-time information regarding the hydrogen refueling status of HRSs and could receive recommended destination HRS notifications on their mobile devices.

3.1 The Queuing Model for the Hydrogen Refueling Station

The queuing and refueling process in the HRS can be described by queuing theory [10]. Each station assumes a random arrival process of HFCVs following a Poisson distribution, while the hydrogen refill time for each HFCV, limited to a short duration, follows a general distribution. Since the hydrogen refueling facilities are numbered, the HFCV queuing and hydrogen refill process at the HRS can be modeled using $M/G/c$ queuing model. Then, the facility utilization rate and average queuing time for HFCVs can be calculated using Eqs. (1) and (2), respectively [11].

$$p = \lambda / (c\mu) \quad (1)$$

$$W_q = \frac{(c\rho)^c \rho}{c!(1-\rho)^2 \lambda} \left[\sum_{i=0}^{c-1} \frac{1}{i!} \left(\frac{\lambda}{\mu}\right)^i + \frac{1}{c!(1-\rho)} \left(\frac{\lambda}{\mu}\right)^c \right]^{-1} \times \frac{1 + (\sqrt{V_T}/E_T)^2}{2} \quad (2)$$

where p represents the average number of HFCVs receiving or waiting for hydrogen refueling service at the HRS, c is the number of hydrogen refueling facilities in the HRS, λ denotes the arrival rate of HFCVs at the HRS per unit time, μ represents the service rate of HFCVs in the HRS per unit time, W_q corresponds to the average queuing time in the $M/G/c$ queuing model, and V_T and E_T are the variance and mean of hydrogen refueling service time, respectively.

The HFCV hydrogen refueling service time consists of both the queuing time and the hydrogen refueling duration. The duration of hydrogen refueling for an HFCV is determined by the remaining hydrogen content in the vehicle before hydrogen prior to refueling, which can be described using β distribution [12], as illustrated in Eq. (3).

$$\varphi \sim \text{Beta}(2.67, 8) \quad (3)$$

Additionally, it is important to mention that the hydrogen refueling facility requires a certain amount of preparation time. Specifically, there must be a minimum time interval of 3 min between two consecutive hydrogen refueling services. This ensures proper functioning and efficiency of the refueling process.

3.2 Constraints in the Optimal Planning Model for Hydrogen Refueling Stations

The optimal planning of hydrogen refueling stations involves various constraints, encompassing transportation limitations of HFCVs, availability of land space,

hydrogen supply constraints at potential sites, and the requirement to uphold the quality of hydrogen refueling services.

First and foremost, it is essential to limit the service radius of the HRS to ensure that HFCVs can reach the station before their fuel cell power is depleted. This constraint can be described using the following Eq. (4).

$$a_{jk} D_{jk} \leq \frac{M \times [\text{SOC}_{\text{alarm}} - E(\text{SOC}_{\text{min}})]}{\varepsilon H_E} \quad (4)$$

where a_{jk} is the variable that indicates whether the HFCVs at the j th ($j = 1, 2, \dots, m$) hydrogen refueling demand point will proceed to the k th ($k = 1, 2, \dots, n$) HRS for hydrogen refueling, D_{jk} represents the distance between the j th hydrogen refueling demand point and the k th HRS (km), M denotes the driving mileage of HFCVs, $\text{SOC}_{\text{alarm}}$ represents the minimum threshold of the state of charge (SOC) alarm for HFCVs, SOC_{min} indicates the lowest SOC of an HFCV when it arrives at HRS, and its expected value is denoted as $E(\text{SOC}_{\text{min}})$, ε is the road curve coefficient in the planning area, and H_E is the fuel cell capacity of HFCVs.

When considering the land space and hydrogen supply constraints in HRSs, it is crucial to consider the restrictions on the number of facilities at each potential site and the daily hydrogen supply capacity at each HRS. These constraints can be described using Eqs. (5) and (6).

$$c_k \leq c_{tk} \quad (5)$$

$$H_{Nk} = \sum_{i=1}^{c_k} H_{ik} \leq H_{tk} \quad (6)$$

where c_k and c_{tk} represent the planned number and the upper limit of hydrogen refueling facilities at the k th HRS, respectively. H_{Nk} and H_{tk} indicate the hydrogen usage amount (kg) and the upper limit of hydrogen usage amount (kg) at the k th HRS, respectively. m_k denotes the number of HFCVs that proceed to the k th HRS for hydrogen refueling, and H_{ik} represents the hydrogen usage amount of the i th facility at the k th HRS (kg).

To uphold service quality at stations, it is essential to establish constraints on the queuing time experienced by HFCV users and the facility utilization rate. These constraints can be represented by Eqs. (7) and (8).

$$W_{qk} \leq W_t \quad (7)$$

$$p_t \leq p_k \leq 1 \quad (8)$$

where W_{qki} represents the average queuing time in the k th HRS, W_t denotes the user-tolerable queuing time. p_k represents the facility utilization rate in the k th HRS,

while p_t denotes the lower limit of facility utilization rate. The values of W_{qk} and p_k can be calculated using Eqs. (2) and (1), respectively.

3.3 Objective Functions in the Optimal Planning Model for Hydrogen Refueling Stations

To minimize the time and hydrogen costs of HFCV users during the search of HRSs, as well as the total number of facilities, the objection functions of the optimal planning model can be represented by Eqs. (9) and (10).

$$\min L = \min \left\{ \left[\sum_{j=1}^m (\varepsilon h_j D_{\min j}) \right] \times \left(\frac{C_t}{V} + \frac{sH_E}{M} \right) \right\} \quad (9)$$

$$\min C_{\text{num}} = \min \sum_{k=1}^n c_k \quad (10)$$

where L represents the travel-to-station cost for HFCV users, which includes the time and hydrogen cost incurred while traveling to the HRS for refueling, ε denotes the road curve coefficient within the planning area, h_j indicates the number of HFCVs at the j th hydrogen refueling demand point, $D_{\min j}$ represents the distance between the j th hydrogen refueling demand point and its nearest construction site (km), C_t corresponds to the average travel time cost per hour for HFCV users, which can be calculated based on their monthly income and monthly working time. V denotes the average driving speed of HFCVs (km/h), s represents the hydrogen refueling price per kilogram, H_E represents the fuel cell capacity of HFCVs, and M denotes the driving mileage of HFCVs (km). The term C_{num} in Eq. (10) represents the total number of hydrogen refueling facilities in all HRSs, while c_k indicates the planning number of facilities at the k th HRS.

In summary, Eqs. (4)–(10) form the HRS planning optimization model for HRSs, as depicted in Eq. (11).

$$\left\{ \begin{array}{l} \min L = \min \left\{ \left[\sum_{j=1}^m \varepsilon h_j D_{\min j} \right] \times \left(\frac{C_t}{V} + \frac{sH_E}{M} \right) \right\} \\ \min C_{\text{num}} = \min \sum_{k=1}^n c_k \\ \text{s.t. } a_{jk} D_{jk} \leq \frac{M \times [\text{SOC}_{\text{alarm}} - E(\text{SOC}_{\min})]}{\varepsilon H_E}, \\ c_k \leq c_{tk}, H_{Nk} \leq H_{tk}, W_{qk} \leq W_t, p_t \leq p_k \leq 1, j = 1, 2 \dots m, k = 1, 2 \dots n \end{array} \right. \quad (11)$$

3.4 The Improved Optimal Planning Model for Hydrogen Refueling Stations with Considering the Balance of Resource Utilization Among Stations

With the optimal planning model of HRSs discussed in Sect. 3.3, the planning area would be divided into several service areas, but each HRS has the same service radius. However, this approach may result in an inequitable distribution of hydrogen refueling demands among HRSs. To address this issue, the $D_{\min j}$ in Eq. (9) should be modified to allow HFCVs at each demand point to choose any HRS within their reachable range. Additionally, in order to enhance the balance of resource utilization among HRSs, an additional objective function is proposed to minimize the variance of resource utilization rate across stations, as indicated in Eq. (12).

$$\min R_T = \min \left[\sum_{k=1}^{n'} \left(\frac{c_k}{c_{tk}} - \frac{1}{n'} \sum_{k=1}^{n'} \frac{c_k}{c_{tk}} \right)^2 \right] \quad (12)$$

where R_T represents the variance of the resource utilization rate in HRSs, n' indicates the planning construction number of HRSs, and c_k and c_{tk} represent the planning number and the upper limit of facilities in the k th HRS, respectively.

Then, the optimal planning model for HRSs in Eq. (11) can be improved as:

$$\left\{ \begin{array}{l} \min L = \min \left\{ \left[\sum_{j=1}^m (\varepsilon h_j D_{opt j}) \right] \times \left(\frac{C_t}{V} + \frac{s H_E}{M} \right) \right\} \\ \min C_{\text{num}} = \min \sum_{k=1}^n c_k \\ \min R_T = \min \left[\sum_{k=1}^{n'} \left(\frac{c_k}{c_{tk}} - \frac{1}{n'} \sum_{k=1}^{n'} \frac{c_k}{c_{tk}} \right) \right] \\ \text{s.t. } D_{\text{opt}} = \{ D_{jk} | j = 1, 2, \dots, m, k = 1, 2, \dots, n \} \\ a_{jk} D_{jk} \leq \frac{M \times [\text{SOC}_{\text{alarm}} - E(\text{SOC}_{\text{min}})]}{\varepsilon H_E} \\ c_k \leq c_{tk}, H_{Nk} \leq H_{tk}, W_{qk} \leq W_t, P_t \leq P_k \leq 1, j = 1, 2 \dots, m, k = 1, 2 \dots, n \end{array} \right. \quad (13)$$

where $D_{\min j}$ in Eq. (9) is replaced by $D_{opt j}$, which can be selected from D_{jk} , and D_{jk} represents the distance between the j th hydrogen refueling demand point and the k th HRS (km). This means that HFCVs at each demand point have the flexibility to choose any HRS within their range.

The improved optimal planning model for HRSs illustrated in Eq. (13) aims to make the location of HRS and the hydrogen refueling facility configuration more rational and minimize the disparities in resource utilization rates among HRSs.

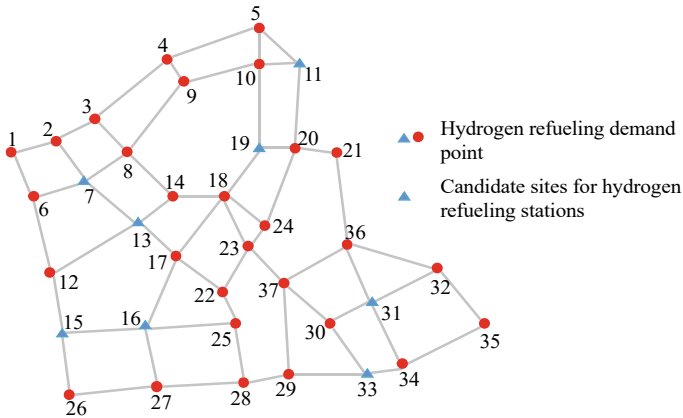


Fig. 4 Distribution of the hydrogen refueling demand points and candidate sites for hydrogen refueling stations

4 Case Study

In this section, an area where the candidate sites and hydrogen refueling demand points have been identified is chosen for planning using both the optimal planning model and improved optimal planning model for HRSs. Subsequently, the results obtained from these two models are compared and analyzed.

4.1 Parameters Setting

The planning area consists of 37 junctions, each of which serves as a hydrogen refueling demand point. The location and constraints of 11 candidate sites are presented in Fig. 4 and Table 1. The average travel time cost per hour for HFCV users is \$10, and the hydrogen refueling price per kilogram is \$1. The HFCV has a fuel cell capacity of 5 kg, enabling a driving mileage of 500 km and the average driving speed of HFCVs in the planning area is 40 km/h. The in-station facility provides a hydrogen refueling speed of 20 kg/h. The lower limitation for facility utilization rate is set at 0.3, and the upper limitation for queuing time at the station is 10 min.

4.2 Results of Optimal Planning

Using the optimal planning model for HRSs presented in Eq. (11), the optimal HRS planning can be obtained, as depicted in Fig. 5 (Scheme I). However, a different optimal HRS planning is achieved when employing the improved optimal planning

Table 1 Location and construction constraints of candidate sites for hydrogen refueling stations

Location no.	Number limitation of facilities	Hydrogen supply limitation per day (kg)
3	5	250
7	4	500
11	6	250
13	4	200
15	5	400
16	6	250
19	4	200
27	4	400
31	4	250
33	4	250
34	4	400

model for HRSs illustrated in Eq. (13), as shown in Fig. 6 (Scheme II). A comparison of the two planning schemes is presented in Table 2. Additionally, the comparison of in-station service quality can be observed in Fig. 7.

Comparing Figs. 5 and 6 reveals that demand points 13, 14, 17, 18, 22, 23, and 24 have changed their target HRSs, and in the improved planning scheme, the station No. 13 has been removed. As indicated in Table 2, the optimal HRS planning scheme obtained according to the optimal planning model for HRSs requires the installation of 11 hydrogen refueling facilities in 8 HRSs. On the other hand, the HRS planning scheme derived from the improved optimal planning model for HRSs only necessitates 10 facilities in 7 HRSs. In other words, the improved model can satisfy

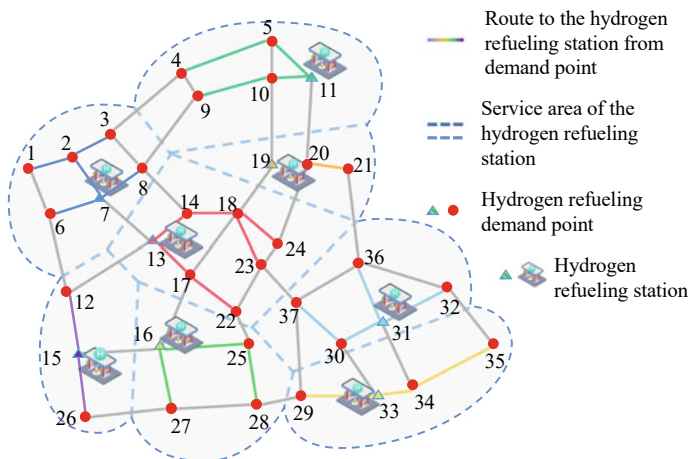


Fig. 5 Diagram illustrating hydrogen refueling station selection and service area division in Scheme I

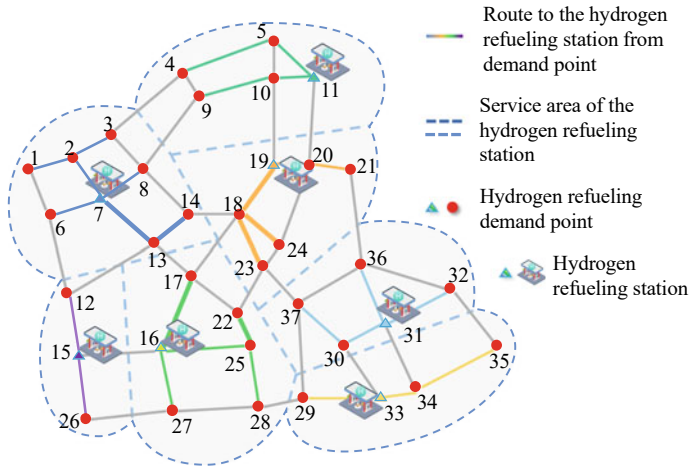


Fig. 6 Diagram illustrating hydrogen refueling station selection and service area division in Scheme II with considering the balance of resource utilization among stations

Table 2 Optimal planning schemes for hydrogen refueling stations with and without considering the balance of resource utilization among stations

The optimal planning scheme	Without considering the balance of resource utilization among stations (Scheme I)		With considering the balance of resource utilization among stations (Scheme II)	
	The hydrogen refueling station No.	The number of in-station facilities	The hydrogen refueling station No.	The number of in-station facilities
/	7	2	7	2
	11	2	11	2
	13	2	×	×
	15	1	15	1
	16	1	16	2
	19	1	19	1
	31	1	31	1
	33	1	33	1
Total number	8	11	7	10

the same hydrogen refueling demands with fewer HRSs and facilities. Furthermore, Fig. 7 demonstrates that the improved model contributes to the balanced utilization of in-station hydrogen refueling resources.

The results demonstrate that the improved model can also reduce the HRS construction cost and promote the facility utilization rate. Notably, HRS No. 19 experienced a significant increase in its facility utilization rate from 0.30 to 0.56, making

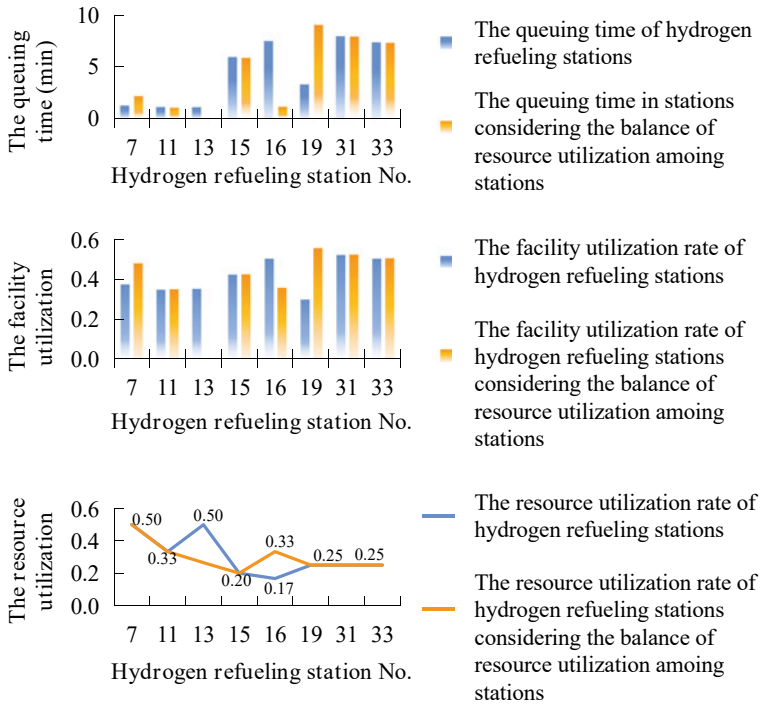


Fig. 7 Service quality comparison with and without considering the balance of resource utilization among stations

it more profitable. This achievement holds great importance for the sustainable development of HRSs and HFCVs.

However, it can also reveal that queuing time is inversely proportional to the facility utilization rate. Although the improved scheme exhibits higher facility utilization rates, it is accompanied by longer queuing times due to a decrease in the total number of facilities. This indicates that the number of facilities plays a crucial role in the balance between queuing time and facility utilization rate. Further exploration of this relationship will be conducted in future research.

5 Conclusions and Following Works

This paper introduces an improved optimal planning model for HRSs with the objective of reducing the resource utilization rate gap among different construction sites. In addition to minimizing the total number of hydrogen refueling facilities and the travel-to-station costs for hydrogen fuel cell vehicle users, an additional objective function is incorporated to minimize the variance of the resource utilization rate

in stations. By adjusting the service area of HRSs, the improved planning scheme requires fewer stations and facilities, resulting in smaller differences in resource utilization rate across stations. However, there are several factors that still need to be considered in future work to further improve the optimization method for HRS planning. These factors include the classification of HFCVs, the solution algorithm for the optimization model, and the impact on power grid. Attention will be given to these areas for further investigation and improvement.

Acknowledgements This work is financially supported by a Science and Technology Project of the State Grid Corporation of China entitled “Research on the Coordinated Operation and Market Mechanism of Integrated Source-Network-Load-Storage Projects in Power Systems” (Project No.: 1400-202333325A-1-1-ZN; Contract No.: SGTYHT/21-JS-223).

References

1. Apostolou D, Xydis G (2019) A literature review on hydrogen refueling stations and infrastructure. Current status and future prospects. *Renew Sustain Energy Rev* 113:109292
2. Isaac N, Saha AK (2023) A review of the optimization strategies and methods used to locate hydrogen fuel refueling stations. *Energies* 16(5):2171
3. Sun HR, He CM, Yu XZ, Wu MX, Ling YH (2019) Optimal siting and sizing of hydrogen refueling stations considering distributed hydrogen production and cost reduction for regional consumers. *Int J Energy Res* 43(9):4184–4200
4. Fuse M, Noguchi H, Seya H (2021) Near-term location planning of hydrogen refueling stations in Yokohama City. *Int J Hydrogen Energy* 46(23):12272–12279
5. Bae S, Lee E, Han J (2020) Multi-period planning of hydrogen supply network for refuelling hydrogen fuel cell vehicles in urban areas. *Sustainability* 12(10):4114
6. Lin R, Ye Z, Guo Z, Wu BD (2020) Hydrogen station location optimization based on multiple data sources. *Int J Hydrogen Energy* 45(17):10270–10279
7. Li L, Chami ZA, Manier H, Manier MA, Xue J (2021) Incorporating fuel delivery in network design for hydrogen fueling stations: formulation and two metaheuristic approaches. *Transport Res Part E Logistics Transport Rev* 152:102384
8. Samuelsen S, Shaffer B, Grigg J, Lane B, Reed J (2020) Performance of a hydrogen refueling station in the early years of commercial fuel cell vehicle deployment. *Int J Hydrogen Energy* 45(56):31341–31352
9. Kurtz J, Bradley T, Winkler E, Gearhart C (2020) Predicting demand for hydrogen station fueling. *Int J Hydrogen Energy* 45(56):32298–32310
10. Brown T, Kisting H (2022) Analysis of customer queuing at hydrogen stations. *Int J Hydrogen Energy* 47(39):17107–17120
11. Choi BD, Shin YW (1991) Transient diffusion-approximation for M/G/m system. *Oper Res* 34(3):306–328
12. Zhao L (2015) Optimization of compressed hydrogen supply system in hydrogen refueling station using thermodynamics and queuing theory. Ph.D. dissertation, College of Chemical and Biological Engineering, University Zhejiang

A Communication Technique to Improve the Performance of On Load Tap Changer of Transformer with Renewable Load



Roshni M. Virkhare, Prema Daigavane, and Hari Kumar Naidu

Abstract The regulating power transformers are utilized in industrial applications and electrical energy networks in on load tap changer. The technological advancements of mixed-type of load, including wind power on load tap changer (OLTC) are addressed in this work. The discussion and presentation of OLTC applications include the general switching concepts for OLTCs. However, on load tap changers of the mechanical type were used in the past, they had a number of drawbacks and restrictions, including arcing, expensive maintenance, poor response times, and other issues. Electronic or solid-state tap changers were adopted to solve these issues. A quick running OLTC regulator with solid-state component are added up as tap changer for controllability which consists of insulated gate bipolar transistor and thyristor to reduce arcing instead of mechanical switches. This work describes the development of a quick OLTC solid-state regulator, with results in MATLAB demonstrating the enhanced on load tap changing performance, which will be useful to the industrial experts and R&D professionals, since the drawback of under load tap changer of cumulative effect of change in load, voltage and the impedance effect is addressed by this novel technique to safeguard against voltage stability of the power system. The performance evaluation can be done by digital signal processing communication from remote.

Keywords On load tap changer · Transmission line · Resistive load · Wind generator · MATLAB

R. M. Virkhare (✉) · P. Daigavane · H. Kumar Naidu
G. H. Raisoni College of Engineering, Nagpur, India
e-mail: roshni.virkhare.mtechps@ghrce.raisoni.net

P. Daigavane
e-mail: prema.daigavane@raisoni.net

H. Kumar Naidu
e-mail: harikumar.naidu@raisoni.net

1 Introduction

On load tap changers (OLTC) of power transformers are a crucial component for any contemporary power system. For almost 90 years, the primary elements of electrical networks and industrial applications have been power transformers with on load tap changers (OLTCs). By continuously changing the transformer ratio under load, OLTC regulates the voltage and shift phase. The two switching principles employed for load transfer operations are resistor-type OLTC and reactor-type OLTC since the beginning of tap changer development. The mechanical OLTC regulates the voltage dips but creates contact arcs in the diverter switches while changing the taps, which creates losses during tap changing and noise disturbances. This could be arrested by the use of solid state on load tap changing method.

2 Overview of the System

The functional block diagram of OLTC transformer is shown in Fig. 1.

A transformer is an electrical tool that works by using electromagnetic induction, which means electricity is created when a conductor moves through a magnetic field. A transformer has two parts called primary and secondary windings. This is mostly about moving electricity from one place to another. It's important to make sure the right amount of electricity goes into a transformer. Hence, usage of tapping is to make sure the voltage is steady based on how much the transformer can handle. Therefore taps are attached to different parts of the transformers coils, which changes the number of turns in the coils. There are ways to make this thing automatic by OLTC. Figure 2 shows the internal details of three-phase on load tap changer.

Figure 3 shows the on load tap changer mechanism with diverter switches.

Figure 4 shows tap position in on load tap changer.

Figure 5 shows the next position after tap change in on load tap changer.

3 Three-Phase OLTC Regulating Transformer Yg/(D1)-17 Positions

The three-phase tap-changing transformer (two windings) is used to represent a transformer with two coils that can be adjusted to change the voltage. The first step that changes the tap on the first part is that coil 1 is joined in the shape of a Y. Further, it connects winding 2 in two different ways either wye or delta.

The way the tap is set and controlled decides how much the transformer can increase or decrease its voltage level. When set the mode to tap control, the input signal controls where the tap is. When the voltage regulation mode is turned on, an

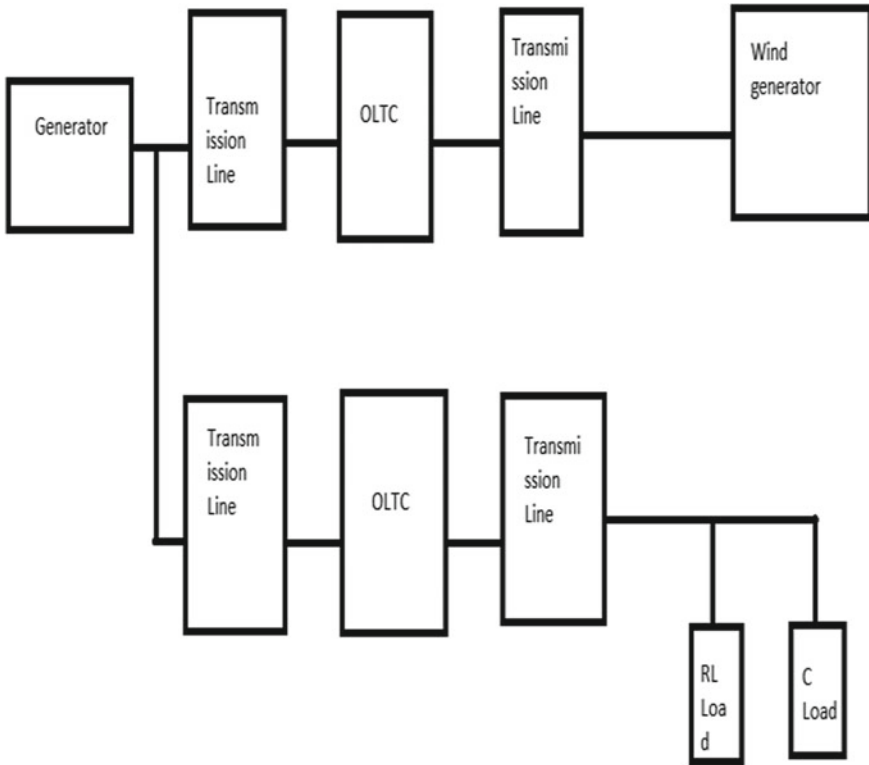


Fig. 1 Functional block diagram of OLTC transformer

internal device compares the input signal with a reference voltage and controls the position of the tap as shown in Fig. 6.

3.1 Three-Phase Mutual Inductance

In three phase, mutual inductance block helps in the setup of three-phase power systems more easily. Instead of using complicated terms, it uses simpler ones called positive and zero sequence resistances and inductances.



Fig. 2 On load tap changer

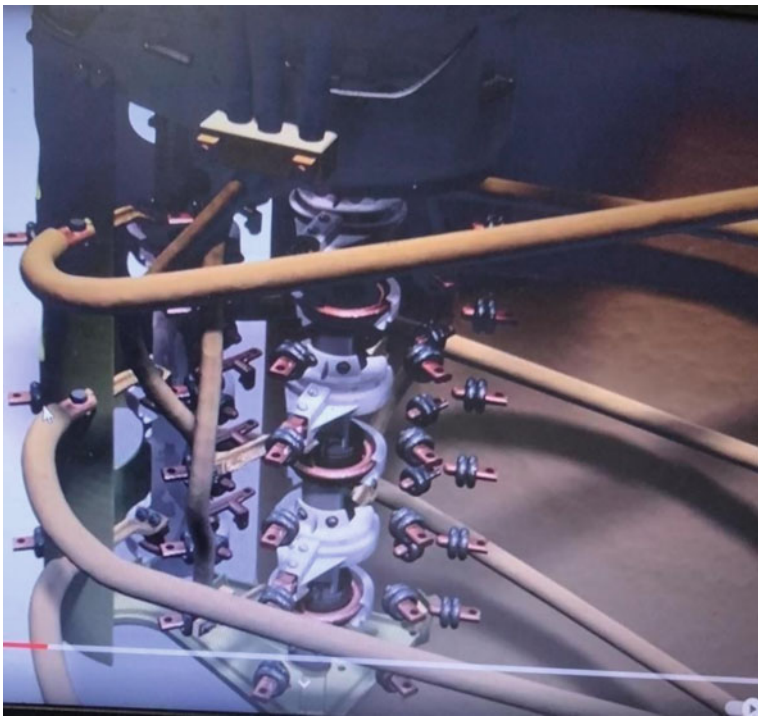


Fig. 3 On load tap changer diverter

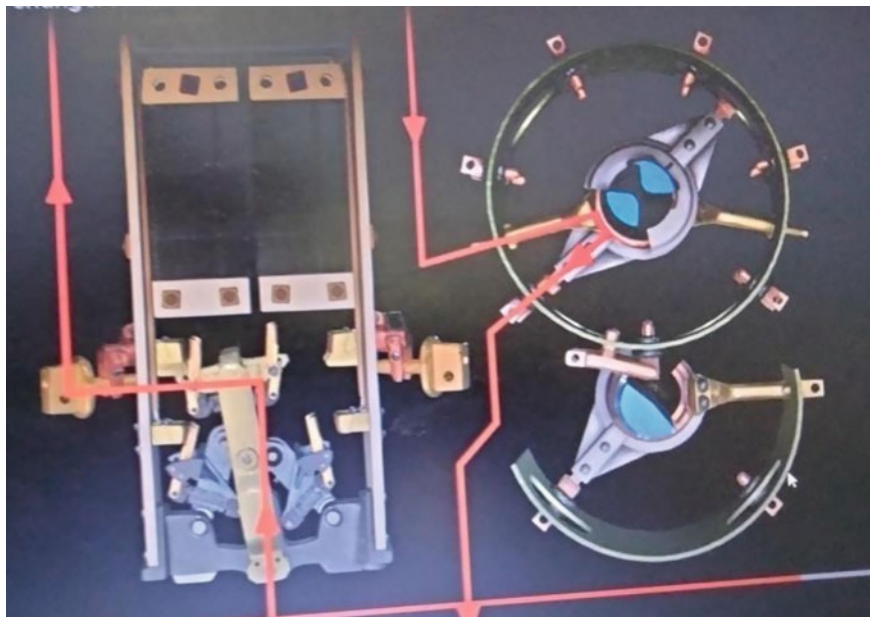


Fig. 4 Tap position in on load tap changer

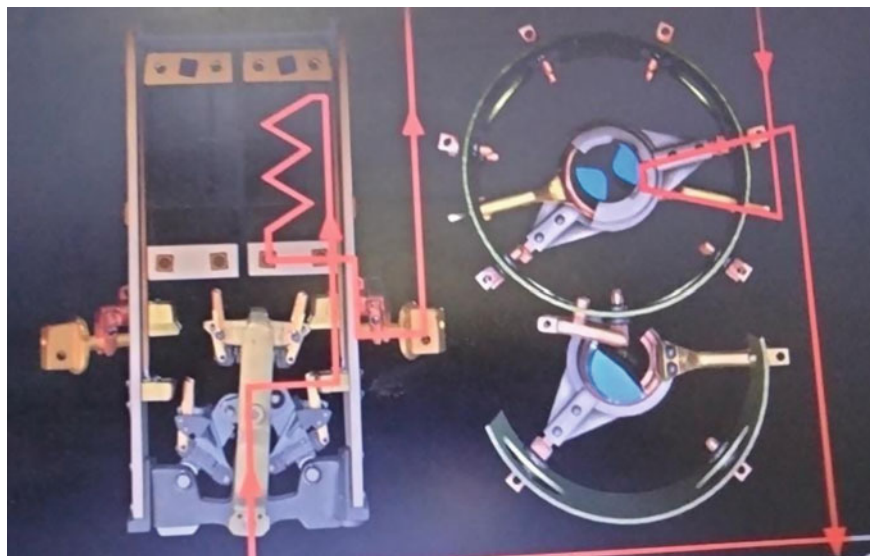


Fig. 5 Tap changing position in on load tap changer

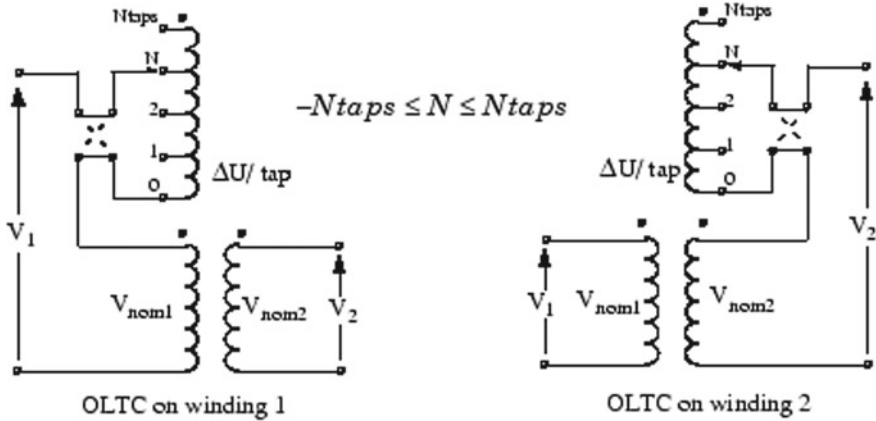


Fig. 6 On load tap changer

4 Wind Turbine and the Doubly-Fed Induction Generator System

The process of changing electricity from AC to DC and back to AC is split into two parts. The rotor and grid are converters that use power electronics to change a DC voltage into an AC voltage. A capacitor connected to the DC side works like a battery for DC power. A special tool called an inductor L is used to join grid to the grid. The rotor is connected to the grid with slip rings and brushes, while the stator is directly connected. The wind turbine turns the power from the wind into electricity, which goes into a generator. The generator then sends the electricity to the power grid. The control system monitors the convertor rotor and grid how to move by sending signals called pitch angle and voltage commands as shown in Fig. 7.

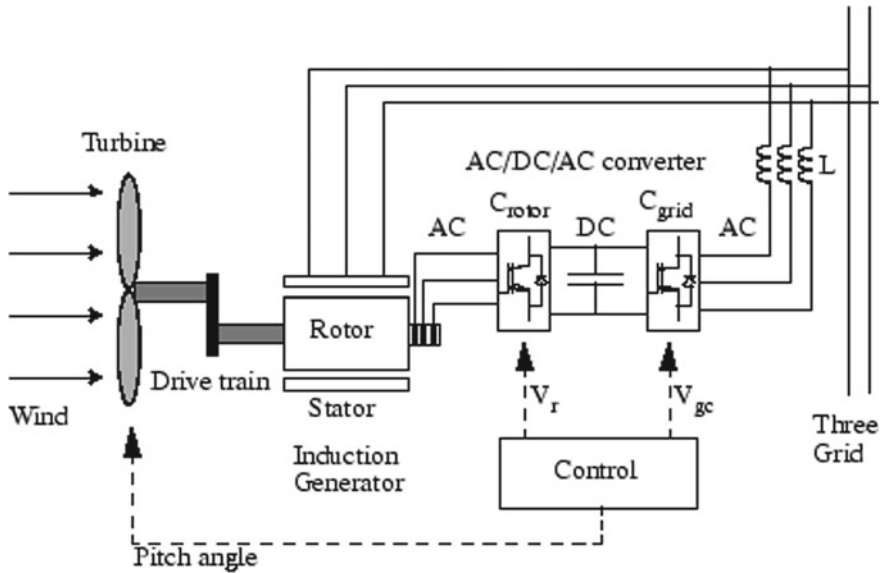


Fig. 7 Block diagram of wind turbine

5 The MATLAB On Load Tap Changer Simulation Model is Shown in Fig. 8

See Fig. 8.

6 Smooth Variation of Voltage

Phasor simulation of on load tap changer (OLTC) regulating transformer gives smooth variation of transformer taps 1 and 2, voltage in per unit B1, B2 and B4, reactive power Mvar B1, B3 and real power MW B1, B3. The results obtained by this new technique which are shown in Fig. 9 address the variation of three-phase smooth voltage by introducing positive and zero sequence resistances and inductances.

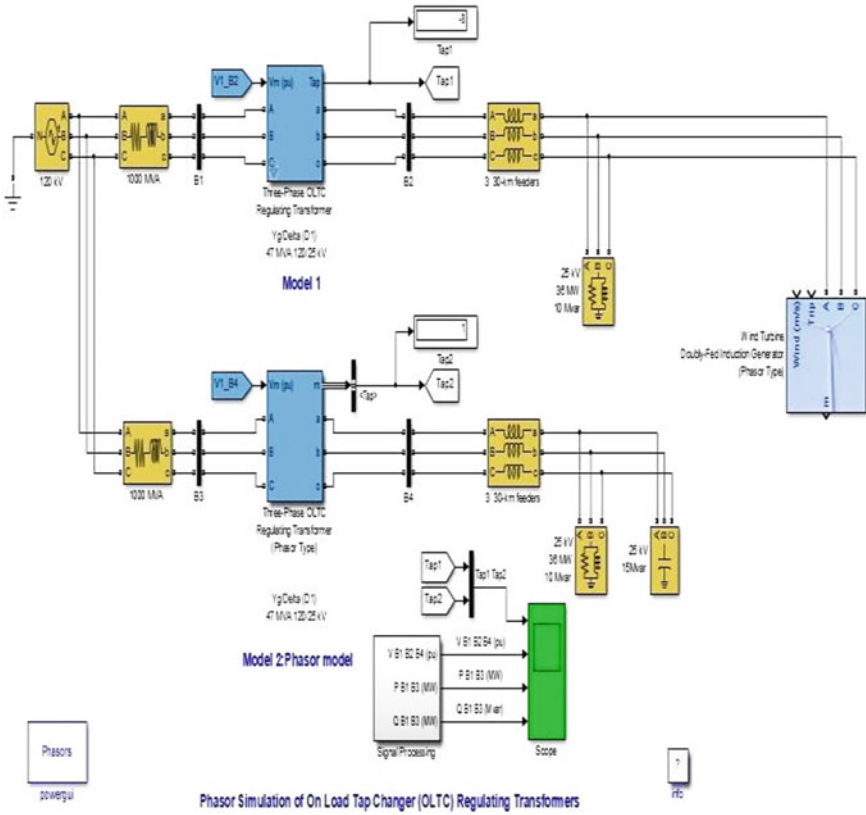


Fig. 8 On load tap changer simulation model

7 Analog to Digital Signal with PCM

On line performance monitoring from remote through digital communication system of on load tap changer process is as shown in Figs. 2 and 3.

The steps in conversion of analog to digital signal goes through filtration to limit bandwidth and thereafter PCM encoder consisting of sampling, quantizing and encoding. The elements of PCM system are transmitter, transmission path and receiver as shown in Fig. 10.

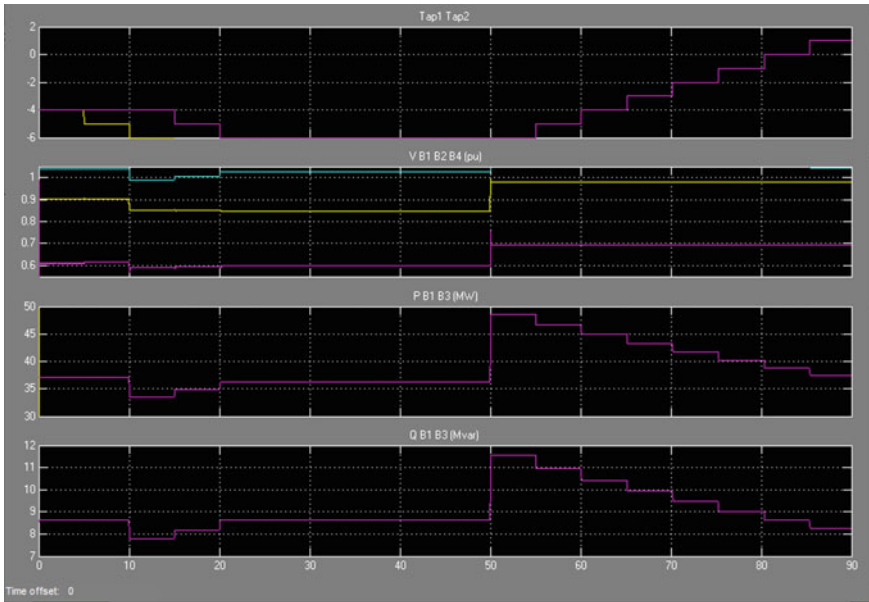


Fig. 9 Smooth variation of voltage

8 Results

In a radial transmission line connected through the transformer and if the load is being increased gradually. The action of the under load tap changer (ULTC) is to maintain the voltage at load terminal by changing the transformation ratio $N1/N2$, when applied voltage is same; however, when the load impedance is reflected on primary side of the transformer the input impedance gets reduced and it further reduces the voltage, and in turn, ULTC tries to increase the voltage which is cumulative phenomena results into voltage stability situation.

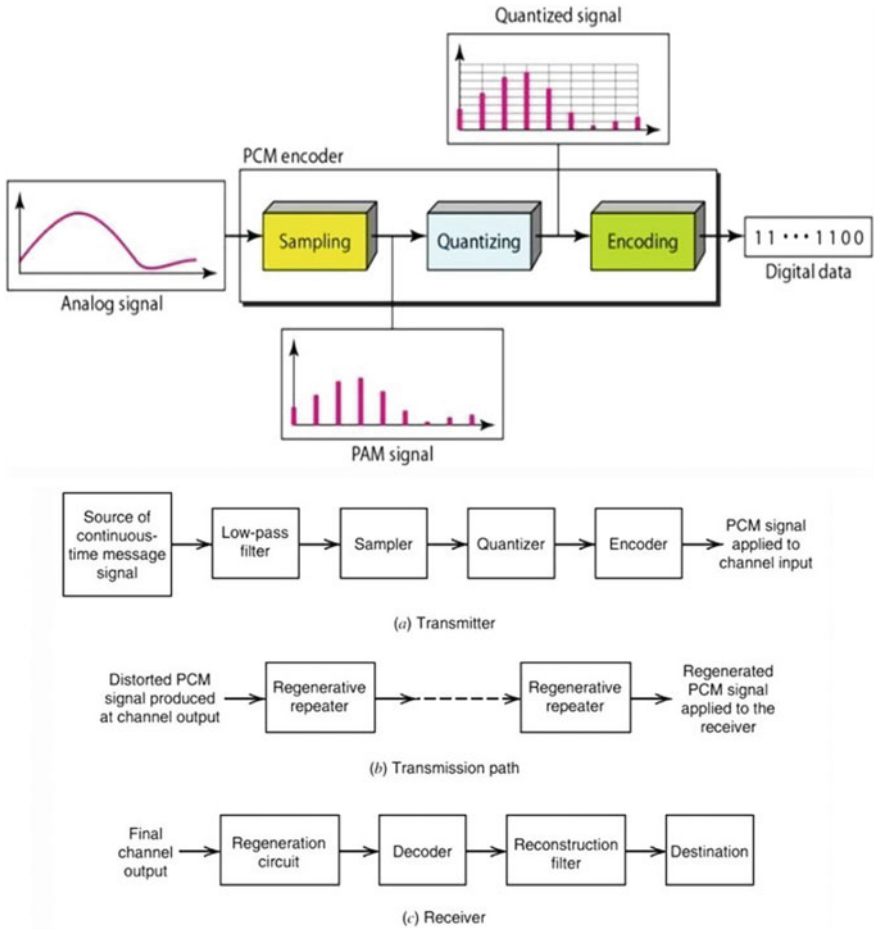


Fig. 10 Analog to digital signal with PCM

9 Conclusion

The performance result indicates that the on load tap changer operates satisfactorily during the load including wind energy, without causing sparking in the diverter chamber and thus protecting the transformer from fire hazards. Also the results indicate the input impedance effect reduction due to change of load has little effect on the performance of on load tap changer due to the introduction of positive and zero sequence resistances and inductances. The performance evaluation can be done from remote by digital communication technique.

References

1. Faiz J, Siahkolah B (2003) New solid state on load tap changers technology for distribution transformer. *IEEE Trans Power Deliv* 18(1):136–141
2. Mailah NF, Bashi SM, Meng WH (2003) Microcontroller based semiconductor tap changer for power transformer. In: *IEEE Bologna Power Tech Conference*. Bologna, Italy, pp 1–6
3. Noorbakhshi M, Afzalian A (2007) Design and PLC based implementation of supervisory control for under-load tap changing transformer. In: *International conference on control, automation and system*, Seoul, Korea, pp 901–906
4. Sarkar M, Subudhi B, Daigavane P (2019) Robust PI damping controller design using wide-area signal for inter-area oscillation in power system. In: *2019 8th international conference on power systems (ICPS)*. IEEE, pp 1–6
5. Faiz J, Siahkolah B (2007) New controller for an electronic tap changer-Part II: measurement algorithm and test results. *IEEE Power Deliv* 22(1)
6. Echavarria R, Claudio A, Cotorogea M (2007) Analysis, design, and implementation of a fast on load tap changing regulators. *IEEE Power Electron* 22(2)
7. Shanqi H (2016) Research and application of contactless load automatic voltage regulating distribution transformer. *China Electr Power (TechnologyEdition)* 1:53–55
8. Chaudhari M, Babu K, Khubalkar SW, Daigavane P (2019) Off-grid hybrid solar power conditioning unit for critical and non-critical loads. In: *2019 international conference on intelligent computing and control systems (ICCS)*. IEEE
9. Faiz J, Kolah B (2006) Differences between conventional and electronic tap-changers and modifications of controller. *IEEE Trans Power Deliv*
10. Mailah NF, Bashi SM, Meng WH (2003) Microcontroller based semiconductor tap changer for power transformer. In: *Paper accepted for presentation at 2003 IEEE Bologna power tech conference*, Bologna, Italy
11. Monroy-Berjillos D, Gomez-Exposito A (2007) Improving the voltage regulation of secondary feeder by applying solid state tap-changer to MV/LV transformers. In: *9th international conference. Electrical power quality and utilization*. Barcelona
12. Ding T, Bo R, Bie Z, Wang X (2017) Optimal selection of phase shifting transformer adjustment in optimal power flow. *IEEE Transaction* 2017
13. Verboomen J, Hertzen DV, Kling WL, Belmans R (2005) Phase shifting transformers: principles and applications. In: *International conference on future power system*
14. Tirupathi R et al (2022) Application of phase shifting transformer in Indian power system. *Int J Comput Electr Eng* 4(2) (April 2012, computing, communication and networking 23–25 Dec 2022)
15. Shaikh RAJ, Naidu H, Kokate PA (2021) Evolutionary computing and mobile sustainable networks: proceedings of ICECMSN 2020. *Evolutionary computing and mobile sustainable networks*. Springer, Singapore
16. Tembhare M, Naidu H, Kokate P (2020) A review study on the multiple and useful application of fiber optic illumination system. *IEEE*
17. Katole DN, Daigavane MB, Gawande SP, Daigavane PM (2018) Improved single phase instantaneous pq theory for DVR compensating nonlinear load. In: *2018 IEEE international conference on power electronics, drives and energy systems (PEDES)*. IEEE, pp 1–6
18. Ramtekkar P, Naidu H, Dudhe S (2021) A novel wireless fire containment and extinguishing system to save life and destruction of property *IEEE*. In: *2021 international conference on computational intelligence and computing applications (ICCICA)*
19. Sheikh SA, Naidu HK (2021) Train accident prevention using fuzzy logic based MEMS system, *IEEE*. In: *2021 5th international conference on electrical, electronics, communication, computer technologies and optimization techniques (ICEECCOT)*
20. Lanjewar A, Khubalkar SW, Junghare AS (2018) Comparative analysis of two loop integer and fractional order PID controller for inverted pendulum. *ICSEDPS 2018, GHRCE, Nagpur*

21. Sao S, Dudhe S, Naidu H (2022) A novel IoT based sample collection method for baby meconium aspiration syndrome and analyzing suction pressure of the system, IEEE. In: 2022 10th international conference on emerging trends in engineering and technology—signal and information processing (ICETET-SIP-22)
22. Jagyasi DM, Naidu HK (2022) A review of MVDC circuit breaker sub-station supply for electrical traction vehicle. In: 2022 10th international conference on emerging trends in engineering and technology—signal and information processing (ICETET-SIP-22). IEEE
23. Kapse SSS, Daigavane MB, Daigavane PM (2020) Optimal localization and sizing of UPFC to solve the reactive power dispatch problem under unbalanced conditions. IETE J Res 66(3):396–413
24. Wath AK, Daigavane PM, Naidu HK (2022) A technique of image and fuzzy logic to solve traffic congestion at signal. In: IEEE 2023. Published in 2022 international conference on smart generation computing, communication and networking 23–25 Dec 2022

Performance Analysis of Wind Power Forecasting via System Advisor Model Software



J. Sathyaraj  and V. Sankardoss 

Abstract This paper discusses the analyses of rapid power consumption of the most popular renewable energy sources: solar energy, wind energy, hydro energy, tidal energy, geothermal energy, and biomass energy. The renewable power generation capacity has significantly increased in recent years. One of the significant renewable energy sources is wind power. The world's and India's total installed wind power capacity was 840.9 and 41.93 GW up to 2022. Wind power plays an important role in converting kinetic energy into electrical energy. The wind analyses of three different lands of each location were selected in the United States of America (USA) like Northwestern Indiana for flat land, Northwestern Montana for canyon lands, and Eastern North Carolina for offshore. The National Renewable Energy Laboratory (NREL) has been developed by system advisor model (SAM) software. The SAM software is used to perform the wind analysis of various locations.

Keywords Renewable energy sources · System advisor model · Wind forecasting · Wind power

1 Introduction

Energy is in the form of heat energy, light energy, chemical energy, electrical energy, nuclear energy, etc. Electrical energy has a more important role in day-to-day life. It is the primary input of economic activity, because of socioeconomic and rapid population. Conventional energy sources are exhausted due to the demand for fossil fuels. For that reason to find alternative energy sources for non-conventional energy sources such as wind energy, solar energy, biomass energy, etc. [1]. The developed and developing countries are concentrated on wind energy and have clean, and safe

J. Sathyaraj (✉) · V. Sankardoss

School of Electrical Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India
e-mail: sathyaraj.j2022@vitstudent.ac.in

V. Sankardoss

e-mail: vsankardoss@vit.ac.in

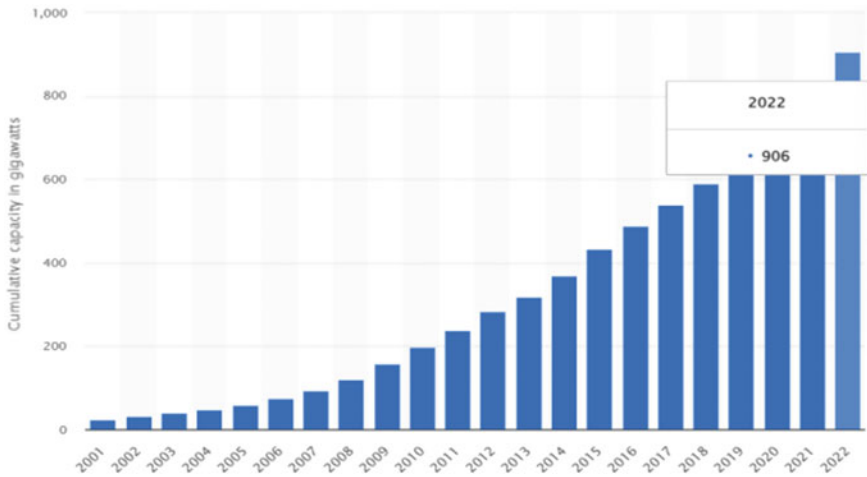


Fig. 1 Worldwide wind power capacity from 2001 to 2022

[2]. The total amount of wind power generated was more than 2100 TWh in the 2022 survey [3].

The cumulative installed wind power capacity worldwide amounted to approximately 906 GW in 2022. Up to 2021, around 842 GW of the world's total wind power generating capacity was generated from wind energy as shown in Fig. 1 [4].

Many researches are focusing on wind forecasting to identify the wind power potential at various locations. This research paper discusses on developing the wind system model and predicting the wind power using the system advisor model. The system advisor model is easy to analyze the annual wind power at different locations [5–7].

This paper is organized as follows; Sect. 2 introduces the system advisor model. Sections 3 and 4 analyze the wind forecasting performance in different case studies. Finally, Sect. 5 deals with the summary of this work and its future.

2 Methodology of System Advisor Model

The system advisor model software is used to carry out the wind forecasting simulation. The researcher is used to choose the different locations using SAM software. This research investigated wind forecasting at three different locations in the USA.

The system advisor model can calculate the hourly electrical output of a single wind turbine or wind farm. SAM is a desktop application used for renewable energy sources analyses to evaluate system design and financial options. The model of the system advisor, a software development kit (SDK), is a collection of software



Fig. 2 Three different locations (Indiana, Montana, and North Carolina)

development tools that make it possible to write programs in C++, C#, JAVA, and MATLAB or Python, which run SAM Simulation Core (SSC) models [8–10].

Wind forecasting is based on power purchase agreement (PPA) using the financial model and performance model. This research work is chosen for three different wind farmland models that are located at geographical locations such as flat land (Indiana), canyon land (Montana), and offshore (North Carolina) as shown in Fig. 2 [11] and Figs. 12, 13, 14 and 15 show the settings of locations. The wind forecasting parameters vary depending upon the geographical locations.

3 Wind Farmland Model

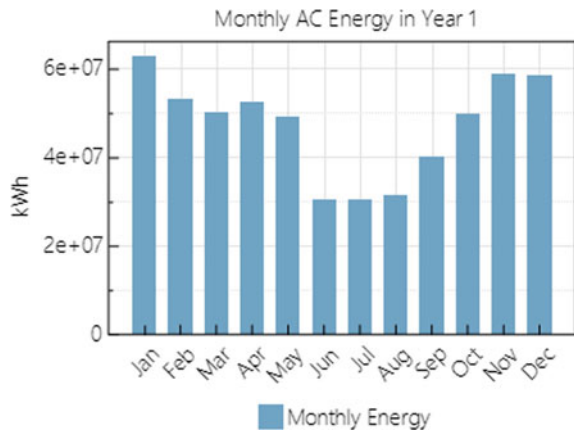
3.1 Flat Land

The performance analysis of the flat land wind turbine model was selected in SAM software. The financial model is chosen in a single ownership function at a power purchase agreement. The flat land is in Northwestern Indiana (NW-IN). The Northwestern Indiana data such as wind speed (m/s), wind direction (deg), wind power (kW), air temperature (°C), etc., is loaded into system advisor model software. The simulated wind power result is summarized as given in Table 1. The monthly and annual wind energy production are shown in Figs. 3 and 4, respectively, and the cash flow of flat land is shown in Fig. 5.

Table 1 Summary of flat land wind power

Metric	Value
Annual AC energy in year 1	5.67083e+08 kWh
Capacity factor in year 1	32.3677%
PPA price in year 1	4 cents/kWh
LPPA levelized PPA price nominal	4.33226 cents/kWh
LPPA levelized PPA price real	3.43969 cents/kWh
LCOE levelized cost of energy nominal	4.47687 cents/kWh
LCOE levelized cost of energy real	3.55451 cents/kWh
NPV net present value	\$ - 8.01618e+06
IRR internal rate of return	6.35597%
Year IRR is achieved	20%
IRR at the end of the project	7.96069%
Net capital cost	\$ 2.79454e+08
Equity	\$ 1.40376e+08
Size of debt	\$ 1.39079e+08
Debt percent	49.768%

Fig. 3 Flat land monthly AC energy production



3.2 Canyon Land

The performance analysis of the canyon land wind turbine model was selected in SAM software. The single ownership function for power purchase agreements is based on the financial model.

The canyon land is in Northwestern Montana (NW-MT). The Northwestern Montana data such as wind speed (m/s), wind direction (deg), wind power (kW), air temperature (°C), etc., is loaded in SAM software. The simulated wind power result is summarized as given in Table 2. The monthly and annual wind energy

Fig. 4 Flat land annual AC energy production

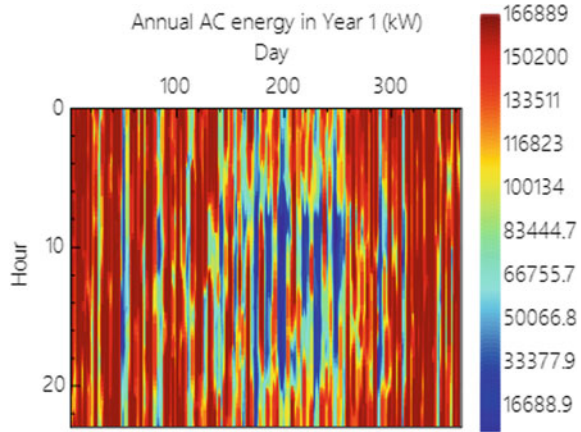


Fig. 5 Cash flow for flat land



production are shown in Figs. 6 and 7, respectively, and the cash flow of canyon land is shown in Fig. 8.

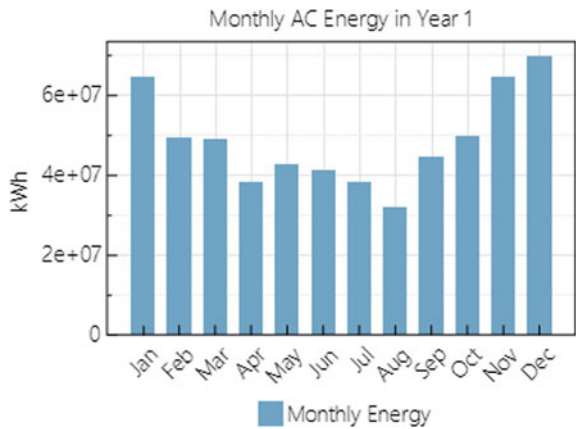
3.3 Offshore Land

The performance analysis of the offshore land wind turbine model was selected in SAM software. The single ownership function for power purchase agreements is based on the financial model. The offshore land is in Eastern North Carolina (E-NC). The Eastern North Carolina data such as wind speed (m/s), wind direction (deg), wind power (kW), air temperature ($^{\circ}\text{C}$), etc., is loaded into system advisor model software. The simulated wind power result is summarized as given in Table 3. The

Table 2 Summary of canyon land wind power

Metric	Value
Annual AC energy in year 1	5.8366e+08 kWh
Capacity factor in year 1	33.3139%
PPA price in year 1	4 cents/kWh
LPPA levelized PPA price nominal	4.33226 cents/kWh
LPPA levelized PPA price real	3.43969 cents/kWh
LCOE levelized cost of energy nominal	4.30971 cents/kWh
LCOE levelized cost of energy real	3.42179 cents/kWh
NPV net present value	\$ 1.28642e+06
IRR internal rate of return	7.83278%
Year IRR is achieved	20%
IRR at the end of the project	9.24424%
Net capital cost	\$ 2.79901e+08
Equity	\$ 1.33859e+08
Size of debt	\$ 1.46042e+08
Debt percent	52.1764%

Fig. 6 Canyon land monthly AC energy production



monthly and annual wind energy production are shown in Figs. 9 and 10, respectively, and the cash flow of offshore land is shown in Fig. 11.

Fig. 7 Canyon land annual AC energy production

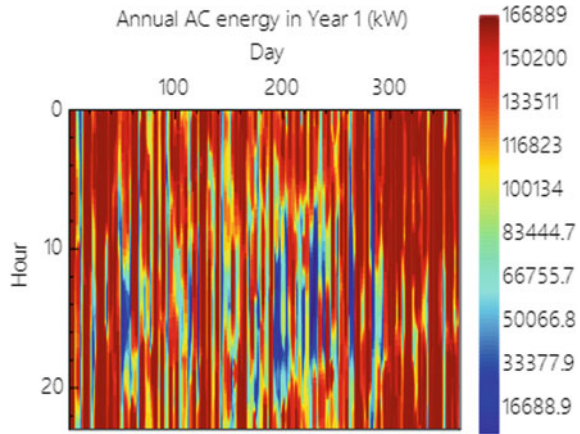
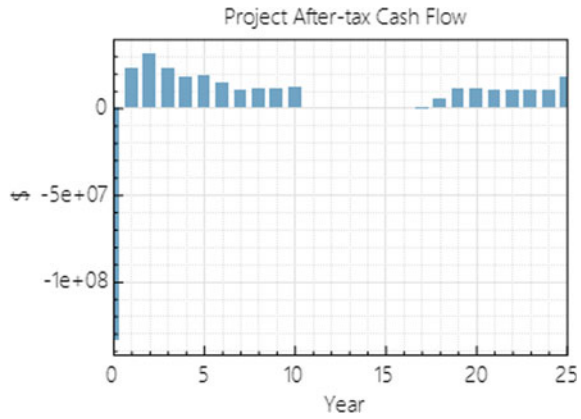


Fig. 8 Cash flow for canyon land



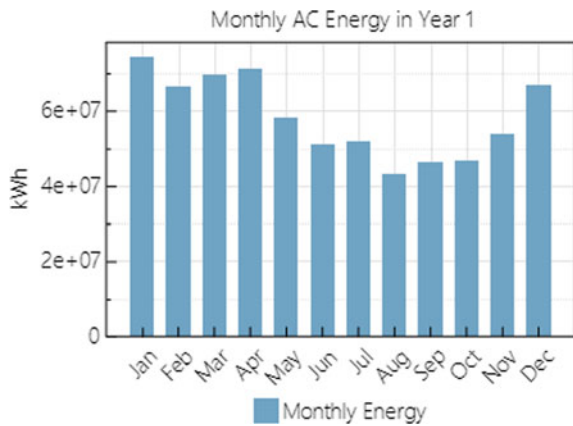
4 Result and Discussion

The wind turbine capacity factor can be set to any level between 0 and 100% by selecting the appropriate rotor and generator sizes. However, the combination of rotor and generator sizes for each geographical location is the maximum for annual wind energy generation [12–14]. In Sect. 3, we investigated wind power forecasting using flat land, canyon land, and offshore land. The capacity factor of wind power forecasting observed in the flat land is 32.3677%, canyon land is 33.3139% and offshore land is 39.9115%. In flat land wind power production is maximum for January month and the June, July, and August months of wind power production is the minimum. The canyon land wind power production is maximum for December and the minimum for August month. The offshore land has maximum wind power production for January and a minimum for August month.

Table 3 Summary of offshore land wind power

Metric	Value
Annual AC energy in year 1	6.99249e+08 kWh
Capacity factor in year 1	39.9115%
PPA price in year 1	4 cents/kWh
LPPA levelized PPA price nominal	4.33226 cents/kWh
LPPA levelized PPA price real	3.43969 cents/kWh
LCOE levelized cost of energy nominal	3.36446 cents/kWh
LCOE levelized cost of energy real	2.67128 cents/kWh
NPV net present value	\$ 6.61513e+07
IRR internal rate of return	23.0639%
Year IRR is achieved	20%
IRR at the end of the project	23.336%
Net capital cost	\$ 2.83015e+08
Equity	\$ 8.84191e+07
Size of debt	\$ 1.94595e+08
Debt percent	68.7581%

Fig. 9 Offshore land monthly AC energy production



The yearly wind forecasting analyzed for levelized cost of energy (LCOE) has 3.55451 cents/kWh for the flat land, 3.42179 cents/kWh for the canyon land, and 2.67128 cents/kWh for the offshore land. The levelized PPA cost is 3.43969 cents/kWh same for all lands. The flat land wind power generation is 5.67086×10^8 kWh per year, canyon land wind power generation is 5.8366×10^8 kWh per year, and offshore land wind power generation is around 6.99249×10^8 kWh per year.

Fig. 10 Offshore land annual AC energy production

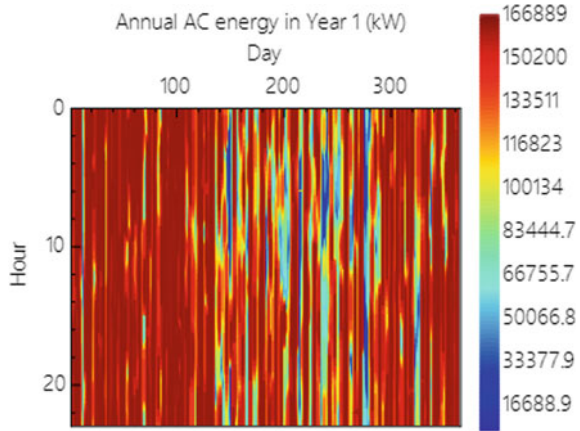
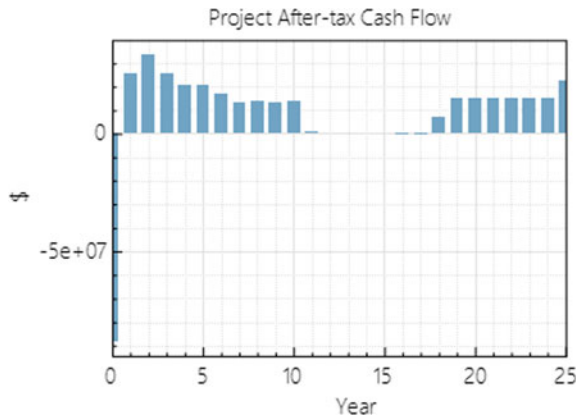


Fig. 11 Cash flow for offshore land



5 Conclusion

We observed that the estimated monthly average wind power is maximum for the month of November, December, and January and the minimum for the month of July and August. The system advisor model simulation result shows that for offshore land of Eastern North Carolina, annual wind energy generation is the maximum and the minimum for Northwestern Indiana. This research reveals that offshore land wind forecasting is the best economy for wind power generation. The system advisor model software helps to design new wind power plants.

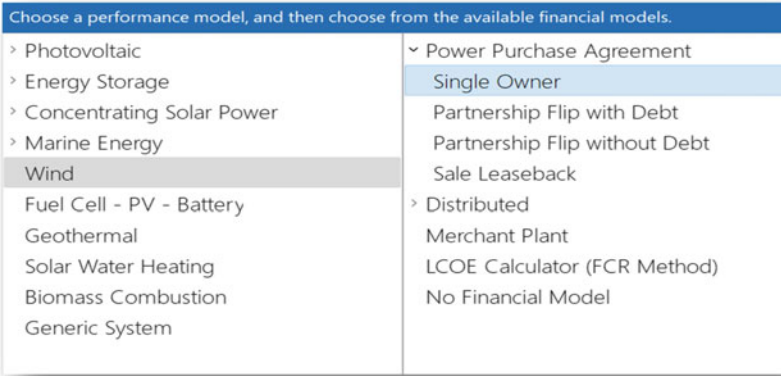


Fig. 12 Selecting wind performance and financial model in SAM

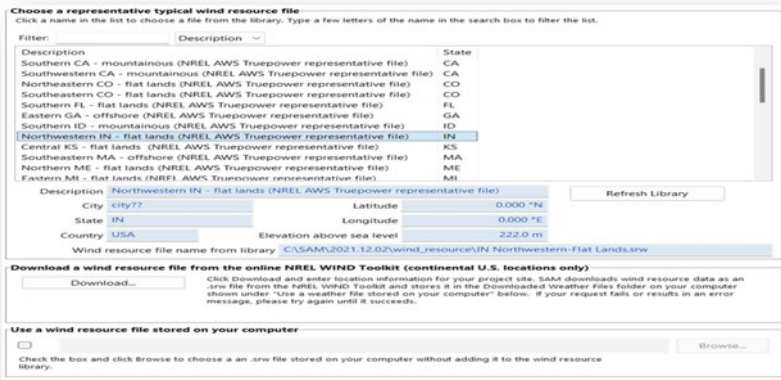


Fig. 13 Selecting Northwestern Indiana location for conducting simulations in SAM

Appendix

See Figs. 12, 13, 14 and 15.

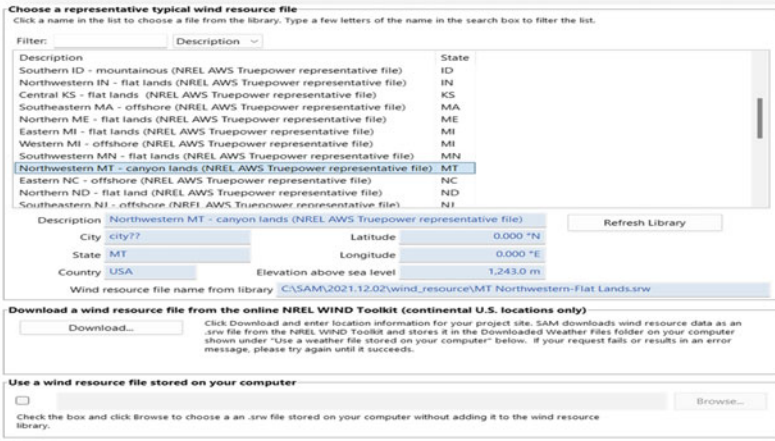


Fig. 14 Selecting Northwestern Montana location for conducting simulations in SAM

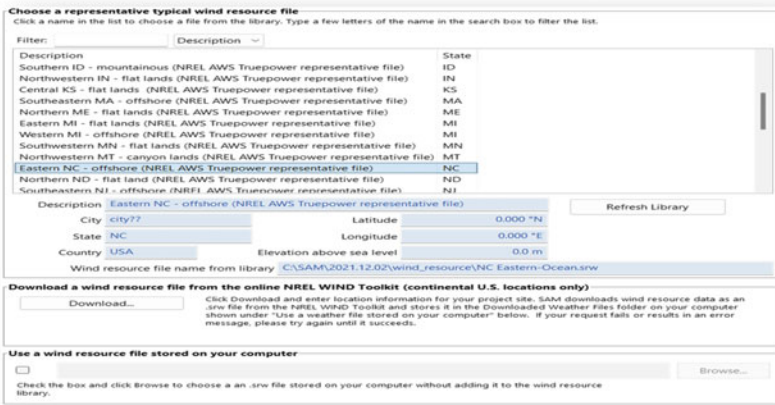


Fig. 15 Selecting Eastern North Carolina location for conducting simulations in SAM

References

1. Rahman A, Farrok O, Haque MM (2022) Environmental impact of renewable energy source based electrical power plants: solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic. *Renew Sustain Energy Rev* 161:112279
2. Zhang K et al (2021) Modeling and utilization of biomass-to-syngas for industrial multi-energy systems. *CSEE J Power Energy Syst.* 7(5):932–942

3. Chen Y, Lin H (2022) Overview of the development of offshore wind power generation in China. *Sustain Energy Technol Assess* 53:102766
4. <https://www.statista.com/statistics/268363/installed-wind-power-capacity-worldwide/>
5. Dobos A, Neises T, Wagner M (2014) Advances in CSP simulation technology in the system advisor model. *Energy Procedia* 49:2482–2489
6. Surendra, H, Sudhindra, K, Archana, H, Madhusudhan, K (2022) Performance analysis of standalone PV system using system advisor model—a case study. In: 2022 IEEE international conference on current development in engineering and technology (CCET). IEEE, pp 1–4
7. Hamilton WT, Neises TW (2023) Dispatch optimization of electric thermal energy storage within system advisor model. *J Energy Storage* 61:106786
8. Freeman J, Jorgenson J, Gilman P, Ferguson T (2014) Reference manual for the system advisor model’s wind power performance model. National Renewable Energy Lab. (NREL), Golden, CO (United States)
9. Freeman JM, DiOrio N, Blair N, Guittet D, Gilman P, Janzou S (2019) Improvement and validation of the system advisor model. National Renewable Energy Lab. (NREL), Golden, CO (United States)
10. Shahzad U (2022) Critical analysis for solar cell models using system advisor model simulations. *J Electr Eng Electron Control Comput Sci* 9(1):23–32
11. <https://education.nationalgeographic.org/resource/united-states-regions/>
12. Boccard N (2009) Capacity factor of wind power realized values vs. estimates. *Energy Policy* 37(7):2679–2688
13. Albadi MH, El-Saadany EF (2009) Wind turbines capacity factor modeling—a novel approach. *IEEE Trans Power Syst* 24(3):1637–1638. <https://doi.org/10.1109/TPWRS.2009.2023274>
14. Pao LY, Johnson KE (2009) A tutorial on the dynamics and control of wind turbines and wind farms. In: 2009 American control conference. IEEE, pp 2076–2089

Performance of Fuzzy MPPT for a Distributed Photovoltaic System Under Changing Environmental Condition



P. Rajakumar, M. Senthil Kumar, A. Sakthidasan, R. Sundar, Baburao Pasupulati, and S. Amosedinakaran

Abstract In recent times, power generation through distributed generation (DG) technique is increasingly adopted around the globe. Particularly photovoltaic (PV)-based DG has been used widely because of the vast availability of solar energy. However, the electrical power delivered by PV cells varies since their characteristics are completely dependent on ambient factors. As a result, having a unique technique to extract the maximum power available from sun irradiation is critical. Such a technique is known as maximum power point tracking (MPPT). In the present study, a simulation analysis is performed on a 5 kW standalone DG PV plant using a fuzzy controller-based MPPT technique. The performance of fuzzy MPPT techniques is evaluated under distinct operating conditions. The simulation study is then extended to include perturb and observe (P&O) technique to evaluate the effectiveness of the fuzzy MPPT controller. The comparative study revealed that fuzzy MPPT provided better results. The simulation investigation is carried out using MATLAB/Simulink software version 2020a.

Keywords Distributed generation · Maximum power point tracking · Fuzzy controller · Photovoltaic system

P. Rajakumar (✉) · B. Pasupulati · S. Amosedinakaran
Department of EEE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India
e-mail: drajakumar@veltech.edu.in

M. Senthil Kumar
Department of EEE, Sona College of Technology, Salem, Tamilnadu, India

A. Sakthidasan
Department of EEE, University College of Engineering Arni, Thiruvannamalai, Tamilnadu, India

R. Sundar
Department of EEE, Muthayammal Engineering College, Namakkal, Tamilnadu, India

1 Introduction

The usage of alternative sources of energy has increased in recent times to meet the ever-growing electrical demand. This is mainly due to compensate for the increase in fuel cost and emission of greenhouse gases. Also, many countries insist on giving incentives for motivating the use of renewable-based power generation. This made a tremendous change in the adoption of renewable energy in the residential and commercial sectors. Among the available resources of renewable energy, solar photovoltaic system is the most popular and prepared one since the source of energy for PV systems is clean and abundant. Because of this, it has been widely used in residential, commercial and industrial purposes. Even solar powers have been utilized for military purposes. Generally, solar power generation ranges from a few kW to large MW systems. In large solar power plants, power loss occurs between generating stations and consumers through transmission and distribution lines. Due to the poor conversion efficiency of solar cells, the cost of the solar plant will increase even further. To reduce the losses, it is preferred to have a power generation closer to the consumer load side. Such a concept is known as distributed generation. A distributed generation is a concept where the power generation is located closer to or on the load centre. Distributed generation uses various sources of energy for the generation of power near load centres. Among them, solar power is widely preferred because of its availability and environmental sustainability [1–3].

Due to the poor conversion efficiency of solar cells, it is essential to have some unique and efficient technique to avail maximum source of power from PV cells under given condition. The power generated from a single PV cell purely relies on factors such as angle of sunlight incident, solar irradiance, environmental temperature and type of load. The typical characteristic curve of a PV cell is nonlinear since its output depends on available irradiance and ambient temperature at a particular location. Due to this nonlinear behaviour of a PV cell, the output power fluctuates in accordance with environmental factors. As a result, MPPT is used between solar PV systems and load/inverter to improve PV cell conversion efficiency. Also, MPPT accommodates a DC-to-DC power converter for effective transfer of power from PV system to load/inverter at required voltage level. MPPT and DC-to-DC power converter help together to achieve maximum power extraction from solar PV cells under given environmental conditions. MPPT can be implemented by changing the duty cycle of the semiconductor switch or by utilizing a closed-loop controller to regulate the PV voltage and current. A brief literature about the adoption of MPPT for PV systems is presented below.

References [4–9] numerous MPPT algorithms were adopted. Among these methods, P&O and InC methods are employed frequently because of its simplicity of design and implementation. Both the methods employ a fixed step size in the operating variable. In PV MPPT choice of step size variation plays an important role in achieving desired performance characteristics. If a small step size is adopted, then tracking speed gets reduced. On the other side, larger step size selection increased

the tracking speed but the system experienced unnecessary power losses due to fluctuation of maximum power point (MPP).

An effective MPPT algorithm should have high tracking speed to reach MPP without producing fluctuation at steady state. For such a technique, adoption of variable step size could produce desired results. If the operating point is far from the target operating point, this strategy leverages larger step size variation in the operational variable. Similarly, smaller step size variation is adopted if the operating point is near the MPP region. This method enables MPPT to reach the desired operating point faster and with less oscillation at steady state. Artificial intelligent techniques are likely to adjust such step size variation according to the operating condition. Fuzzy logic and neurofuzzy systems have been addressed in [10, 11] for such applications. Moreover, FLC-based techniques can also be suitable for nonlinear systems. The important aspect of this FLC technique is that it does not require precise knowledge on the mathematical representation of the system to achieve the required control performance [12, 13]. This feature has made FLC-based MPPT an effective system in the solar energy management for better tracking of MPP. In this paper, a main focus is given on the implementation of FLC-based MPPT for standalone 5 kW DG PV plant. The performance of FLC MPPT is related to the conventional P&O method for different environmental factors.

The content of the remaining paper is presented as follows: Sect. 2 briefs about the PV system configuration. The concept of P&O and FLC MPPT techniques is presented in Sect. 3. In Sect. 4, the simulation test results of the P&O and FLC MPPT methods for various operating situations are presented and examined. Then, Sect. 5 summarizes the finding of this paper work as a conclusion.

2 PV System Configuration

Figure 1 depicts the overall system setup used in the present study. The PV system consists of a 5 kW PV array, a DC–DC converter, and an MPPT controller.

2.1 PV Array

A 5 kW PV array system is modelled for simulation using a single diode photovoltaic cell circuit model, as shown in Fig. 2 [14]. The mathematical description for output current (I_{PV}) of a single PV cell is given in (1).

$$I_{PV} = I_{SC} - I_0 \left\{ \exp \left[\frac{q(V + IR_S)}{nkT_K} \right] - 1 \right\} - \frac{V + R_S I}{R_{SH}}, \quad (1)$$

where (I_{SC} , I_0) points to short circuit current and saturation current of PV cell; q reflects the electron charge; n and k refer ideality factor and Boltzmann's constant,

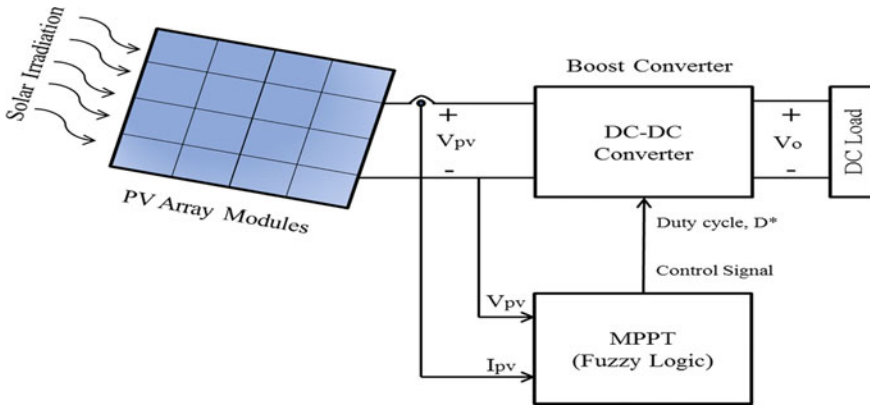


Fig. 1 Configuration of PV system

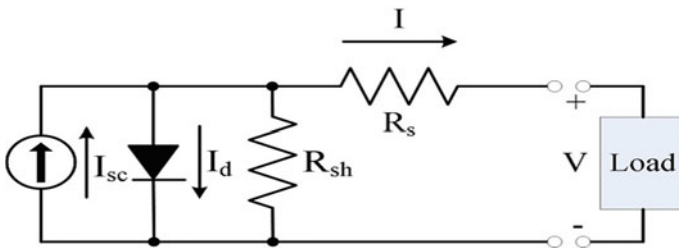


Fig. 2 Single PV cell equivalent circuit model

respectively; T_K is the PV cell temperature; (R_S, R_{SH}) denotes to equivalent series and shunt resistance.

From (1), it is inferred that output current of a PV cell depends on temperature and solar irradiance. In this work, a 5 kW PV array system is modelled using 1Soltech 1STH245WH PV module. Table 1 shows the electrical specification of 1Soltech 1STH245WH PV unit. The 5 kW PV system is modelled using PV modules from NREL System Advisor Model. A single 1Soltech 1STH245WH PV module can deliver maximum power of 245 W for a 1000 W/m^2 solar irradiance and 25°C temperature. For modelling a 5 kW PV system, 21 PV modules are arranged in three parallel strings and each string carries seven series of connected PV modules.

Figures 3 and 4 show the characteristics of a 5 kW photovoltaic system under various operating circumstances. Figure 3a exemplifies the relation between P and V for variable solar irradiance level. Figure 3b demonstrates the correlation between I and V for the variable solar irradiance incident. Similarly, Fig. 4a and b illustrates the characteristic curves of PV array systems for variable temperature conditions.

Table 1 1Soltech 1STH245WH PV module electrical parameter

S. No	Parameters	Values
1	Open circuit voltage (V_{oc})	37.2 V
2	Maximum voltage (V_{pm})	30.2 V
3	Short circuit current (I_{sc})	8.62 A
4	Maximum current (I_{mp})	8.1 A
5	Maximum power (P_{mp})	244.62 W
6	Number of cells	60

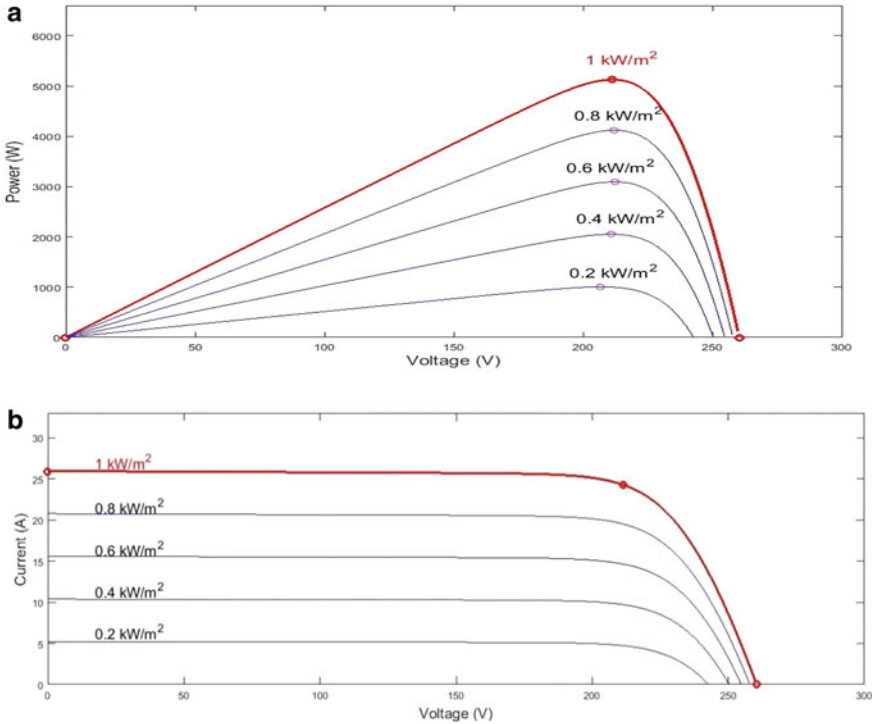


Fig. 3 **a** P–V characteristics for variable irradiance. **b** I–V characteristics for variable irradiance

2.2 Converter Circuit

The maximum power extraction from PV system is ensured through the connection of converter (DC–DC) in the middle of the solar photovoltaic system and the load or inverter circuit. A converter, on the other hand, varies the magnitude of PV output voltage to the necessary voltage level. A typical DC–DC converter is classified as buck, boost, or buck–boost. In this paper, boost converter topology is selected since the output voltage is fixed to 370 V. Equation (2) expresses the correlation between

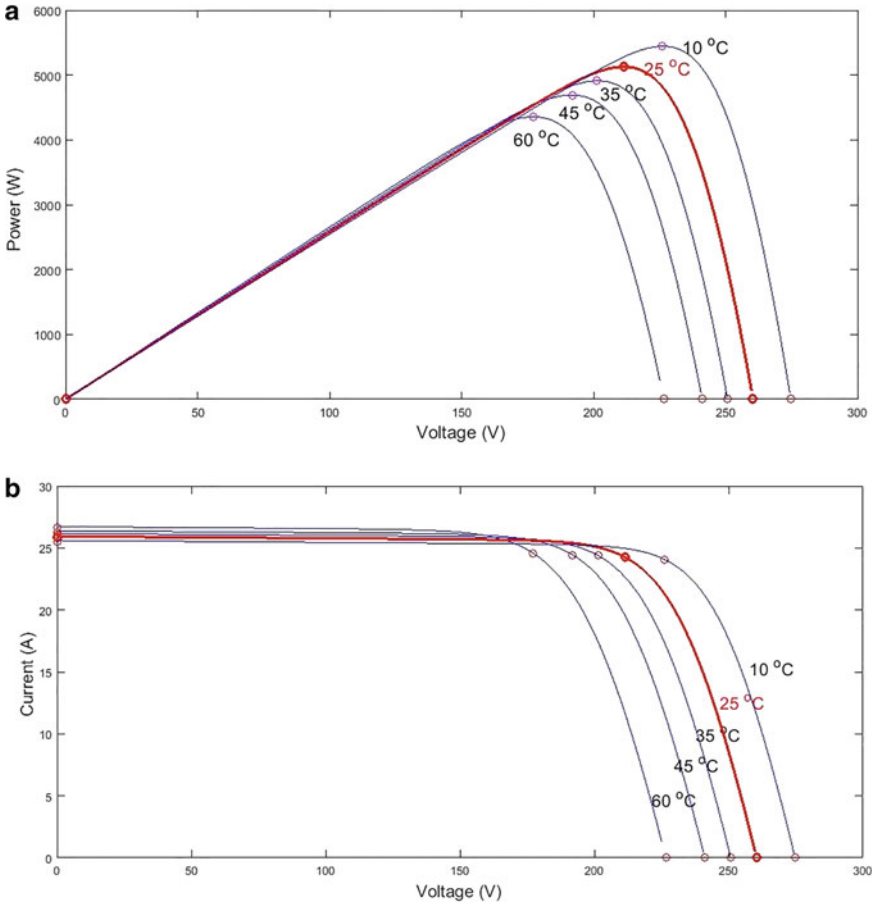


Fig. 4 a P–V characteristics for variable temperatures. b I–V characteristics for variable temperatures

input (V_{PV}) and output (V_O) voltage of boost converter.

$$\frac{V_O}{V_{PV}} = \frac{1}{1 - D} \tag{2}$$

where D refers to duty cycle.

Figure 5 and Table 2 show the circuit topology and design parameters for a boost converter, respectively.

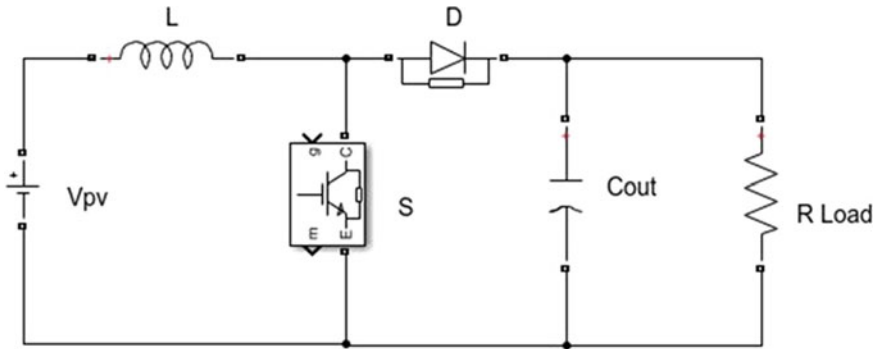


Fig. 5 Topology of boost converter

Table 2 Boost converter design parameters module electrical parameter

S. No	Parameters	Values
1	Input voltage, V_{pv}	180–210 V
2	Output voltage, V_O	370 V
3	Output capacitor, C_{out}	1000 μ F
4	Inductor, L	470 μ H
5	Switching frequency, F_s	10 kHz

3 MPPT Algorithm

The MPPT technology effectively extracts electricity from sunlight. P&O, InC and FLC approaches are three prominent and widely used MPPT strategies in PV systems [15]. In this article, fuzzy and P&O MPPT algorithms are used to test their effectiveness on a 5 kW PV DG unit under various climatic situations.

3.1 P&O MPPT Algorithm

It is a basic and widely used MPPT technique that works by changing the operational parameter. The operating parameter can be a PV array voltage or current. P&O algorithm functions by sampling PV array power for a sample time m interval. The PV array power of two consecutive sampling intervals of m and $(m - 1)$ is computed and related with each other. If the power of the PV array during the m th interval is greater than the power during the $m - 1$ interval, the operating parameter is sampled as the previous interval. If the difference of PV array power during two consecutive sampling intervals is negative, then direction of perturbation is reversed. The above process is repeated till the desired operating point. Figure 6 depicts the flowchart for P&O algorithm.

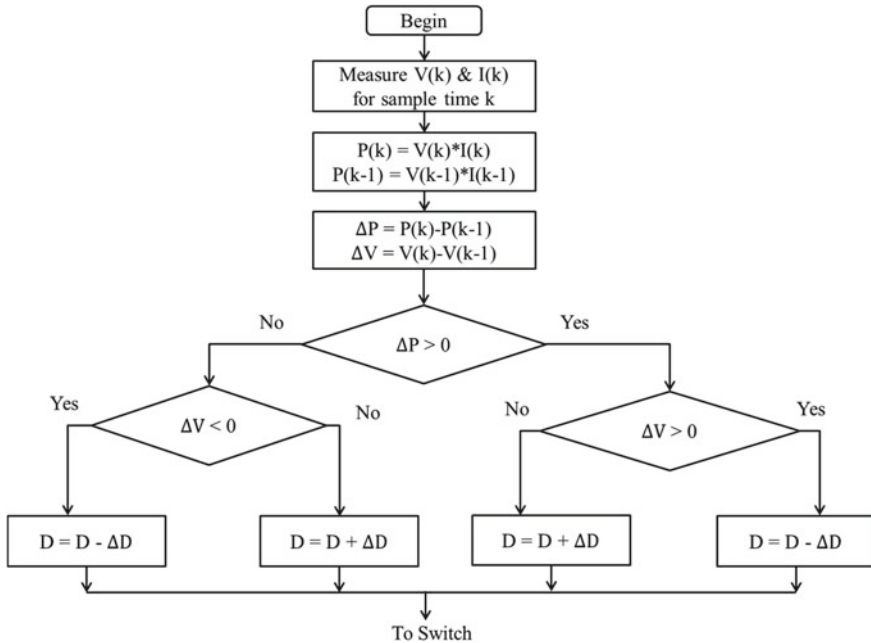


Fig. 6 P&O algorithm-based MPPT technique

The system parameter is perturbed during each sample time of P&O MPPT cycle. This leads to sustained output power oscillation around the MPP region even after attaining the desired operating point. This results in unnecessary power loss and poor conversion efficiency. The power loss will be more predominant in slowly varying environmental conditions. Also, the P&O MPPT algorithm may wrongly track MPP when environmental factors change rapidly.

3.2 Fuzzy Logic MPPT

The typical MPPT approach (P&O) has difficulty tracking MPP in varying environmental conditions and also offers poor conversion efficiency. As a result, the FLC-supported MPPT method is implemented for the photovoltaic (PV) system to have better tracking performance and conversion efficiency under variable environmental circumstances. A typical FLC has four key modules.

Fuzzifier: Converts crisp values into fuzzy input sets.

Fuzzy rules: IF-THEN sentences are employed to define a set of rules for controllers.

Inference: A group of rules are employed to map input to output.

Defuzzifier: Maps output to crisp value.

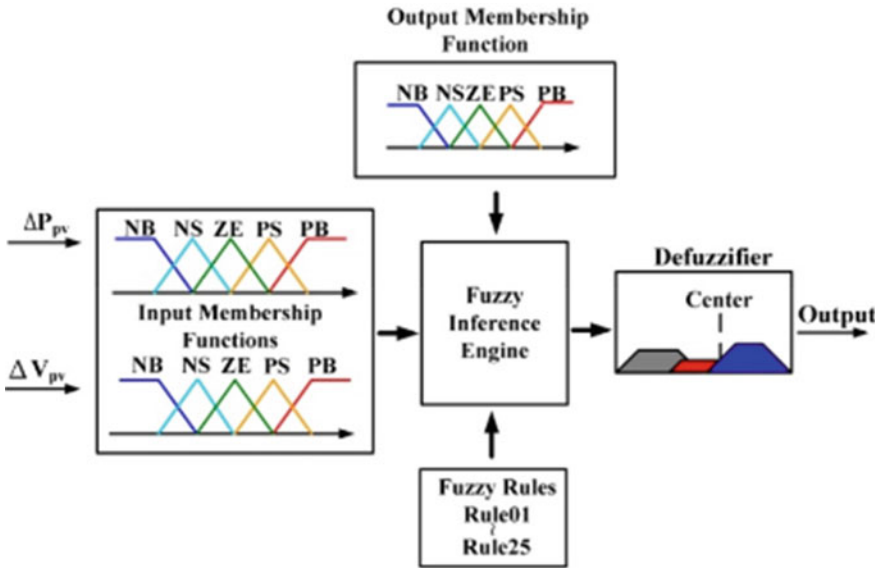


Fig. 7 Components of FLC

FLC components are shown in Fig. 7.

FLC uses different kinds of input and output variables [14]. Here, PV voltage and power are engaged as inputs, while duty cycle (δ) is referred as the output. The choice of PV power and voltage as FLC control variables simplifies design. Equations 3 and 4 define the expression for the input variable.

$$\Delta V_{pv} = V(m) - V(m - 1) \tag{3}$$

$$\Delta P_{pv} = P(m) - P(m - 1) \tag{4}$$

where ΔP_{pv} and ΔV_{pv} refer to variation in input variables for consecutive sampling intervals. The input variables are continuously sampled at m interval time.

Selection of membership values for input and output variables of FLC has a crucial role to achieve better MPP tracking performance. The mapping of membership function values for ΔP_{pv} , ΔV_{pv} and δ is presented in Fig. 8.

The membership's functions are distributed into five distinctive linguistic variables such as BN, SN, Z, BP and SP, where SN and BN refer to negative small and negative big, respectively; SP and BP point to positive small and positive big, respectively, whereas Z points to Zero. Figure 8 illustrates mapping of linguistic variables. Total of 25 rules are incorporated into the inference system for five linguistic variables. Table 3 presents the complete sets of fuzzy rules. Then, defuzzification is done using the centre of gravity (COG) method. The mathematical expression for COG method is given in (5).

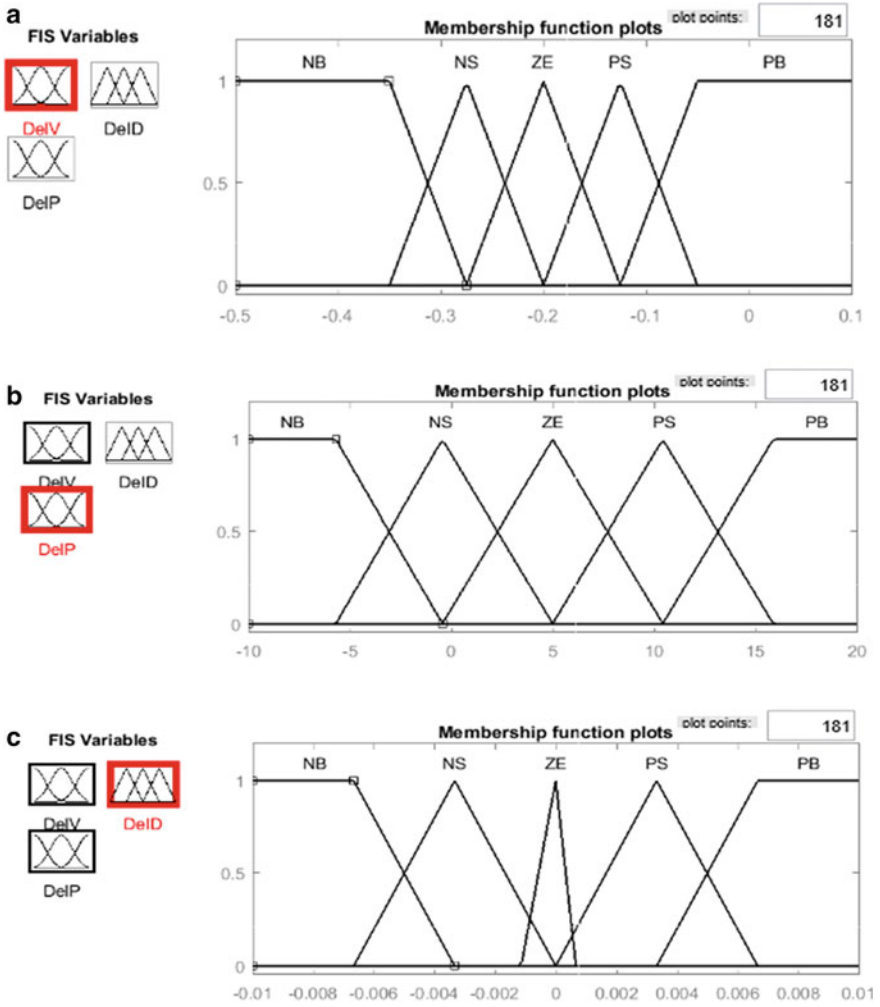


Fig. 8 a ΔV_{PV} membership values. b ΔP_{PV} membership values. c ΔD membership values

$$Y_{COG} = \frac{\sum_{i=1}^n S_i(T_i)T_i}{\sum_{i=1}^n S_i(T_i)} \tag{5}$$

where S_i points to the inference result; T_i refers to output.

Table 3 Complete rule base of FLC system

ΔP_{pv}	ΔV_{pv}				
	BN	SN	Z	SP	BP
BN	SP	BP	BN	BN	SN
SN	SP	SP	SN	SN	SN
Z	Z	Z	Z	Z	Z
SP	SN	SN	SP	SP	SP
BP	SN	BN	BP	BP	SP

4 Simulation Outcomes

The simulation is implemented using MATLAB/Simulink software version 2020a by considering the following assumptions:

- DC resistive load is considered for analysis.
- Converter efficiency is considered as 100%.

A 5 kW PV array system modelled to deliver power to DC resistive load via converter (boost) and MPPT controller. The simulation model for the overall system is presented in Fig. 9. Also, a Simulink model for FLC MPPT method is displayed in Fig. 10. PV system performance with P&O and FLC MPPT approaches is evaluated at various solar irradiance (I_r) and temperature (T) levels. Figures 11 and 12 depict variations in solar irradiance and temperature, respectively. The simulation lasts for two seconds.

4.1 P&O Algorithm

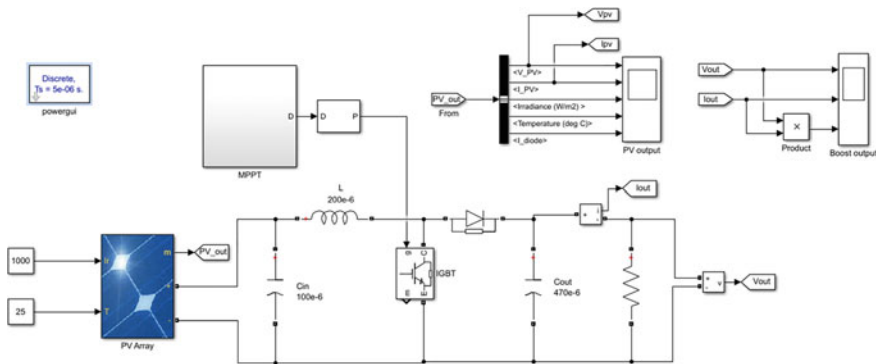


Fig. 9 Overall PV system Simulink model

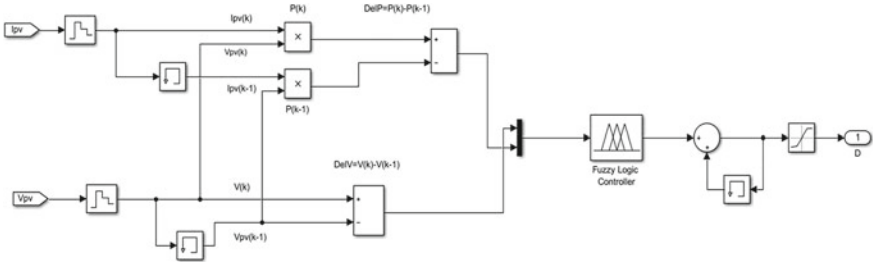


Fig. 10 FLC MPPT Simulink model

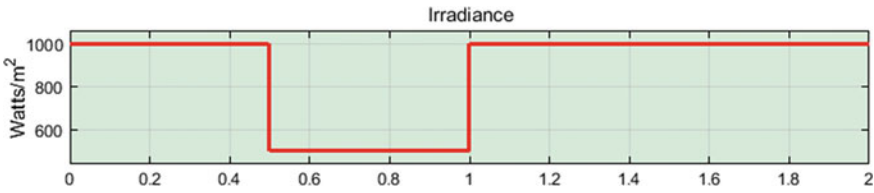


Fig. 11 Solar irradiance variation

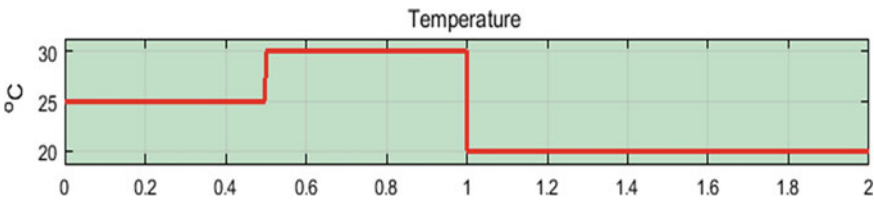


Fig. 12 Temperature variation

Case 1: Variable Irradiance but Constant Temperature (25 °C) Initially, the sun irradiation is held constant at 1000 W/m² between 0 and 0.5 s. However, at time $t = 0.5$ s, the radiation level is lowered to 500 W/m² and remains constant until time $t = 1$ s. Then, at time $t = 1$ s, solar irradiance is once again raised to 1000 W/m². Figure 13 illustrates simulation results for PV systems with P&O MPPT controller for changing irradiance conditions. From Fig. 13, it was evident that the P&O MPPT controller somehow tracks the MPP but produces a sustained oscillation in PV output power.

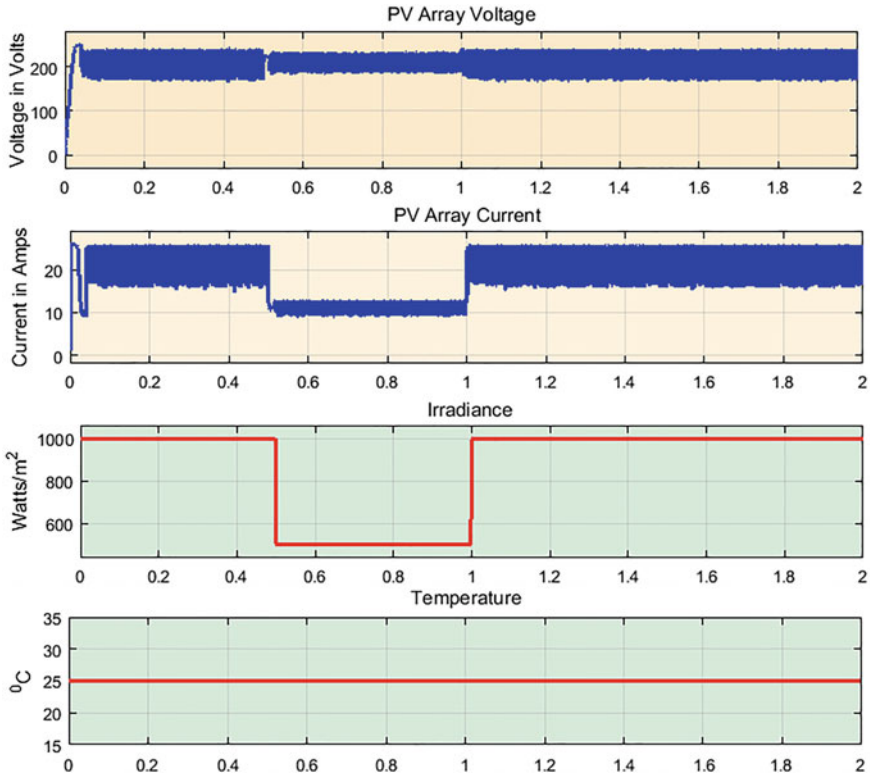


Fig. 13 PV output with P&O MPPT under variable irradiance

Case 2: Variable Temperature but Constant Irradiance (1000 W/m²) For this case, the initial temperature of the solar cell is assumed as 25 °C for a time interval 0–0.5 s. Then, at time $t = 0.5$ s the temperature is raised to 30 °C and is maintained till $t = 1$ s. At time $t = 1$ s, it falls to 20 °C and the same is maintained till $t = 2$ s. Figure 14 depicts the simulation test result of the P&O MPPT controller for variable temperature. Here also, a sustained power oscillation took place around the MPP region. Therefore, referring to Figs. 13 and 14, it was witnessed that there was a sustained power oscillation around MPP when using the P&O MPPT controller. As a result, unnecessary power losses were experienced in the PV system.

4.2 FLC MPPT

Case 1: Variable Irradiance but Constant Temperature (25 °C) PV system output for a FLC MPPT technique under different irradiance is presented in Fig. 15. It was obvious from Fig. 15 that FLC MPPT technique effectively tracks down the MPP

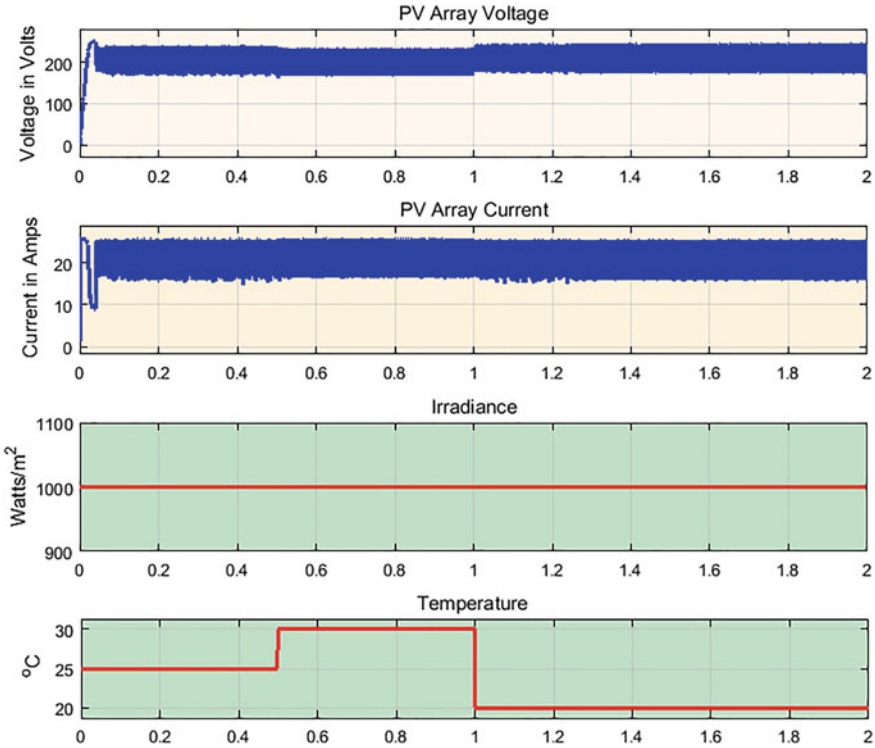


Fig. 14 PV output with P&O MPPT under variable temperature

irrespective of solar irradiance variation. Furthermore, compared to P&O MPPT controllers, power oscillation around the MPP zone is greatly decreased.

Case 2: Variable Temperature but Constant Irradiance (1000 W/m²) Figure 16 illustrates the simulation test result of PV system for FLC MPPT controller for variable temperature. Here also, power oscillations are reduced with FLC MPPT technique but not as much referred to as case 1 condition. The simulation results reported above demonstrated that the FLC-based MPPT approach outperforms the P&O MPPT controller in terms of outcomes and response under varied conditions. In addition to this, FLC MPPT technique registers minimum power losses at steady state. Table 4 summarizes the results of the P&O and FLC MPPT approaches for various sun irradiance and temperature conditions.

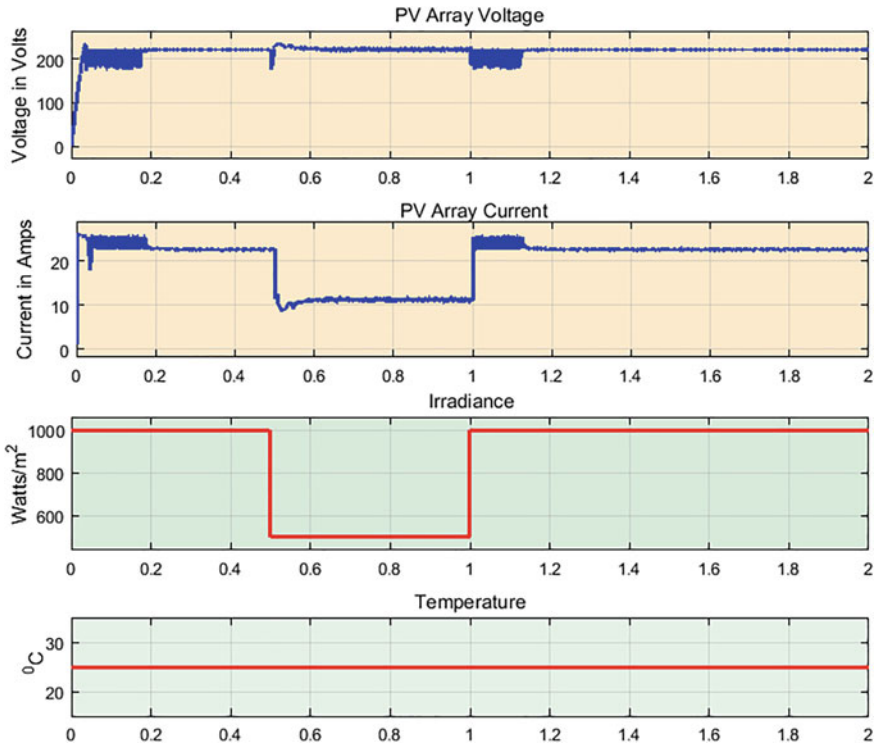


Fig. 15 PV output with FLC MPPT under variable irradiance

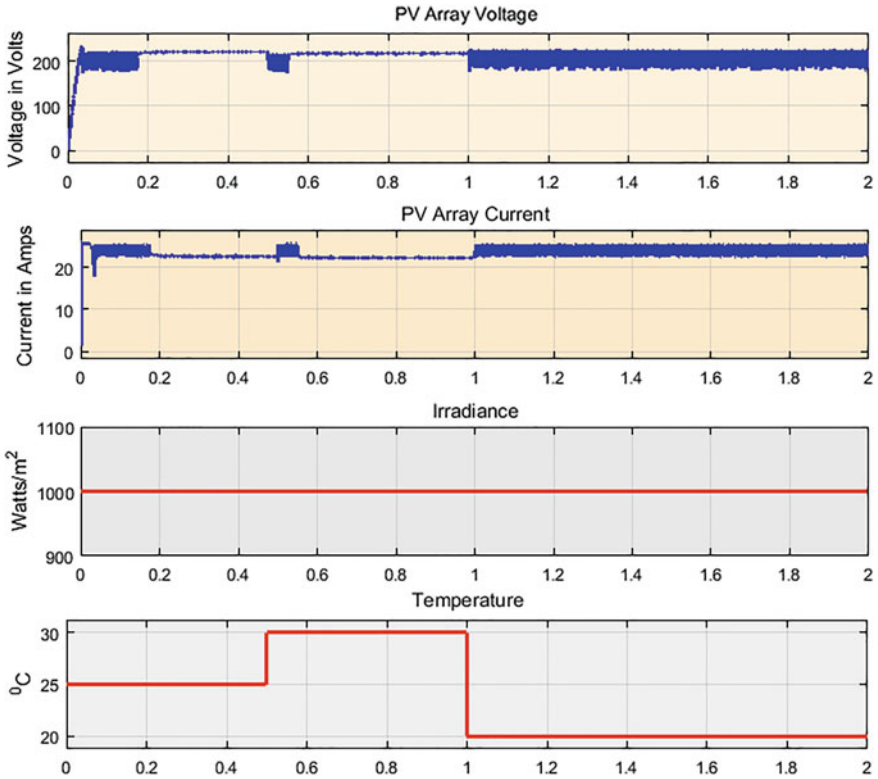


Fig. 16 PV output with FLC MPPT under variable temperature

Table 4 PV output under different solar irradiance and temperature

Variable	MPP parameters					
	P&O MPPT			FLC MPPT		
	V_{pv} (V)	I_{pv} (A)	P_{pv} (W)	V_{pv} (V)	I_{pv} (A)	P_{pv} (W)
Ir (W/m²)	Different solar irradiance					
1000	213.1	21.8	4585	221.2	22.62	5003
500	210.1	11.76	2458	222.6	11.14	2480
T (°C)	Different temperature					
25	213.1	21.8	4585	221.2	22.62	5003
30	212.3	20.4	4331	220.4	22.54	4968
20	210.2	20.1	4225	222.1	22.60	5019

5 Conclusion

The simulation performance of the FLC MPPT technique is investigated in this research for a 5 kW distributed photovoltaic system under fluctuating solar radiation and temperatures. The suggested PV DG system model is implemented in the MATLAB/Simulink 2020a software platform. The simulation test results of FLC and P&O MPPT techniques were related to each other. For a changing operating condition, the P&O MPPT method produced sustained power oscillation around the MPP region. But, FLC MPPT controller has significantly reduced the steady-state oscillation with better tracking response compared to P&O technique. FLC offers simpler design and it does not need knowledge about the system model. Also, inclusion of variable step size in the control variable of FLC provided improved results than P&O technique.

References

1. Lee JS, Lee KB (2013) Variable DC-link voltage algorithm with a wide range of maximum power point tracking for a two-string PV system. *Energies* 6:58–78
2. Shen CL, Tsai CT (2012) Double-linear approximation algorithm to achieve maximum power point tracking for photovoltaic arrays. *Energies* 5:1982–1997
3. Yau HT, Wu CH (2011) Comparison of extreme-seeking control techniques for maximum power point tracking in photovoltaic systems. *Energies* 4:2180–2195
4. Pandey A, Dasgupta N, Mukerjee AK (2008) High-performance algorithms for drift avoidance and fast tracking in solar MPPT system. *IEEE Trans Energy Convers* 23:681–689
5. D'Souza NS, Lopes LAC, Liu X (2010) Comparative study of variable size perturbation and observation maximum power point trackers for PV systems. *Electr Power Syst Res* 80:296–305
6. Abdelsalam AK, Massoud AM, Ahmed S (2011) High-performance adaptive perturb and observe MPPT technique for photovoltaic-based microgrids. *IEEE Trans Power Electron* 26:1010–1021
7. Liu F, Duan S, Liu F (2008) Variable step size INC MPPT method for PV systems. *IEEE Trans Ind Electron* 55:2622–2628
8. Lalili D, Mellit A, Lourci N (2011) Input output feedback linearization control and variable step size MPPT algorithm of a grid-connected photovoltaic inverter. *Renew Energy* 36:3282–3291
9. Mei Q, Shan M, Liu L (2011) Novel improved variable step-size incremental-resistance MPPT method for PV systems. *IEEE Trans Ind Electron* 58:2427–2434
10. Jiang Y, Qahouq JAA, Haskew TA (2013) Adaptive step size with adaptive perturbation frequency digital MPPT controller for a single-sensor photovoltaic solar system. *IEEE Trans Power Electron* 28:3195–3205
11. Mohd Zainuri MAA, Mohd Radzi MA, Abdul Rahim N (2012) Adaptive P&O fuzzy control MPPT for PV boost DC–DC converter. In: *Proceedings of the 2012 IEEE international conference on power and energy (PECon)*, Kota Kinabalu Sabah, Malaysia, pp 524–529
12. Yong T, Xia B, Xu Z (2014) Modified asymmetrical variable step size incremental conductance maximum power point tracking method for photovoltaic systems. *J Power Electron* 14:156–164
13. Liu C-L, Chen J-H, Liu Y-H, Yang Z-Z (2014) An asymmetrical fuzzy-logic-control-based MPPT algorithm for photovoltaic systems. *Energies* 7:2177–2193
14. Shiau J-K, Wei Y-C, Chen B-C (2015) A study on the fuzzy-logic-based solar power MPPT algorithms using different fuzzy input variables. *Algorithms* 8:100–127

15. Subudhi B, Pradhan R (2013) A comparative study on maximum power point tracking techniques for photovoltaic power system. *IEEE Trans Sustain Energy* 4(1):89–98

Enhancing Photovoltaic Connector Reliability: A Comparative Review of Studies with Practical Recommendations



Kartik Kapoor, Muthusamy Eswaramoorthy, Devendra Goyal, P. G. Nikhil, and Prem Prakash Bharti

Abstract The failure of photovoltaic connectors influences the performance of solar photovoltaic power plants. Various studies have been carried out across the globe on its reliability, degradation, cross-mating, and fire hazards caused due to its failure. Most reliability studies on photovoltaic connectors have used accelerated stress testing procedures based on IEC 62852 in a controlled environment, and degradation studies are based on the arc fault assessment. Advanced testing procedures have been employed to determine the material's morphology, including corrosion in metallic contacts and fretting, to develop a reliability model for predicting connector lifetime and a degradation model for the connector. A comparative analysis has been carried out in this paper of all the studies, and it is highlighted that if these studies were also carried out on field-aged PV connectors, it would have brought a clearer perspective on connector reliability and degradation. A practical field study on connectors carried out in this paper on an existing photovoltaic plant highlights the practical issues a connector faces, probable causes of its failure, and remedial measures in safety, operation, and maintenance. It concludes to implement a standard design of connectors

K. Kapoor (✉) · M. Eswaramoorthy
School of Mechanical Engineering, Shri Mata Vaishno Devi University, Katra, Jammu and Kashmir 182230, India
e-mail: 20dme002@smvdu.ac.in

M. Eswaramoorthy
e-mail: m.eswaramoorthy@smvdu.ac.in

D. Goyal
Hi- Physix Laboratory Private Limited, Ranjangaon, Pune 412220, India
e-mail: svrphplindia@gmail.com

P. G. Nikhil
Solar Radiation Assessment Division, National Institute of Solar Energy, Gurugram, New Delhi, India
e-mail: nikhilpg.nise2@gmail.com

P. P. Bharti
Skill Council of Green Jobs, New Delhi, India
e-mail: prem@sscgj.in

globally, a performance standard for PV connectors, and an industry-driven approach for training in installing and maintaining connectors.

Keywords Photovoltaic connector (PV) · Reliability · Degradation · Life cycle analysis · Accelerated stress testing · Contact resistance · Fretting · Corrosion · Electrical contact

Abbreviations

AC	Alternating current
AST	Accelerated Stress Testing
BoS	Balance of Systems
C-AST	Combined-accelerated stress testing
DC	Direct current
EDX	Energy-Dispersive X-ray Spectroscopy
EN	European standards
I_{sc}	Short circuit current
IEC	International Electrotechnical Commission
IR	Infrared imaging
kWp	Kilowatt peak
MWp	Megawatt peak
NEC	National Electrical Codes
PID	Potential-Induced Degradation
PV	Photovoltaic
Riso	Insulation resistance
SCB	String combiner box
SEM	Scanning Electron Microscopy
STC	Standard Test Condition
TI	Thermal imaging
TUV	Technischer Überwachungsverein
UL	Underwriter laboratories
UV	Ultraviolet
V_{oc}	Open circuit voltage
XCT	X-ray Computed Tomography

1 Introduction

1.1 Background

The connection of the solar cables takes place with standardized PV connectors and is used to connect and disconnect between circuits quickly. For the system to perform efficiently, solar connectors must provide low-resistance connections to avoid the circuit's resistive losses in the circuit. They should be designed to have an operational life of 25 years. The industry still lacks a standardized connector design and is presently based on a male part, an integral plug that encases a contact, and a female part, which is a socket with an extended contact. These connectors are weather-resistant, UV-proof, designed for continuous outdoor use, and widely used in series and parallel connections between photovoltaic devices. They terminate each other with a notched interlock which prevents the cables from being unintentionally pulled apart, and it requires a specialized tool to disconnect, as depicted in Fig. 1. However, the bridging position of the connector also implies that it may be the most unsafe hidden danger in solar systems. During the active operation of solar PV systems, solar connectors are overlooked, but when they fail, they lead to heavy damage to the system, due to which the maintenance costs become humongous [1].

The PV generation system comprises components to include PV modules, combiner boxes, protection devices, inverters, and transformers. PV connectors establish connections between these components and cables as described in the DC side of a power plant as depicted in Fig. 2. The solar PV panels are attached to solar cables to carry the electricity generated to the solar circuit. Failure of these interconnecting components affects the power plant's generation capacity and is a

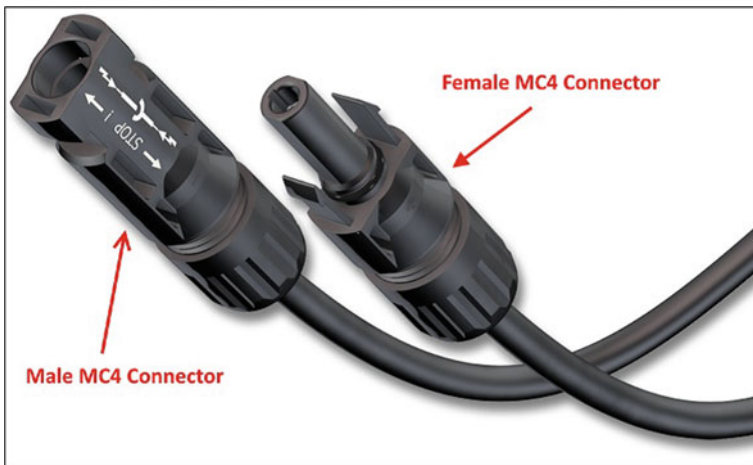


Fig. 1 A solar PV connector with male and female parts

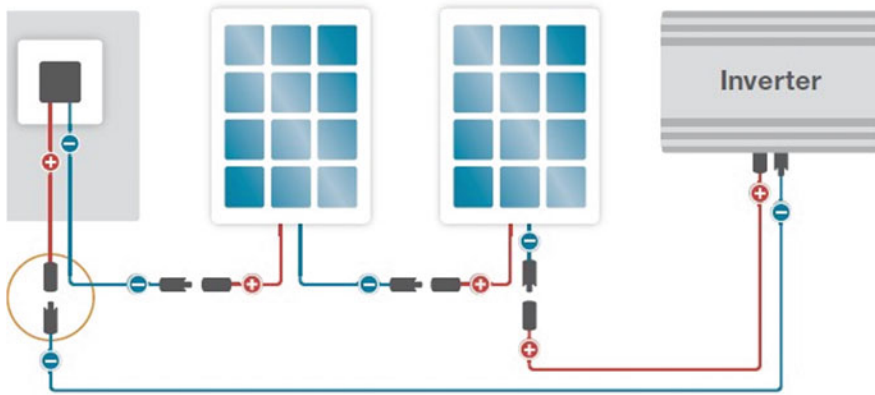


Fig. 2 DC side of a grid-tied PV plant showing interconnections with PV connectors

considerable safety threat leading to fire and shock hazards. Failures in such components can also result in open circuit and short circuit faults and leakage-driven impacts [2].

Elaborate efforts and research have been carried out on PV module and inverter reliability and degradation, but the same cannot be said for PV connectors. These connectors are a part of the BoS in a PV plant and are often neglected as they cost 0.5% of the initial investment of a plant. The industry also believes that the quantifiable replacement rate of connectors is 1% of the strings per year; however, the operator takes it to a lifetime of 25 years as per existing warranties and codes, which seems to be an overkill. Furthermore, it has been ascertained through various studies that 12% of the PV module failures [3] and 29% of total fire incidents [4, 5] reported are due to the failure of PV connectors which is a very high percentage, making them a critical component in the system.

1.2 Literature Review

The subcomponents of photovoltaic (PV) power plants are exposed to some of the severest outdoor conditions. Most components are subjected directly to the environment and various stresses (micro- and macro-environment). Additional characteristics such as local site conditions, installation workmanship, quality, and maintenance influence the probability of such hazards. A sizeable quantity of connectors is scaled according to the PV power plant size and layout. In a 10 kW_p solar photovoltaic plant, there are usually 60–90 such connectors. The life and operation of a connector depend on many factors, whether it is the correct installation and maintenance, quality of the product, and no cross-mating to its working states [6]. All these factors are essential for the standard operation of a solar PV plant. While the occasional open circuit failure is an issue for the operators, however, degradation caused

due to an increase in contact resistance further leads to resistive heating due to high $I^2 R$ loss and ultimately causes arcing and, finally, burning of the connector. Most of these failures are due to poor workmanship quality (poor connections) caused due to moisture and dust ingress. It can permanently damage modules and, eventually, causes a loss of power output. The annual potential power loss observed in one study due to an increase in contact resistance of a connector was estimated to be 140 W-h per string [7]. In some studies, arc faults are documented with the increase of contact resistance of connectors, and its prevention is identified as a critical knowledge gap [8].

Material properties of the PV connector casing are primarily plastic which also plays a critical role in determining the rate of the severity of these failures [9]. As insulators, PV connectors are subjected to high electrical fields throughout their service life, besides undergoing mechanical and thermal stresses during installation and maintenance. Depending on environmental factors like temperature, humidity, dust, pollution, and UV exposure [10], these stresses vary from one geographical location to another. Connectors have three primary failure mechanisms wear, corrosion, and stress. Corrosion represents the degradation of the contact interface through the ingress of dust and moisture [11], whereas stress refers to the gradual change in material morphology with age. Wear in a connector leads to fretting, which is the relative movement of the contact interface. The report [9] has studied five different plastic materials for candidacy for PV connectors and tested them based on environmental stresses, including flame retardance and impact resistance. The best overall results were observed on polycarbonate/silicone copolymers.

Primarily international standards or codes for connectors are IEC 62852 [12], UL 6703 [13], add IS 16781 [14], which mainly focus on safety standards by applying various functional metrics tests, namely current resistance, electrical insulation, mechanical force, and accelerated stress test. Connector durability testing is mainly concerned with realistically stressing products to predict long-term performance and with routes to failure mechanisms, rates of performance, or loss [15]. These methods find out about the safety characteristics. The universal methodology is to apply higher stress levels than actual site conditions—but over a shorter period to forecast longer-term performance and mortality rate. This methodology carries the probable caution that it may cause failures that might not occur naturally but is beneficial for studying failure modes and product robustness. The accelerated testing approach is primarily used for reliability and degradation. It uses near-normal stresses but is applied over a much shorter period to age the material to its useful operational life. In addition, a loss of material or product durability may lead to catastrophic failure, i.e., loss of reliability. There are some advanced detection techniques such as X-ray computed tomography (XCT), energy-dispersive X-ray spectroscopy (EDX), scanning electron microscopy (SEM), and cross-sectioning used for identifying the material morphology and corrosion in a connector.

1.3 Motivation

During the ongoing research work on the energy and safety audit of plants on the field, we noticed that there is a large number of ground and arc faults that have occurred in the plants, and most of the plant operators do not pay attention to the connector maintenance being a slight cost in their capital investment. So when we were doing the literature survey on the topic, it came to light that there has been a limited study on connectors. Therefore, to contribute to this nascent but critical component of BoS, we have analyzed the finding and studies on the PV connectors for the operators, manufacturers, and policymakers to better understand the issue and to find the best possible solution to tackle this potential overgrowing hazard.

The key contributions to this manuscript are:

- A detailed analysis of the present ongoing studies of PV connectors.
- The critical issues in a connector and its remedial measures for the environment.
- Issues observed with connectors in our present ongoing field study.
- The significant shortcomings/limitations in the ongoing studies and recommendations to overcome these.

2 Studies on PV Connectors

2.1 Studies on the Reliability and Degradation of Connectors

Most of the studies on solar PV connectors are based on accelerated stress testing in laboratories and are primarily done on new control sample connectors accelerated aged to 25 years under the control environment. In studies [16] on the reliability and degradation of photovoltaic connectors, no corrosive effects of dust on the connector pins through damp heat accelerated testing at 85 °C/85% RH was observed. However, it led to a minimal increase in resistance in the first 100–250 h of testing. A report on PV module failures published by International Energy Alliance (IEA) [17] states that the failure of the PV connectors is caused by improper fitting or wrongly crimped connectors that lead to power loss in a whole string or even cause electric arcs and fires. There is an increase in contact resistance of PV connectors with time due to varied field conditions, which results in higher power loss in the DC field of the solar plant. The findings estimate a contact resistance of 5 mΩ between connectors which increases with time and can lead to an increase in temperature along with the change in dimension due to expansion and contraction, which further leads to a change in stiffness and morphology, eventually leading to the fast rundown state of connector. Corrosion occurs through contamination ingress at the contact interface [18, 19]. The reliability of connectors under accelerated fretting testing showed a linear increase in contact resistance for the first 10,000 thermal cycles; however, suddenly, a sharp increase led to its failure. It was also studied that a lubricant reduces the friction

coefficient of the mated connector interface, thus offering protection against fretting and corrosion to the connector interface [20].

Scandia National Laboratories [21] reported on the reliability of connectors tested with the accelerated stress testing method on 75 new connectors from three manufacturers. The comparison highlights only a 9% difference in contact resistance between manufacturers. The corrosion effects of grime on connector pins were also studied when these were subjected to 450 hours of damp heat testing; the results show a slight increase in resistance within the first 100 hours, and later no changes were recorded. Also, Sandia National Laboratories [22] reported the reliability based on the arc fault risk of the connector. The study shows that the new connector is resilient to corrosion after thousands of hours of damp heat cycles and corrosion tests. It also revealed that arc fault risk in a connector depends on its design geometry, and a degradation model was made to formulate a data-driven plan for further connector maintenance. Three connector designs were studied [23] for their reliability on arc fault diagnosis, and it was observed that during the mixed flow gas corrosion test, the connectors exhibit Joule heating that can increase the arcing behavior and suggested that the connector temperature monitoring in the field is most appropriate arc fault prognostic method. In a separate study done at NREL [24], field samples of connectors were put under accelerated stress testing to clarify the environmental stressors and effect of mechanical perturbation especially designed for BoS components. It was brought out in the study that maximum degradation occurs at the metal pins of both the male and female connectors with mechanical perturbation. In this study, it was also highlighted that degradation was obtained on connector pins that were caused due to oxidation, inelastic deformation, and corrosion in both the field-aged and accelerated-aged samples. The failure location was identified as the hottest in both pins.

2.2 Fire Incidents Attributed to Connector Failure

According to statistics [25], from 2010 to 2017, there were 58 solar system fire accidents in the UK, of which 27% were attributed to connector failures. The prime cause of fires is electrical arcing, where a current flows across an air gap by ionizing the air. The components responsible for PV fires were DC connectors and isolators, inverters, PV modules, and junction boxes. The causes of identified electric arcs were attributed to the fact that the components were either incorrectly specified, poorly installed, or had manufacturing defects.

In a similar study in Germany from 1995 to 2012, 180 photovoltaic system fire accidents occurred, of which 24% of incidents were attributed to connector failures [26]. In a study carried out in Japan from 2008 to 2017, 127 photovoltaic system fire accidents occurred, and the primary cause identified was panels and interconnections, including connectors [27]. A numerical review of fire sources in 75 PV systems, which caught fire, shows that the chance of the quick connector causing the fire (29%) is nearly as high as for the rest of the module (34%) or other parts of the PV

system (37%). However, no such data is available in India, which states that these incidents have not either been reported in India or that developers and installers are unaware or do not want to report such incidents. As of December 2021, there were an assessed 3.5 billion PV connections globally. Each connector represents a potential point of failure, but only a tiny fraction of them are regularly monitored. Fires in operating assets with faulty connectors are preventable, but only with the correct inspection, design, certification, and testing.

2.3 *Cross-Mating of Connectors*

The various standards elaborate non-cross-mating of PV connectors. To ensure safety, connectors get tested as part of their certification process through EN 62852 and UL6703. UL connector certification only applies if products from the same family have been mated. However, different brands of connectors, one female of brand X and one male of brand Y, do not typically go through this process, which means that the term compatible is only on the shape of the termination. Therefore, ideally, connectors must never be cross-mated unless tested and certified together as a single component. UL1703 STP Module norm requires module manufacturers to list specific manufacturer and model numbers of all allowable PV connectors on their module and assembly instructions.

- IEC 60364-7-712:2017 [28] states that ‘male and female connectors shall be of the same type from the same manufacturer’.
- The installation norm IEC 62548, Para 7.3.9 Plugs, sockets, and connectors [29] specifies that plugs and socket connectors mated in a PV system shall be the same type from the same manufacturer to make a connection.

As part of a technical assessment of solar project risk, IEA [30] assessed different PV module failure modes—according to the severity of effects, probability of occurrence, and detectability and ranked PV module failures according to a risk priority number (RPN). According to this failure modes and effects analysis (FMEA), failures associated with different connectors characterize the highest risk to PV system performance and safety. In addition, it highlighted the adverse consequences of joining different connectors and highlighted regular inspection and monitoring during the operation and maintenance phase, and clauses for similar and spare connectors during tendering stage in modules, inverters, and cable runs (Fig. 3).

Generally, cross-mated pair of connectors will connect and pass electricity without issues. However, the mismatching of design dimensions and material compatibilities over time can further lead to loss of weather tightness and increased contact resistance, leading to eventual disastrous connector failure. According to the article published in a reputed solar magazine [31], the severity of cross-mating of connectors, with dreadful consequences, including burnt connections, arcing failures, and structural fires, has been highlighted.

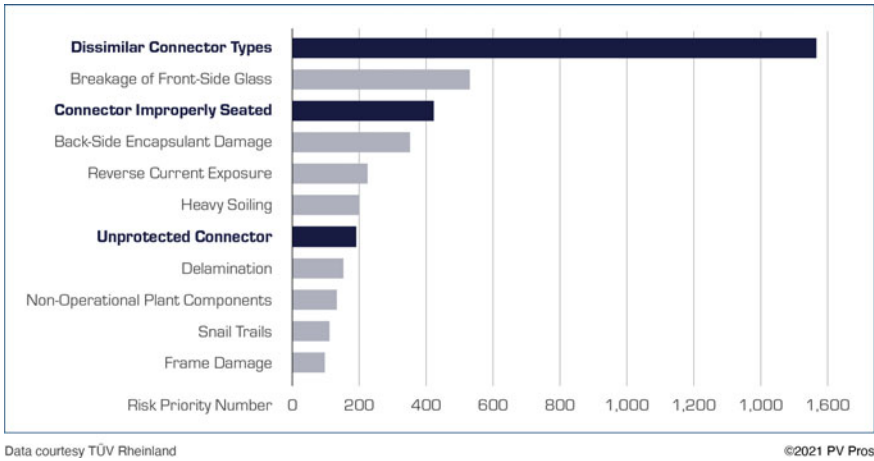


Fig. 3 FMEA rating for PV module failures [30]

The study carried out by a reputed industry brand [32] for PV connectors explains that using combinations of different PV connectors can cause severe damage and faults due to the chemical and thermal incompatibility of the material. Different thermal expansion parameters of the metal contact can also lead to contact corrosion after some time. This leads to the PV system being at-risk for people and the environment. Each product manufacturer designs and builds PV connectors to its specifications, and the material commonly used is either polycarbonate or polypropylene, having different material properties and thermal coefficients of thermal expansion. These differences lead to failures and some of which will lead to fires.

3 Connector Issues and Diagnostic Measures

3.1 Root Cause Failure Analysis (RCFA) of Connectors

Root cause and failure analysis (RCFA) investigates the origin of failure in a product, process, or service and uses the information to develop and implement a change to prevent the recurrence of the failure. Studies and articles [17, 31, 33–36] have highlighted the root causes of failures of operational connectors in PV plants:

- Due to improper workmanship, the connectors are not fitted correctly or are improperly crimped during installation and maintenance leading to the ingress of moisture and dust, which causes the heating of connectors.
- When a module is being replaced or installed, cable and connectors remain hanging for more than a few days, and during this time, the site might suffer rain and wind/storms, which leads to dust deposition and moisture ingress in connector

pins. This leads to connectors being corroded at a faster rate and being more prone to increased resistance which leads to overheating, leading to its failure.

- Improper size of DC cables leads to improper sealing at the end gap of the connector, which further leads to the ingress of moisture and dust in the connector and may cause an insulation failure of a connector. This leads to isolation failures in the long run due to short circuits. An isolation fault produces potentially lethal voltages in conducting part of the system, which may damage an inverter and a safety hazard.
- Finding similar-looking connectors in the present market is easy, as no standard design exists. A slight difference in design can lead to moisture and dust ingress, and its IP 68 rating, which should be certified, is lost. Problems that occur due to incompatibilities of low-quality plastic materials used in the manufacturing of DC connectors lead to degradation of the casing of the connector and can cause a temperature rise, thus leading to overheating of the connector.
- Cross-mating male and female connectors during maintenance increase the loss of performance due to mismatch losses and may cause defects and faults such as corrosion, degradation of a material due to heating, and electrical arcing in the long run, thereby causing a fire incident.

3.2 Improper Quality and Size of PV Connectors Lead to the Following Issues [36]

Reliability

Solar PV connectors are subject to work for 25 years in extreme weather conditions; hence, low-quality materials cannot endure such stresses.

Mismatch loss

PV connectors using different materials or brands for a single PV string will create a mismatch of connections, resulting in a poor connection.

Fire hazards

Low-quality material may cause a rise in connector temperature, which can result in overheating of the system and lead to fire which can cause water and dust to ingress and can cause electrocution.

Chances of electrocution

When a connector with poor connection or insulation is touched, it reduces its insulating properties, which may cause electrocution. It is recommended to use proper size cables for interconnection and proper sealing at the endcap of the connector to prevent moisture and dust ingress.

Financial Loss

In identifying faulty connectors and their repair/replacement, extra manpower costs will be incurred. Moreover, faulty PV connectors lead to a loss in power generation and revenue for solar PV power plants.

3.3 Connectors Failure Detection Challenges

Determining connector quality in the field is a challenge. For example, by visual inspection, some of the exterior faults can be observed, but it does not confirm the quality of the internal connection. However, connectors can safely be unplugged when not under load and cannot be disassembled without compromising their integrity. Faulty connectors have a high resistance than correctly assembled connectors. It causes $I^2 R$ losses and produces heat, and decreases energy yield. As a result, high temperature is the key warning sign of imminent failure. If significant issues exist during installation, thermal identifiers will be seen when the plant is commissioned. However, thermal identifiers may not appear, even when dormant defects exist until the plant has operated for some time. Bad connectors can be challenging to detect when poor wire management is seen in the plant. It is pertinent to highlight that even field inspectors find tracing and detecting these connectors challenging, even with handheld thermal cameras.

3.4 Diagnostic Methods for Detection of Connector Failures

In field test methods, only visual inspection and infrared imaging are available to find out about the probable failures of connectors, whereas there are tests available in laboratories that are routine and advanced to detect the cause of failures.

Visual inspection

It is done on connectors to identify the following issues: separation or loose connection, connector with loose back nuts or not fitted properly, improper bending of connector wires in a string, the wrong polarity of connectors in a string, signs of heat deformation and evidence of cross-mating. A push-pull test will be performed on the connector to check whether the male and female parts are jammed due to dust or any other visual effect.

Infrared imaging

It is performed with a handheld long-wave infrared camera on connectors to identify the heat signatures of these components. Though there is no technical standard at what temperature the connector is a failure, in a recent study by Heliovolta and PVEL [36], absolute temperature failure occurs when the connector displays a

temperature beyond 85–95 °C. Differential temperature failures are project specific when connectors operate below the threshold absolute temperature and above 60 °C. These connectors still present performance and safety risks.

Routine non-destructive testing

It is carried out in laboratories, as per IEC 62852-2015 includes leakage current test, voltage withstand test, Ingress Protection (IP) test and contact resistance test. The contact resistance test uses a megohmmeter to measure resistive heating and changes to resistance from thermal effects, whereas leakage current/insulation resistance tests evaluate connector body integrity. As per IEC 62852, connectors are subjected to accelerated stress testing in high temperature/humidity conditions and thermally cycling in a laboratory environment to accelerate the failure of suspected bad connectors.

Advanced forensic testing

It is a destructive procedure conducted in laboratories and requires sophisticated equipment. Neither thermal imaging nor visual inspection reveals the inner workings of connectors. These forensics and material analyze connectors to determine the root causes of failure, such as manufacturing defects, installation handling, inadequate standards, environmental stress, and extreme weather. PVEL and Scandia National Laboratories [36, 37] recommend the following techniques for advanced analysis:

X-ray computed tomography (XCT)

It creates a 3D image of the connector, including metal and plastic components, by acquiring radioactive projections from different angles. It is done to inspect, evaluate, and analyze a product.

Scanning electron microscopy (SEM)

It is used to identify any contaminants in the metal's microstructure and evidence of corrosion and arcing products on the surface. It can also identify morphology and chemical composition. SEM scans the sample with an electron beam to produce a 2D image of an object for analysis. It is used for microanalysis and failure analysis of solid inorganic materials.

Energy-dispersive X-ray spectroscopy (EDX)

It checks the integrity of the interface between mating halves. It is an X-ray technique used to identify the elemental composition of materials. It investigates for evidence of corrosion, arcing, melting, and other failure indicators, including the indication of mismatch and the overall quantity and quality of the metal alloys. The technique includes the removal of the plastic sheath to expose the metal connector pins, followed by visual inspection for signs of damage and preparation of samples for microscopic analysis. Applications include material and product research troubleshooting, deformation, adhesion, bonding and delamination investigations, color issues, product imperfection, and defect analysis.

Cross-sectioning

It is done on the connector at the position of the crimp to evaluate connector pin crimp quality.

4 Review Results

This section presents the key findings of the various studies undertaken in Sect. 2. The core results cover the reliability and degradation of the connectors to develop a model to predict the connector lifetime. Fires due to failure of connectors are covered with a point to highlight that even if the initial capital cost of a connector is very low *visa-vi* compared to other components, it can create havoc in the long term on project sustenance if it is not installed or maintained correctly.

The reliability studies have been carried out in phases, and the research has been taken ahead in each phase. The outcomes of most of the studies have suggested that more research needs to be carried out on field-aged connectors. A comparative analysis of field-aged connectors must be conducted with the connectors on the accelerated stress testing to develop a correct reliability and degradation model. The various studies conducted across the globe reveal that the connector, though a minuscule part of PV plant plays a significant role in running the system to its optimum energy efficiency. If it is not maintained in the longer run, it can lead to faults and in extreme cases, a safety and fire hazard. Most of the fires caused due to connector failures have been due to electric arcing. The best way to detect this inside the connector is by using thermal imaging. Various studies globally have highlighted that the PV operator/installer is not aware of the exact causes of fires until a specialized team conducts the forensics, and most operators generally do not report the burning of PV connectors and replace them without going deep into the cause of the fire. However, studies have ascertained that to achieve a low PV levelized cost of energy (LCOE), testing before installation and an advanced monitoring system for early fault detection must be part of the PV initial investment.

Cross-mating of connectors is now a big issue in the solar PV industry, as there is no single general design of connector casing. PV connectors require seamless matching to the function for which it has been designed. If these do not match, it can cause a misalignment in design dimensions and material compatibilities over time, leading to mismatch losses with burnt connections, arcing, and fires. Moreover, most of the codes are standards for safety. However, they do not recommend the cross-mating of connectors; even the certification agencies like TUV and UL have also warned against the cross-mating of connectors. However, to obviate this, a performance criteria standard for PV connectors is recommended to be issued by specialized agencies.

5 Issues Observed with Connectors in Our Present Ongoing Field Study

5.1 Testing of PV Connectors in the Field

DC PV plug connectors are the essential parts of an installation about safety, yet they are often overlooked, compromising the array's performance and putting the system's safety at risk. The PV connectors' testing is primarily based on visual inspection and thermal infrared imaging. PV connectors that showed an increased temperature were brought to the laboratory for contact resistance measurement. Results are highlighted in the succeeding paragraphs.

Visual Inspection for PV plug connectors

Visual inspection of the plant is one of the essential activities in an audit, as it involves technical and non-technical parameter analyses of the solar PV plant. Visual inspection allows for identifying the defects visible by the naked eye due to either improper workmanship quality or non-adherence to standard technical norms or environmental factors. It finds the fundamental issues preventing a plant from functioning reliably and safely. The visual inspection is carried out per IEC 61215:2016 and the NREL visual inspection checklist [38]. Plant layout, module, SCB, inverter, and transformer technical specifications are checked according to the design. Soiling, shading, encapsulation discoloration, and corrosion of contacts and interconnectors are also inspected on modules along with PV plug connectors and DC cable layout. It also includes system functional testing, which provides for proper start-up and shutdown operation of the whole plant to check whether it operates under acceptable operational limits. Various scenarios of failures of PV connectors and their likely causes were analyzed during site visits.

During the site visits to different plants located in different regions of India, it was observed that no standard design is being followed on-site and, in the industry, as well for connectors. The connectors vary in size, shape, and material, as specified in the technical data sheet. These connectors (1000–1500 V) vary in IP rating, which varies from IP 65 to IP 68 from different manufacturers. The PV connector's insulation material (i.e., casing and back nut) is manufactured using thermoplastics but with various materials ranging from PC, PPE, PA, or PPO, as mentioned in Table 1. The samples of the connectors, which are field-aged between 2 and 10 years, have been collected from different regions of the country with varied environmental and weather conditions for further tests, which is the next phase of the study.

It is pertinent to mention that all these thermoplastics have been well-researched, and their thermal and chemical properties are well-defined. Which material is better suited for PV connectors in the harshest conditions (cold and most hot regions) is to be ascertained. We must specify the material suited for the region, so a standard design is essential. It was observed in the market analysis that several counterfeits are available in the market. Moreover, it has been observed on-site that there is a lack of standard signs and markings required in the PV connector, which is a variance from

Table 1 Different types of connectors installed at different sites

S. No	Make	Region	Location	Operational life (Yrs)	Type of plant (ground mounted/ rooftop)	Material of connector
1	Manufacturer 'A' (China) [39]	High altitude arid, cold region	Leh (Leh & Ladakh)	2–8	Ground mounted (100 kWp)	PC/PA
2	Manufacturer 'B' (Switzerland) [40]	Humid subtropical	Deogarh (Jharkhand)	2–10	Ground mounted (5 MWp)	PC/PA
3	Manufacturer 'B' (Switzerland) [40]	Tropical wet and dry	Pune (Maharashtra)	2–8	Ground mounted (2 MWp)	PC/PA
4	Manufacturer 'C' (India)	Tropical wet and dry	Pune (Maharashtra)	2–6	Rooftop (50 kWp)	PPO
5	Manufacturer 'C' (India) [41]	Subtropical humid	Dehradun (Uttarakhand)	2–6	Ground mounted (20 MWp)	PPO
6	Manufacturer 'D' (Japan) [42]	Hot semiarid and humid	Tirunelveli (Tamil Nadu)	2–8	Ground mounted solar park (100 MWp)	Not specified
7	Manufacturer 'B' (Switzerland) [40]	Hot semiarid and humid	Tirunelveli (Tamil Nadu)	2–6	Ground mounted solar park (100 MWp)	PC/PA
8	Manufacturer 'A' (China) [39]	Hot and semiarid	Delhi	2–6	Rooftop 100 kWp	PC/PA
9	Unknown	Hot and humid	Cochin (Kerala)	2	2 MWp Grnd Mtd	
10	Manufacturer 'E' (China) [43]	Hot and semiarid	Jaipur (Rajasthan)	2–6	2.5 MWp Grnd Mtd	PC/PPO
11	Manufacturer 'F' (China) [44]	Humid subtropical	Imphal (Manipur)	4	1 MWp Grnd Mtd	Not specified
12	Manufacturer 'G' (USA) [45]	Hot and humid	Chennai (Tamil Nadu)	4	1 MWp ground mounted	PPO
13	Manufacturer 'B' (Switzerland) [40]	Hot and semiarid desert	Jaisalmer (Rajasthan)	10	5 MW ground mounted	PC/PA

(continued)

Table 1 (continued)

S. No	Make	Region	Location	Operational life (Yrs)	Type of plant (ground mounted/ rooftop)	Material of connector
14	Manufacturer 'C' (India) [41]	Hot and semiarid desert	Jaisalmer (Rajasthan)	10	5 MW ground mounted	PPO
15	Manufacturer 'D' (Japan) [42]	Hot and semiarid desert	Jaisalmer (Rajasthan)	10	5 MW ground mounted	Not specified
16	Manufacturer 'H' (USA)	Hot and semiarid desert	Jaisalmer (Rajasthan)	10	5 MW ground mounted	PPE
17	Manufacturer 'J' (Germany) [46]	Hot and semiarid desert	Jaisalmer (Rajasthan)	10	5 MW ground mounted	Not specified

the electrical standards of DC connectors. Even connectors that claim to be certified and passed through testing procedures vary in price between Rs 40 and Rs 150. This price difference for a male and female plugs with no moving parts is enormous and impossible without compromising quality.

These PV plug connectors require constant maintenance to avoid getting jammed. A pull-out test is carried out to check their functionality. During one such visit to a 2.5 MWp grid-tied solar PV plant site in Pune, India, 63 PV plug connectors at modules, strings, and SCB were found to be jammed and could not be pulled out. Wires and connectors were seen hanging from the module mounting structure, leading to stress in the connectors. Some of the connectors at the string and SCB are loosely tightened and are prone to dust and moisture ingress. Wire and connector discipline at most sites have been poor, though it has not been caused during the installation but during the operation and maintenance activities. The main reasons are poor workmanship which causes dust and moisture ingress in the PV connector, as depicted in Fig. 4. It will lead to the development of mechanical stress at a later stage due to improper bending.

Thermal or Infrared Imaging (TI or IR)

Thermal imaging allows for analyzing the PV plant components without interrupting the power plant operation [46, 47]. Using thermal (or IR) imaging analysis, the direct approach behind fault identifications is that any error or loss in power plant components causes unusual temperature variations compared to normal operating conditions [48]. Thermal imaging is used to identify such hotspots in many modules without making direct contact due to soiling, shading, and micro-cracks. TI is carried out during the audit on modules and BoS, including PV connectors. As observed during thermal imaging, a temperature increase of more than 50 °C was detected on



Fig. 4 a Connector hanging on the MMS at Deogarh b connector bend and hanging on rooftop plant site at Delhi c loosely tighten PV plug connectors at SCB marked by an arrow at Pune d hanging connectors and birds nest on MMS e hanging of connectors and wires at Dehradun plant site f connectors bunched together at Chennai site

some PV connectors, which may lead to excessive heating and faults in the system. These PV plug connectors in the strings had hotspots and were heavily soiled. Those connectors which showed hotspots had also some hotspots seen in the solar cell of module as shown in Fig. 5. In extreme cases, if it is not maintained, it can lead to burning out of the connector and damaging the module, and causing a fire incident inside the plant.

Contact Resistance of PV Plug Connectors

The contact resistance of those 10 PV plug connector samples showing temperatures more than 50 °C through thermal imaging was measured in the laboratory. The connectors where vegetation growth had increased, considerable soiling and hotspots were observed, and connectors were loosely tightened due to improper workmanship and had high contact resistance of 0.8–1.9 mΩ beyond permissible limits of 0.5 mΩ.

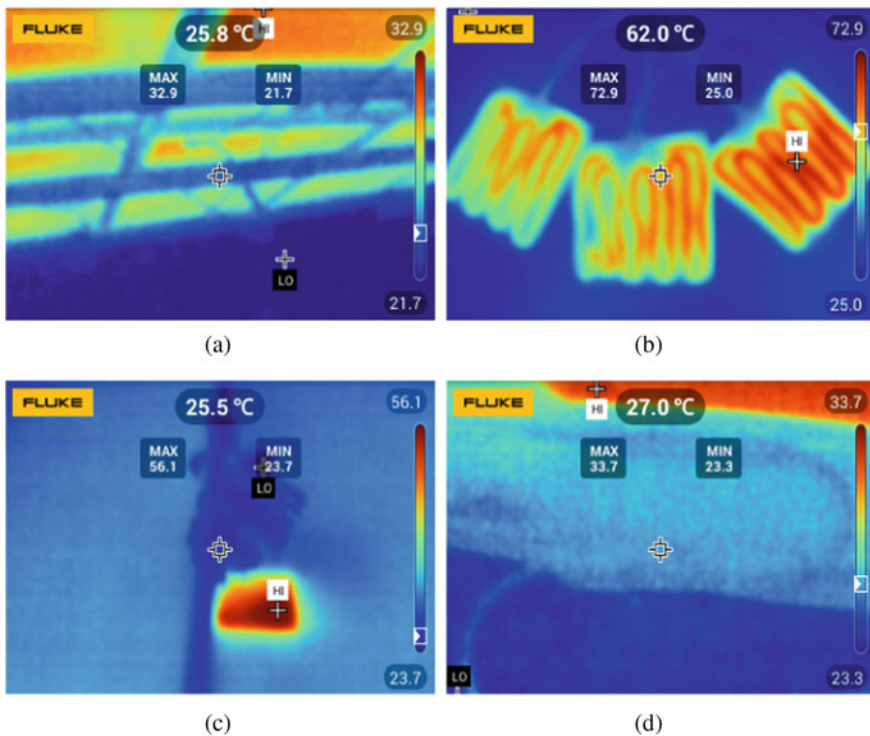


Fig. 5 a Infrared images showing connectors in connector hanging on the MMS b hot connectors which are bunched together forming hotspots c hotspots caused in module which had hotspots on connectors d hotspot on connector aligned with MMS

6 Limitations of the Present Studies

Studies based on accelerated stress testing under laboratory conditions may not be wholly reliable, as the connectors are exposed to different environmental, weather (temperature, humidity, dust, pollution, and UV exposure) and service conditions (bending and improper workmanship) in the field, which may lead to an early failure as supposed to be represented in a controlled environment. The IEC tests predominantly use the above approach. However, this approach's 85 °C/85%RH damp heat test condition is not very realistic regarding actual service. These approaches work best for mechanically related single-stress failures, and response is often nonlinear for chemically induced changes or where degradation is dependent on sequential or overlapping mechanisms. Moreover, weather and climate have multiple interrelated stressors varying continuously in short- and long-term patterns, which is extremely difficult to reproduce in test bed equipment. However, most studies are based on this accelerated stress testing for the aging of connectors to their warranted operational life of 25 years. Moreover, a study by NREL [49] also brings out that many failure mechanisms, as observed in the field, are not screened in current safety and design qualification tests, and there is no capability to project overall module field durability.

Studies on the new/standard design of the connector, which overcomes the issues of the present connector, have not been studied in detail, but most of the studies are based on the existing design and its reliability and degradation issues.

7 Recommendations

- At the field level, to negate the large-scale issues of PV connectors, various essential guidelines need to be issued to stakeholders, installers, and operator levels for specific training on the installation and maintenance of connectors.
- There is a dire need for a standard design for PV connectors prescribing the dimensions, tolerances, and material properties. Various products of different shapes and sizes are available in the market with various counterfeit connectors.
- Training on the operation and maintenance of the PV connectors should be carried out on priority. Being a minuscule part of the system, it is often neglected, and the ill practices of workmanship have been deeply ingrained, which needs to be wiped off by industry-driven training programs.
- The industry will only benefit from the improved tools and field test methods for connector inspection. With the proper perspective and technology, connector inspections can become a standard operating procedure for solar PV systems instead of a reactive approach now being followed for these components.
- There is no technical standard for the performance of connectors. It mainly concerns the safety aspects of connectors based on IEC, UL, EN, BIS codes, and NEC regulations. The academicians and industry should push for a performance

standard to reach better efficiency in the system. It will also negate counterfeits and low-quality products.

8 Conclusion

PV connectors are important subcomponents of a solar photovoltaic power plant, and it is not complicated when installed correctly and will function properly if the required quality connector is installed. However, when it fails, the ramifications are enormous. While routine field inspections can save costly fire incidents, the root causes of connector failure are fundamental and must be addressed on priority. The significant outcome of the studies that have been carried out is that reliability and degradation models in connectors should be based on field-aged connectors with a comparison of the existing combined accelerated stress testing aged samples to get more reliable and accurate results. However, significant fires can be avoided if proper installation practices, operation, and maintenance are followed along with an advanced monitoring system. A performance standard and general design to include a standard material for the casing is required for a connector to negate the various issues in a connector, including cross-mating and quality. In addition, training should be industry-driven at all levels to obviate and detect faults and installation issues early. In addition, the future scope of work includes testing of the above field aged samples collected from various regions of India with different climatic conditions to predict an accurate connector lifetime.

9 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Vyas K (2017) Managing solar cables and connectors: for safety and longevity of PV system. Akshay Urja
2. Aghaei M et al (2022) Review of degradation and failure phenomena in photovoltaic modules. *Renew Sustain Energy Rev* 159:112160. <https://doi.org/10.1016/j.rser.2022.112160>
3. Tevi GJ-P, Faye MÈ, Sene M, Faye I, Blieske U, Seidou Maiga A (2018) Solar photovoltaic panels failures causing power losses: a review. In: 7th international energy and sustainability conference (IESC). pp 1–9. <https://doi.org/10.1109/IESC.2018.8439986>
4. XHPV (2021) “Photovoltaic connector damage and burnout” ranks 2nd in the list of TOP20 technical failures in power plant operation and maintenance

5. Muntwyler U, Schüpbach E (2021) New findings on the PV fire prevention—firefighter strategy for in-Roof PV installations. In: 38th European photovoltaic solar energy conference and exhibition. pp 1275–1277
6. Baschel S, Koubli E, Roy J, Gottschalg R (2018) Impact of component reliability on large scale photovoltaic systems' performance. *Energies (Basel)* 11(6). <https://doi.org/10.3390/en11061579>
7. Photon International (2010) Good connections: market survey on PV connectors. Photon International
8. Kalejs J, Gadomski J, Nobel Z Connector issues in reliability
9. van de Wetering K et al (2012) Investigation of plastic materials for application as PV module connectors. In: 38th IEEE photovoltaic specialists conference. pp 2328–2330. <https://doi.org/10.1109/PVSC.2012.6318064>
10. Dhere NG, Kumar B, Hadagali VV, Pethe SA, Wohlgemuth J, Amin D (2006) PV connector performance in a hot and humid environment. In: IEEE 4th world conference on photovoltaic energy conference, vol 2. pp 2199–2201. <https://doi.org/10.1109/WCPEC.2006.279944>
11. Swingler J, McBride JW (2002) Fretting corrosion and the reliability of multi-contact connector terminals. *IEEE Trans Compon Packag Technol* 25(4):670–676. <https://doi.org/10.1109/TCAPT.2002.808007>
12. EN (2019) EN 62852-2015 connectors for DC-application in photovoltaic systems—safety requirements and tests
13. UL 6703 (2017) Connectors for use in photovoltaic systems. Northbrook, USA
14. Bureau of Indian Standards (2018) IS 16781-2018: connectors for d.c. application in photovoltaic systems—safety requirements and tests
15. Kalejs J (2014) Junction box wiring and connector durability issues in photovoltaic modules. In: *Proc SPIE*, vol 9179. <https://doi.org/10.1117/12.2063488>
16. Bahaj AS, James PAB, McBride JW (2003) Predicting photovoltaic connector lifetime. In: *Proceedings of the 3rd world conference on photovoltaic energy conversion*, vol 3. pp 2833–2836
17. Köntges M et al (2014) Review of failures of photovoltaic modules
18. Pandey S (2017) Evaluation of various PV module cable connectors and analysis of their compatibility. *Int J Curr Eng Technol* 7(5):1721–1727
19. McBride JW (2006) Developments in fretting studies applied to electrical contacts. In: *Electrical contacts—2006. Proceedings of the 52nd IEEE holm conference on electrical contacts*. pp 170–180. <https://doi.org/10.1109/HOLM.2006.284083>
20. Bahaj AB, James P, McBride J (2001) Photovoltaic connector behavior under accelerated fretting testing regimes. In: *Proceedings of the forth-seventh IEEE holm conference on electrical contacts (IEEE Cat. No. 01CH37192)*. pp 203–208. <https://doi.org/10.1109/HOLM.2001.953212>
21. Yang BB et al (2013) Reliability model development for photovoltaic connector lifetime prediction capabilities. In: *IEEE 39th photovoltaic specialists conference (PVSC)*. pp 139–144. <https://doi.org/10.1109/PVSC.2013.6744115>
22. Yang BB et al (2014) Arc fault risk assessment and degradation model development for photovoltaic connectors. In: *IEEE 40th photovoltaic specialist conference (PVSC)*. pp 3549–3555. <https://doi.org/10.1109/PVSC.2014.6924875>
23. Schindelholz E et al (2015) Characterisation of fire hazards of aged photovoltaic balance-of-systems connectors. In: *IEEE 42nd photovoltaic specialist conference (PVSC)*. pp 1–6. <https://doi.org/10.1109/PVSC.2015.7356425>
24. Miller D et al (2022) Evaluating the durability of balance of systems components using combined-accelerated stress testing. In: *IEEE 49th photovoltaics specialists conference (PVSC)*. pp 15–17. <https://doi.org/10.1109/PVSC48317.2022.9938581>
25. BRE National Solar Centre (2018) Fire and solar PV systems—investigations and evidence. Cornwall
26. Namikawa S (2017) Photovoltaics and firefighters' operations: best practices in selected countries: task 12. Japan

27. Japan Photovoltaic Energy Association (JPEA) (2019) Japanese government warns against fire risk from rooftop PV. Consumer Safety Investigation Commission
28. IEC (2017) IEC 60364-7-712:2017: low voltage electrical installations—part 7-712: requirements for special installations or locations—solar photovoltaic (PV) power supply systems
29. IEC standard (2016) IEC 62548:2016:photovoltaic (PV) arrays—design requirements
30. Herz M, Friesen G, Jahn U, Köntges M, Moser D (2022) Task 13 performance, operation and reliability of photovoltaic systems—quantification of technical risks in PV power systems
31. Brearley D (2020) PV connector mating and intermatability in NEC 2020. Mayfield Renewables
32. Staubli Electrical Connectors AG (2021) Cross-mating of PV connectors. Webpage
33. Grablutz C (2016) Top 5 points of failure on solar systems, and how to correct them. Solar Power World
34. Shellenberger G (2021) A guide to addressing fire risks in rooftop solar. PV Magazine
35. Stin Energy PV (2022) Three reasons for the failure of MC4 connectors
36. Karin T, Penalva D, Nagel J (2022) The ultimate safety guide for solar PV connectors
37. Scandia National Laboratories (2021) PV connectors
38. Packard CE, Wohlgemuth JH, Kurtz SR (2012) Development of a visual inspection data collection tool for evaluation of fielded PV module condition. United States. <https://doi.org/10.2172/1050110>
39. Elmex Elmex connector
40. Yukita Yukita connector YS 254/255
41. Zhonghaun Sunter Sunter connector PVZH201
42. Xinhui Xinhui connector PV HCA-30
43. JINKO Solar Jinko JK03Mxy-US connector
44. BIZLINK Bizlink F type connector
45. Weidmuller Weidmuller WM4C connector
46. Álvarez-Tey G, Jiménez-Castañeda R, Carpio J (2017) Analysis of the configuration and the location of thermographic equipment for the inspection in photovoltaic systems. *Infrared Phys Technol* 87:40–46. <https://doi.org/10.1016/j.infrared.2017.09.022>
47. Hong F, Song J, Meng H, Wang R, Fang F, Zhang G (2022) A novel framework on intelligent detection for module defects of PV plant combining the visible and infrared images. *Sol Energy* 236:406–416. <https://doi.org/10.1016/j.solener.2022.03.018>
48. Glavas H, Vukobratovic M, Primorac M, Mustran D (2017) Infrared thermography in the inspection of photovoltaic panels. In: International conference on smart systems and technologies (SST). pp 63–68. <https://doi.org/10.1109/SST.2017.8188671>
49. Kurtz S (2015) Why a qualification test can NOT be used for lifetime assessment and proposal for pathway to standard for lifetime assessment. CA, USA
50. Renhesolar Renhesolar ZJRH 05-6 connector
51. Staubli Staubli MC4 connector

A Comprehensive Study of Power Quality Improvement Techniques in Smart Grids with Renewable Energy Systems



Manpreet Singh and Lakhwinder Singh

Abstract The power quality problems are very important now-a-days in modern power electrification. As the transition to smart grids progresses in traditional electrical power grids, power quality issues are becoming increasingly significant. This paper presents a review of power quality improvement techniques often used in micro-grids. The most recent distributed flexible AC transmission system (D-FACTS) devices-based soft computing techniques are reviewed. The power quality issues that arise with the penetration of renewable energy sources are thoroughly analyzed, along with power quality (PQ) mitigation strategies including several D-FACTS devices and control algorithms such as artificially intelligence-based control algorithms and meta-heuristic optimization methods. For the benefit of engineers and academicians working in this field of study, 25 research publications have been carefully evaluated and organized for rapid reference.

Keywords FACTS technologies · Power quality · Renewable energy sources · Smart grid · Artificial intelligence techniques · Meta-heuristic optimization

1 Introduction

The load side harmonics produced by the power electronics devices perturb the supply current and cause it to wander from the fundamental signal. Large power quality (PQ) difficulties have been brought on by the large input of renewable sources into the power system and grid linked power electronics interfaces. The primary objectives of power quality improvement (PQI) devices are to stop harmonics from propagating to the grid, from being injected into a load, or from being compensated, mostly on the consumer side [1]. The majority of PQI equipment is multifunctional, meaning it can

M. Singh (✉) · L. Singh
Electrical Engineering Department, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib 140407, Punjab, India
e-mail: mnprtsingh95@gmail.com

M. Singh
Research Scholar IKGPTU, Jalandhar, Punjab, India

do many tasks simultaneously with the same hardware, increasing co-effectiveness in addition to being dependable and efficient. This category contains both active and passive power filters. The aforementioned problems may be greatly alleviated by IoT-inspired applications by enabling two-way communication protocols that help turn outdated power grids into contemporary smart grids (SG). The SG is made up of several power configurations, such as distributed generation (DG) or micro-grids (MG), which combine various energy sources with different demand characteristics [2]. The geographic location of the renewable energy production sites, which is typically in remote areas where surface conditions are favorable for solar and wind energy while the transmission systems are relatively weak, is another factor that restricts the integration of these energies. This poses new difficulties for controlling voltage and compensating reactive power. Power quality enhancement devices/techniques are divided into three categories [3]. First generation devices include passive filters, active filters, and hybrid filters. Second generation devices include static synchronous compensator (STATCOM), dynamic voltage restorer (DVR), static var compensator (SVC), and unified power quality conditioner (UPQC). Third generation devices include multifunctional distributed generation (MFDG), electrical springs (ES) and smart impedance, etc. The normal distribution of long PQ disturbances shows that disturbances lasting less than one second occur much more often than others [4].

Smart grids range from advanced AC/lighting systems to emergency appliances and connected devices with agile controllers. Power system operators often use custom power devices (CPDs) and retain some of their generation capacity as spinning reserve (SR) to maintain power levels [5]. These parameters include not only the physical layout of the unified power quality conditioner (UPQC), but also its control architecture and the characteristics of the supply chain to which it is connected. The battery energy storage system (BESS), transformer and inverter are connected to the DVR and installed in series with the load. These components balance active and reactive power requirements to reduce sags and surges. Flexibility that allows the distribution static synchronous compensator (DSTATCOM) network to improve power quality by changing the common point voltage. If it is planned to place the BESS in parallel with the DC bus capacitor, then DSTATCOM can exchange active and reactive power with the network [6].

Over the past decade, intelligence-based evolutionary algorithms have been developed for complex optimization problems, including genetic algorithms (GA), particle swarm optimization (PSO), differential evolution (DE), bacteria foraging (BF) and ant colony enhancement (ACO), artificial bee colonies (ABC), learning-based optimization, and recovery studies [7]. The aforementioned task is completed by utilizing the salp swarm optimization algorithm's (SSA) intelligence to determine the ideal combination of proportional plus integral (PI) controller parameters and the value of the dc-link side capacitance, which results in the least amount of settling time and overshoot under DG injection and load switching conditions. SSA is a more advanced approach for resolving various optimization issues when compared to genetic approach (GA) and particle swarm optimization (PSO) [8]. Singh et al. [9] implemented an 11-level cascaded multilevel inverter (CMLI)-based distribution static compensator (DSTATCOM) to mitigate the harmonic content of an induction

furnace in a steel plant. In another attempt, the comparison of two CPDs, UPQC, and DSTATCOM for harmonic mitigation in an induction furnace has been performed [10]. Only a few researches were devoted to compiling and assessing the features of these new UPQC models, despite the literature being updated with new models and control methods. Therefore, this study will provide an in-depth analysis of the UPQC patterns and the impact of these different structural configurations on the functionality of the device and associated distribution system. The most significant contribution of this review is to provide a review of the literature on the models, control algorithms, and performance of the D-FACTS devices now in use.

2 Power Quality and Facts Devices in Power Networks

Voltage, current, and/or frequency variations caused by power quality issues may lead to device failure or malfunction. Electrical systems must also maintain sinusoidal voltage with consistent amplitude and frequency so that its users are never without the power they need. Unwanted behavior, power outages, and interference with surrounding communication cables may all be brought on by poor power quality. The issue of electrical system quality has become increasingly complicated in modern times. As a consequence, PQ concerns might get greater attention from energy providers and end customers. It might be difficult to maintain network quality within acceptable bounds. The detrimental effect on quality has been well explored [11]. In the Springer, IEEE and Science Direct databases, the quality enhanced AC flexible transmissions search turned up 14,532 scholarly papers from 2015 to April 2023. The quantity of publications published globally between 2015 and April 2023 is shown in Fig. 1. Figure 2 shows the most well-known researchers and research facilities throughout the globe together with the number of publications each has received.

3 Power Quality Enhancement in Renewable Energy Sources Using Artificial Intelligence Techniques

It is widely believed that renewable energy sources (RES) will shortly displace traditional fossil fuel power production, giving rise to a novel idea of power grids driven by DG systems. However, there are still certain challenges to be solved, such as sporadic power production. Because of this, efforts are being made to develop additional technologies, such as energy storage systems (ESS), dependable power electronic devices, processing systems, and low-latency communications [12].

An intelligent fuzzy controller-based AC-DC micro-grid system has been shown by Nafeh et al. [13] to potentially enhance voltage stability and system power quality. These regulators include fuzzy-PI (FPI), fuzzy-PID (FPID), and distributed static

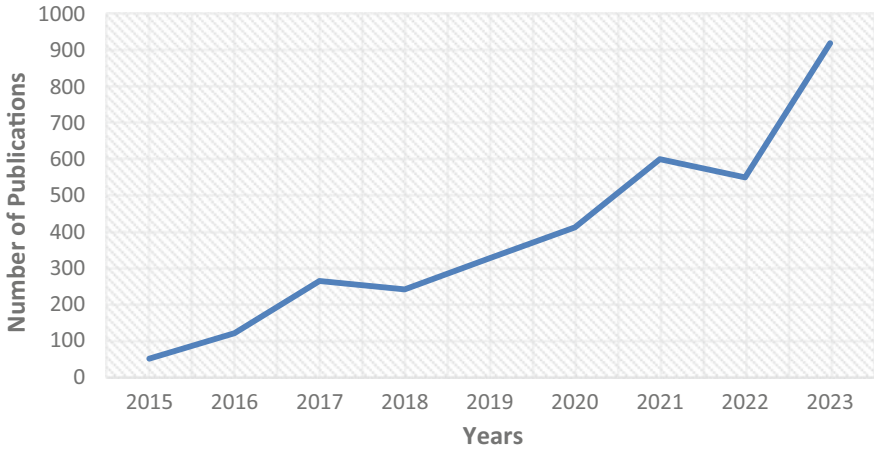


Fig. 1 Number of articles on FACTS devices from 2015 to 2023

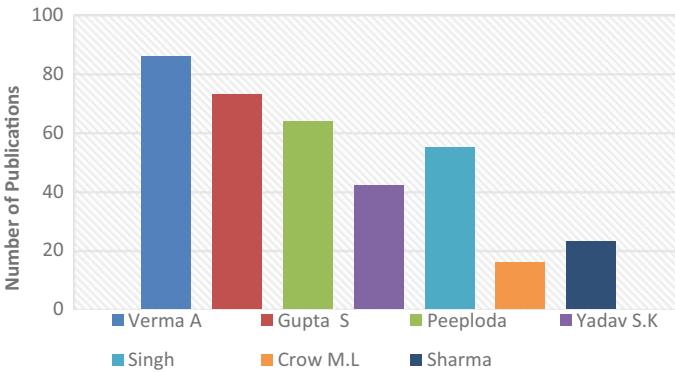


Fig. 2 Outstanding researcher in FACTS devices from 2015 to 2023

synchronous compensators (DSTATCOM) current regulators. Using the capabilities of the suggested system, two case studies simulate unforeseen interruptions and dynamic load changes in a hybrid AC-DC micro-grid including numerous renewable energy sources.

A grid-connected solar and wind photovoltaic system linked to an energy storage system (ESS) and an electric vehicle (EV) was reported by Sarita et al. [14]. Improvement methods include the unified power quality conditioner, generalized unified power flow controller (GUPFC), static var compensator (SVC), fuzzy logic controller unified power flow controller FLC-UPFC, and other AI-based strategies. FLC effectively regulates source-to-load and source-to-source power flow as well as peak and off-peak power usage.

Dheeban and Selvan [15] demonstrate how integrated solar UPQC in distribution systems might enhance power quality using an adaptive neuro-fuzzy inference system (ANFIS). Fuzzy model-based (FMB) controllers boost system performance and aid in generating reference currents by employing language principles to infer system characteristics. PV-UPQC operates well even under a variety of load circumstances. An adaptive neuro-fuzzy inference method reduces harmonic distortion overall by a certain amount.

Micro-grids may enhance power quality in linked electrical distribution networks, as shown by Renduchitala et al. [16]. Power quality problems at the sub-distribution level are resolved using unified power quality criteria. Harmonic currents and voltage imbalances are reduced. Additionally, a network-based adaptive fuzzy inference system that incorporates independent compensating devices in standard coupling stages offers a comprehensive solution for power control.

Grid integrated safety supervisory fuzzy controller (SSFC) was used by Kuchibhatla et al. [17] to build an adaptive approach for PQ analysis in renewable resources. Recurrent neural network (RNN) and cuttlefish algorithm (CFA) are both used in the adaptive approach. For SSFC devices, an adaptive approach is advised to regulate the switching of shunt capacitor banks on risky tuned arm filters. Specific suggested solutions improve demand management and energy efficiency operations for grid utilization. Here, effective power flow regulation with regard to network utilization is taken into account using a power injection model for SSFC.

4 Review Based on Optimization Algorithms

To improve the micro-grid's (MG) overall dynamic performance, Elmetwali et al. [18] provide a compensator with an adaptive switching filter (ASF) and a specially built proportional-integral (PID) controller. The grasshopper optimization algorithm (GOA) is used to establish the PID controller's optimal gain, allowing it to operate adaptively while the MG is running and adapt to changing operating conditions. Rajesh et al. [19] discuss the quality management strategy for RES in a micro-grid system. The recommended method is called the IBSMFO method because it combines an improved bat search (IBS) algorithm with a butterfly flame optimization methodology. In a manner similar to interbreeding, the mutation modifies the bats' dietary patterns. In this case, the moth flame optimization (MFO) algorithm has changed the improved bat search algorithm search procedure in an attempt to reduce the error function. Increasing the efficiency of the system's overall energy storage component is the recommended approach for improving performance.

A hybrid solar, photovoltaic, and wind micro-grid that uses STATCOM was constructed by Bakir et al. [20] to extend the system's stable operational range. The main contribution of this study is to use the genetic algorithm (GA) and bacteria set algorithm (BFA) to optimize the gain settings of the four PI controllers in the STATCOM loop, resulting in enhanced behavior and stability.

In order to improve the voltage profile and reduce harmonic distortion utilizing specific compensating devices in low inertia systems, Sindi et al. [21] developed a multi-micro-grid connection. An adaptive power quality compensator (APQC), which comprises of series and parallel compensators, is one kind of compensating device. The use of a thyristor series controlled capacitor (TCSC) series compensator lowers the transient voltage and dynamic voltage profile. In order to accurately decrease voltage and current harmonic errors, Bharathi and Selvaperumal [22] proposed a modified UPQC PI controller based on the grey wolf optimization (GWO) technique that is incorporated into renewable energy sources like squirrel-cage induction wind turbines (SCIWT). Modified grey wolf optimization (MGWO) was also utilized by UPQC.

To address PQ issues in hybrid renewable energy sources (HRES) systems, Goud et al. [23] proposed atomic search optimization (ASO) with a unified power quality conditioner (UPQC). The main goals of the work are to lessen the PQ problem and balance the load demand in the HRES system. The UPQC system utility is beneficial for the PQ problem. Reddy et al. [24] developed a modified elephant herd optimization (EHO) technique and implemented using a distributed power flow controller (DPFC) to optimize power quality in smart grids. Nonlinear loads are delivered to the system through PQ. On the basis of the recurrent neural network (RNN) algorithm, the EHO method has been modified. Power parameters for PQ disturbances are then recovered after first estimating the projected power flow in the system using the multi-wavelet transform (MWT). Mishra et al. [25] presented a hybrid active power shunt filter (HAPSF) and fractional order low integral derivative controller optimized using grey wolf particle shower hybrid optimization to account for reactive power and harmonics under balanced and unbalanced load conditions.

5 Performance Evaluation

A total of 13 research papers were analyzed as part of this survey. Each research paper uses its own FACTS tool and a different verification algorithm. In this survey, the methods used and to what extent they appreciate the improvement in smart grid's power quality using FACTS devices is analyzed. In the analysis of current research articles presented in Table 1, some methods have low accuracy and computational complexity, and some methods do not effectively highlight areas of power quality improvement. FACTS use several power quality improvements for RES devices. However, the speed of artificial intelligence methods and meta-heuristic approaches needs some improvement.

Figure 3a depicts the review of grid integrated renewable energy sources in power quality improvement. Figure 3b shows the improvement of power quality utilizing FACTS devices like UPQC, STATCOM, and other devices. Figure 3c presents power

Table 1 Overall analysis of survey

Author's name	Renewable energy sources	FACTS devices	Algorithms	Objectives
Nafeh et al. [13]	Wind, PV, fuel cell, and battery	DSTATCOM	Fuzzy-PI (FPI) and fuzzy-PID (FPID)	To maintain unity power factor according to the control loops
Sarita et al. [14]	Wind and PV	UPFC and UPQC	Artificial neural network (ANN)	To facilitate quality upgrades and THD reduction
Dheeban and Selvan [15]	PV	UPQC	Adaptive neuro-fuzzy inference system (ANFIS)	To improve power quality issues
Renduchintala et al. [16]	Wind and PV	UPQC	Adaptive neuro-fuzzy inference system (ANFIS)	To restore the voltage and lessen peak voltage distortions
Kuchibhatla et al. [17]	Wind and PV	SSFC	Recurrent neural network (RNN) algorithm	To improve power quality issues
Elmetwaly et al. [18]	Wind, PV, fuel cell, and battery	DSTATCOM	Grasshopper optimization algorithm (GOA)	To enhance the MGs' overall dynamic performance
Rajesh et al. [19]	PV, wind turbine, fuel cell, and battery	DPFC	Bat search algorithm and moth flame optimization	To improve power quality issues
Bakir et al. [20]	Solar-wind hybrid micro-grid	STATCOM	GA and BFA	To increase the reliability of operation
Sindi et al. [21]	Wind, PV, battery, and fuel cell	APQC and TCSC	Swarm intelligence-based puzzle optimization	To improve voltage profiles and reduce harmonic distortions
Bharathi and Selvaperumal [22]	Wind turbine	UPQC	Modified grey wolf optimization (GWO)	To eliminate the faults occurred during the transmission of power

(continued)

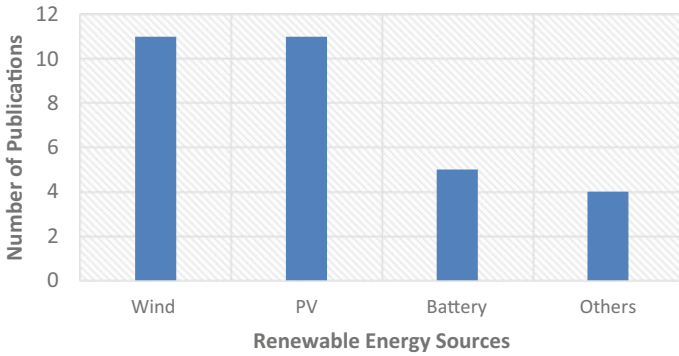
Table 1 (continued)

Author's name	Renewable energy sources	FACTS devices	Algorithms	Objectives
Goud and Rao [23]	PV, wind and battery energy storage system	UPQC	Atom search optimization (ASO)	To mitigate power quality issues and compensate load demand in HRES system
Reddy et al. [24]	Wind, PV	DPFC	EHO algorithm	To improve the power quality issues
Mishra et al. [25]	–	Hybrid shunt active power filter (HSAPF)	PSO-GWO	To decrease the dc-link voltage deviation

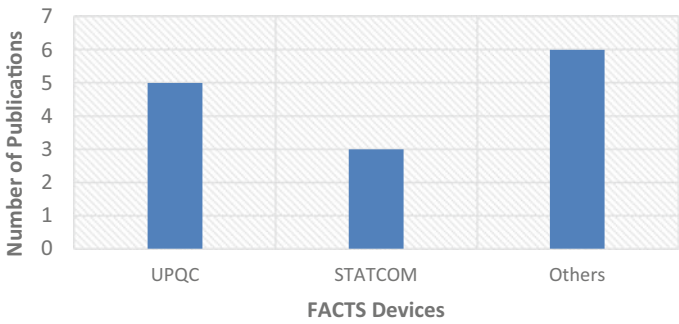
quality improvement in a micro-grid using FACTS devices based on different optimization algorithms. It is clear from Fig. 3c that the most commonly used are the meta-heuristic optimization techniques.

6 Conclusion

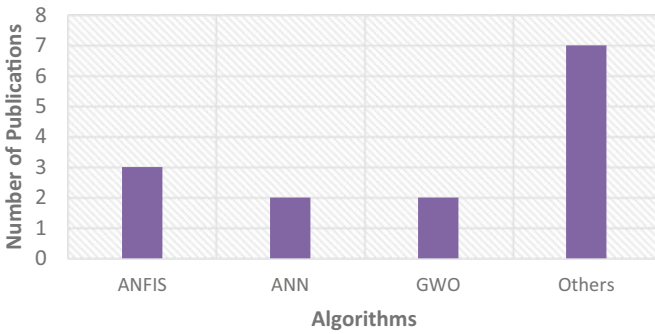
Power electronic converters are becoming more and more crucial for integrating vital technologies into smart grids, including electric rail systems, electric mobility, and renewable energy sources (RES). As a result, power quality becomes more important when creating the latest power electronics solutions to maintain power quality and solve problems. In this paper, power quality improvement techniques usually implemented in micro-grids has been reviewed. Various FACTS devices and optimization techniques used for improving power quality are critical analyzed.



(a)



(b)



(c)

Fig. 3 Review based on **a** renewable energy sources, **b** FACTS devices, and **c** algorithms

References

1. Das SR, Ray PK, Sahoo AK, Singh KK, Dhiman G, Singh A (2021) Artificial intelligence-based grid connected inverters for power quality improvement in smart grid applications. *Comput Electr Eng* 93:1–18
2. Alkahtani AA, Alfalahi STY, Athamneh AA, Shetwi AQ, Mansor MB, Hannan MA, Agelidis VG (2020) Power quality in microgrids including supraharmonics: issues, standards, and mitigations. *IEEE Access* 8:127104–127122
3. Martin VML., Azcondo FJ, Pigazo A (2018) Power quality enhancement in residential smart grids through power factor correction stages. *IEEE Trans Ind Electron* 65(11):8553–8564
4. Liu H, Hussain F, Shen Y, Arif S, Nazir A, Abubakar M (2018) Complex power quality disturbances classification via curvelet transform and deep learning. *Electr Power Syst Res* 163:1–9
5. Subudhi U, Dash S (2021) Detection and classification of power quality disturbances using GWOLM. *J Ind Inf Integr* 22:1–11
6. Wang Y, Wang Y, Huang Y, Yang J, Ma Y, Yu H, Zeng M, Zhang F, Zhang Y (2019) Operation optimization of regional integrated energy system based on the modeling of electricity-thermal-natural gas network. *Appl Energy* 251:1–27
7. Lewis M, Naughton JM, Dominguez CM, Todeschini G, Togneri M, Masters I, Allmark M, Stallard T, Neill S, Brown AG, Robins P (2019) Power variability of tidal-stream energy and implications for electricity supply. *Energy* 183:1061–1074
8. Bollen MH (2000) Understanding power quality problems: voltage sags and interruptions. *IEEE Press series on power engineering*
9. Saggiu TS, Singh L, Gill B (2017) Harmonics mitigation in a steel industry using 11-level cascaded multilevel inverter-based DSTATCOM. *Can J Electr Comput Eng* 40(2):110–115
10. Saggiu TS, Singh L, Gill B, Malik OP (2018) Effectiveness of UPQC in mitigating harmonics generated by an induction furnace. *Electr Power Compon Syst* 46(6):629–636
11. Buzau MM, Aguilera JT, Romero PC, Exposito AG (2019) Hybrid deep neural networks for detection of non-technical losses in electricity smart meters. *IEEE Trans Power Syst* 35(2):1254–1263
12. Sadeghzadeh M, Ahmadi MH, Kahani M, Sakhaeinia H, Chaji H, Chen L (2019) Smart modeling by using artificial intelligent techniques on thermal performance of flate-plate solar collector using nano-fluid. *Energy Sci Eng* 7(5):1649–1658
13. Nafeh AA, Heikal A, Sahiemy RA, Salem WAA (2022) Intelligent fuzzy-based controllers for voltage stability enhancement of AC-DC micro-grid with DSTATCOM. *Alexandria Eng J* 61(3), 2260–2293 (2022)
14. Sarita K, Kumar S, Vardhan ASS, Elavarasan RM, Saket RK, Shafiullah GM, Hossain E (2020) Power enhancement with grid stabilization of renewable energy-based generation system using UPQC-FLC-EVA technique. *IEEE Access* 8:207443–207464
15. Dheeban SS, Selvan NBM (2021) ANFIS-based power quality improvement by photovoltaic integrated UPQC at distribution system. *IETE J Res* 69(5):1–19
16. Renduchintala UK, Pang C, Tatikonda KM, Yang L (2021) ANFIS fuzzy logic based UPQC in interconnected microgrid distribution systems: modeling, simulation and implementation. *J Eng* 2021(1):6–18
17. Kuchibhatla SM, Padmavathi D, Rao RS (2020) Adaptive technique for PQ analysis in renewable sources with grid integrated SSFC. *J Ambient Intell Humaniz Comput* 11:2421–2434
18. Elmetwaly AH, Eldesouky AA, Sallam AA (2020) An adaptive D-FACTS for power quality enhancement in an isolated microgrid. *IEEE Access* 8:57923–57942
19. Rajesh P, Shajin FH, Umasankar L (2021) A novel control scheme for PV/WT/FC/battery to power quality enhancement in micro grid system: a hybrid technique. *Energy Sources, Part A* 1–17
20. Bakir H, Kulaksiz AA (2020) Modelling and voltage control of the solar-wind hybrid micro-grid with optimized STATCOM using GA and BFA. *Eng Sci Technol Int J* 23(3):576–584

21. Sindi HF, Alghamdi S, Rawa M, Omar AI, Elmetwaly AH (2022) Robust control of adaptive power quality compensator in multi-microgrids for power quality enhancement using puzzle optimization algorithm. *Ain Shams Eng J* 1–18
22. Bharathi SLK, Selvaperumal S (2020) MGWO-PI controller for enhanced power flow compensation using unified power quality conditioner in wind turbine squirrel cage induction generator. *Microprocess Microsys* 76:1–8
23. Goud BS, Rao BL (2021) Power quality enhancement in grid-connected PV/wind/battery using UPQC: atom search optimization. *J Electr Eng Technol* 16:821–835
24. Reddy PM, Reddy AS, Sujatha P (2020) A modified EHO algorithm utilized DPFC for power quality enhancement in smart grid. In: *Emerging trends in electrical, communications, and information technologies: proceedings of ICECIT*, vol 569. Springer, pp 321–334
25. 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

Fault Detection at PCC Using Wavelet Theory in Grid-Tied Solar PV Battery-Based AC Microgrid



Sarika S. Kanojia and Aagam Shah

Abstract In power systems, detecting faults and maintaining power quality are very important factors. Today's unpredictable load scenario necessitates the maintenance of voltage profile under fluctuating load conditions. This paper aims to improve and analyze voltage fluctuations power quality during faulty conditions with the aid of the wavelet transform method which accurately detects fault and the type of fault in the proposed AC–DC microgrid. For MPPT, this paper concentrates on advanced artificial intelligence algorithm called particle swarm optimization (PSO). MATLAB software is used to simulate the proposed model, the results have been displayed on the command window and through waveforms. In this system, fault identification is made quick and accurate using ideas from wavelet theory and signal processing. The performance analysis and dynamic modelling of a grid-tied 6.75 kW solar PV system has been done along with a solution to the issue of rapid fault detection at PCC. To increase the stability and dependability of the system, a battery energy storage system has been created and implemented. Using the MATLAB Simulink tool, different curves are obtained for use in result analysis. In this proposed system, different irradiance levels are used in accordance with our daily schedule so that results which are extremely comparable with or very close to reality have been observed.

Keywords Particle swarm optimization (PSO) · Microgrid · Voltage profile improvement · Fault detection · Wavelet theory · Solar PV · Battery energy storage system (BESS)

S. S. Kanojia (✉) · A. Shah
Electrical Engineering Department, Nirma University, Ahmedabad, India
e-mail: sarika.kanojia@nirmauni.ac.in

A. Shah
e-mail: 19bee109@nirmauni.ac.in

1 Introduction

Due to growing population, industrial development, and economic development, a massive increase in demand of electrical energy has been observed. The largest energy demand is yet to take place as due to integration of EV and other more industrial developments. Renewable energy sources such as PV, wind is becoming a more promising source for generation of electrical energy. As electrical grid is a network which has unlimited amount of energy supply, microgrids can be defined as small electrical grids, which can be operated in both modes, i.e. either grid-connected or islanded mode. At many places, grid-connected microgrid is already installed as due to more power cut-off experienced by them, thus initiating microgrid will lead to own generation and own utilization, thus reducing dependency although utilizing free available power, i.e. renewable power. Battery storage unit is necessary for the entire system, as it can act as power backup and also it can be used for smoothing of PV output power fluctuations. Battery bank stores power produced by solar panels and whenever the generation isn't available or not sufficient, the battery supplies power to load or grid. The MPPT algorithm supports sustainable efficiency by dynamically adjusting the voltage to ensure power optimization [1]. In this proposed model, MPPT uses AI methods to get accurate results with efficient use of energy. Particle swarm optimization, i.e. (PSO) method has been used which is a very accurate and advanced method whose algorithm mainly focuses on optimizing the solution by performing iterations. There is a massive change in voltage and current levels at the time of fault, thus fault detection is necessary. As fault detection should be quick and accurate, this paper concentrates on a method known as wavelet transformation for fault detection. Weather conditions are responsible for PV generation, thus an MPPT is a must need so as to give maximum power available at output side. The next sections of the paper are as follows. Section 2 describes configuration of AC microgrid based on its principal operation including fault detection, fault classification, i.e. the whole unit's description. Section 3 shows the system configuration. Section 4 describes the modelling of battery energy storage system. Section 5 comprises of modelling of inverters and filters, types of loads used and its requirement, and the analysis of results with respect to change in load and also about the fault detection method, i.e. wavelet transform. Section 6 comprises of modelling and designing of system with the help of equations, used for calculations. Section 7 comprises of calculations and equations used for simulation and Sect. 8 comprises of simulation results of proposed system with necessary waveforms. Section 9 concludes the paper followed by Sect. 10 which depicts the future scope of the proposed system.

2 System Overview

The AC–DC microgrid model has been designed and simulated in MATLAB Simulink where solar PV generation is fed to PSO-based MPPT. MPPT, is generally included for extracting maximum power available from PV module under certain conditions [2]. The voltage at which PV module produces maximum power is called maximum power point. Through the MPPT, DC–DC boost converter is connected which boosts the voltage level according to our requirement and output available from MPPT or PV module. Irradiance is the most dominant or influencing factor for the output of any PV module. In this system a signal generator is programmed in such a way that it starts from zero irradiance where battery and grid collectively supply the loads connected in microgrid, and attain peak and then gradually reduce to zero, so as to analyze the output in every condition. The temperature is considered constant at 25 °C, as the general analysis is considered by keeping temperature constant, at nominal temperature, i.e. 25 °C. The power post boost converter is available at a common DC link or DC bus, which connects thoroughly to BESS and inverter. A DC load of 1000 is also attached to same DC bus which further directly connects to battery energy storage system via a DC–DC bi-directional converter which acts as boost while discharging of battery and acts as buck while charging the battery. The calculations of DC–DC buck-boost converter and the boost converter post MPPT are carried out with accordance to load requirements and considering tolerance limits. An inverter is connected to the same DC bus in order to convert the generated DC power into equivalent AC form. The power after being converted in AC form is fed to AC load, passing through the filters, whose primary task is to filter the ripples or harmonics. The loads in proposed model include a purely resistive load and an induction motor load. At PCC along with the load connected, fault detector is also connected. As similar to grid, fault arises frequently, which leads to development of instability and disturbances in grid. In this proposed system, fault detector and classifier are introduced. This system is connected to utility grid with breaker in between, for the purpose of islanding the microgrid. The block diagram of the proposed AC/DC microgrid is shown in Fig. 1.

3 System Configuration

A P–N junction fabricated in layer of a semiconductor, forms a complete PV cell structure. Ideal solar cell's structure has a semiconductor diode connected in parallel to a current source with having a series resistance and parallel resistance [3]. In this system, 10 modules per string are connected in series and three strings in parallel. The parameters of the solar module under study are given in Table 1.

The dynamic model of PV array is shown in Fig. 2. The I–V curve and P–V curves of the panel under different irradiance levels are presented in Fig. 3. The PSO-based MPPT controller tracks output power of the PV array in real time. The

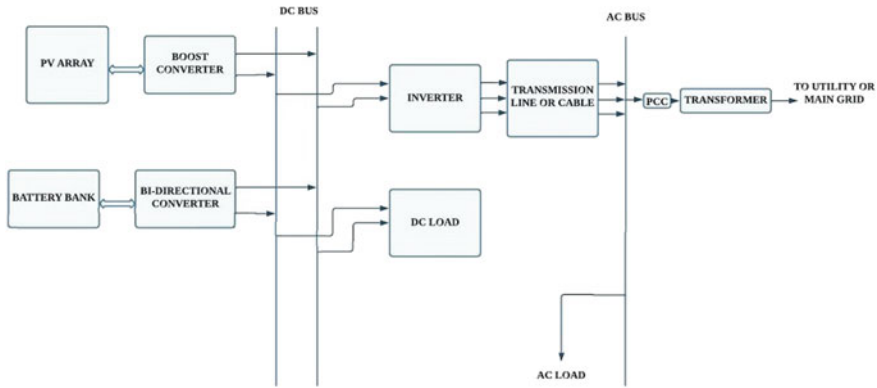


Fig. 1 Proposed AC/DC microgrid system

Table 1 PV array specifications

Parameters	Value
Maximum power	225 W
Cells per module	60
Maximum power voltage	30 V
Maximum power current	7.5 A
Open circuit voltage	73 V
Short circuit current	8.17 A

maximum power point (MPP) is dependent on ambient temperature and insolation shifts [4]. Particle swarm optimization is basically a computational method which optimizes the problem by iteratively trying to improve solution with respect to given measure. Also, with the help of references [5], algorithm based on equations for values of components, i.e. (inductor and capacitor) is used, to calculate the voltage and current of the PV array output. Among various algorithms for solar MPPT, in this paper MPPT has been designed with the help of PSO algorithm. This theory basically works according to the flowchart shown in Fig. 5. The main and greatest advantage of using this method is that there is a huge reduction in oscillation once the steady state is achieved (practically near to zero). Also, it can work under worst environmental conditions, i.e. large fluctuations of insolation and partial shading conditions. As the algorithm is quite simple, thus can be computed very easily and rapidly thus leading to less costly micro-controllers’ involvement [6]. It regulates the PV voltage base on the power delivered. Considering the irradiance levels for a particular day, it increases as approaching to afternoon and then it reduces as evening is approached, thus in this proposed model the irradiance levels also change according to real-time data. The change in irradiance, from 0 to 1000 W/m² and reverse, i.e. 1000 to 0 W/m², is created so as to analyze the performance and to get an analysis very close to real-time results. The irradiance level changes according to waveform shown in Fig. 6.

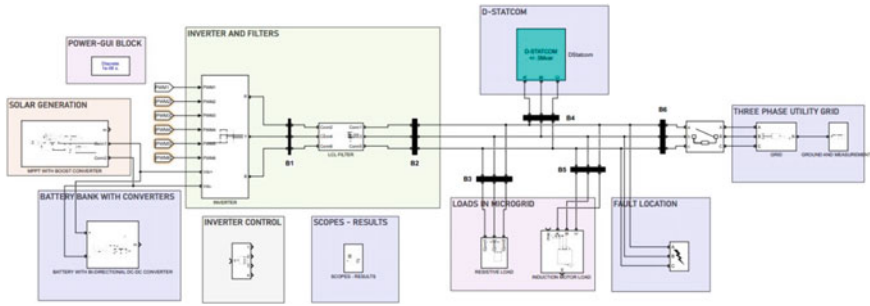


Fig. 2 Detailed schematic of AC/DC microgrid system through MATLAB Simulink

Considering real-time data, the signals shown in Fig. 6 have been fed to the boost converter in order to regulate voltage and get output according to load requirement. The output voltage waveform at the common DC bus obtained is shown in Fig. 8. The output directly is obtained to the DC bus-common link and through it a load of 1000Ω is also connected. The calculations and design parameters of boost converter are described in Sect. 7.

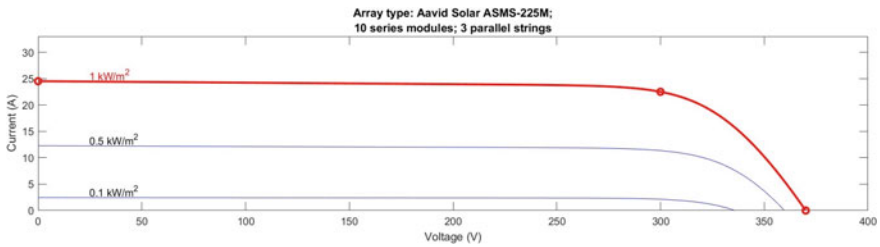


Fig. 3 Dynamic model of PV array

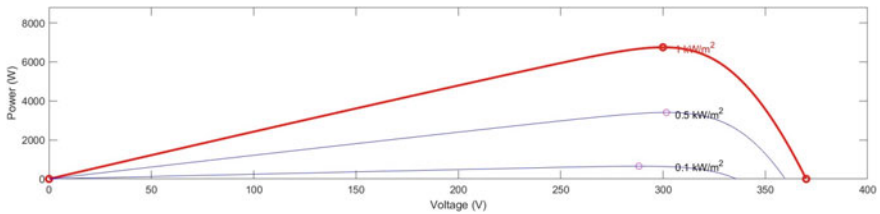


Fig. 4 P-V curve of solar panel

Fig. 5 Working of PSO-based MPPT [7]

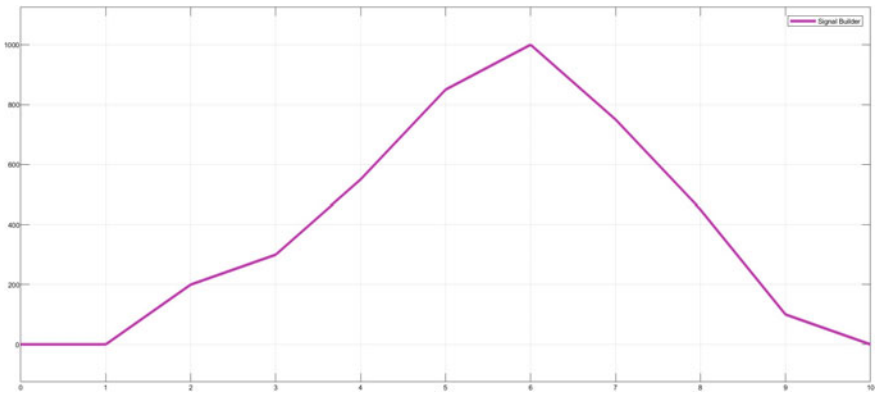
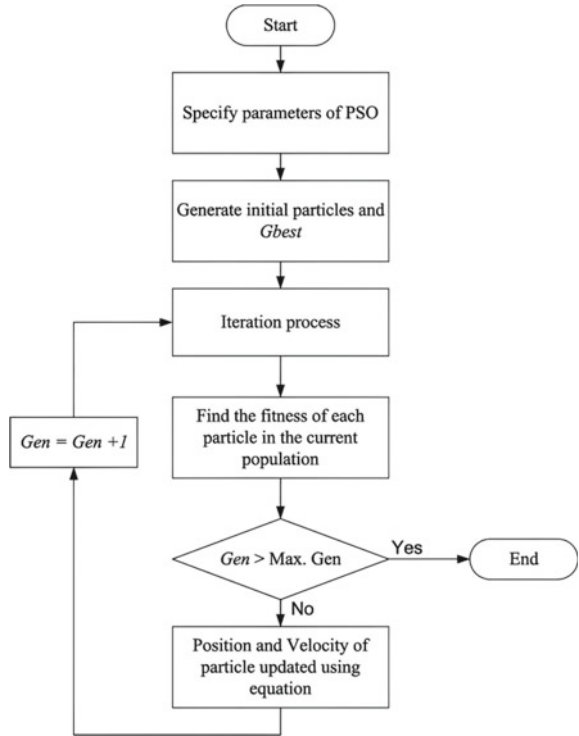
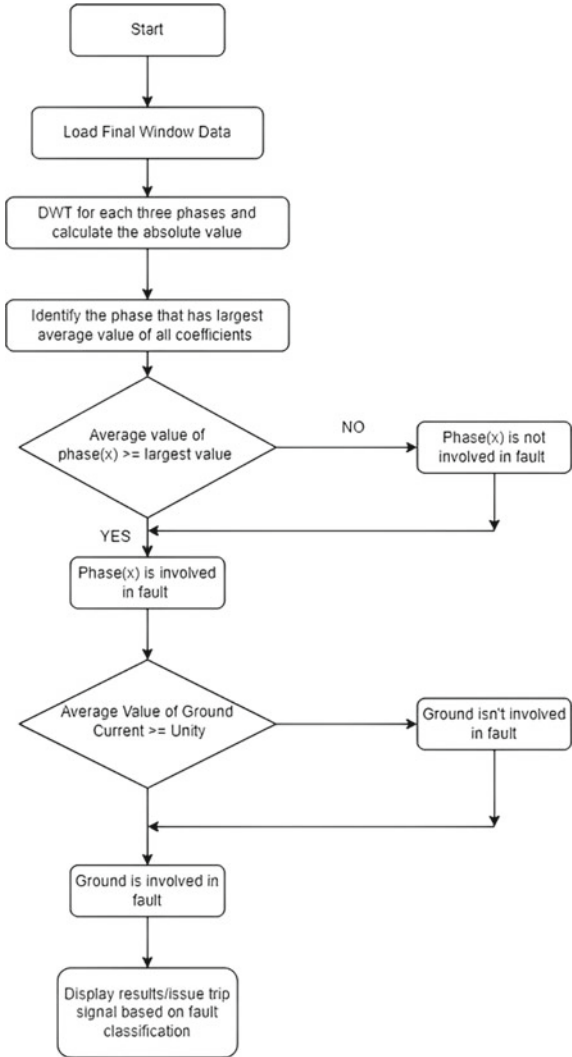


Fig. 6 Irradiance level to solar panels

Fig. 7 Flowchart of wavelet transform method for fault detection



4 Modelling of BESS

The battery’s capacity and lifespan are influenced by a number of factors which includes temperature, depth of discharge (DOD), and discharging rate. Due to their durable lifecycle properties and proven safety record, lithium-ion phosphate batteries have been employed in this application. Table 2 displays the battery specifications. Peak shifting, peak smoothing, and peak shaving are some of the major problems which can be solved with the help of BESS. The generated solar power may lead to fluctuations due to transient cloud shadows on the PV array, which is smoothed

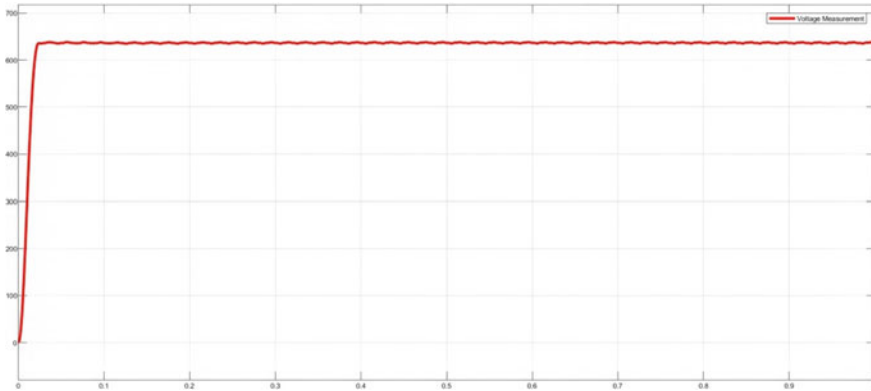


Fig. 8 Voltage of DC common link/DC bus

using a technique called smoothing. This small-scale BEES-based smoothing is implemented using a moving average configuration [8, 9]. Mostly statistical softwares are used which include basic statistics, measurement analysis, regression, and graphics. In this system the battery SOC value is predetermined and according to the SOC of the battery, the power flow from PV will take place. As power flow from PV to grid or to use for charging the battery are determined by the SOC value, the excess power which remains is used in charging the battery if SOC is lower than the predetermined value and PV has ample power generation available. Also, if further the excess power remains, then the breaker is closed and the power is supplied to grid. On the contrary if SOC indicates the fully charged condition then the battery along with PV will supply to microgrid and the remaining or excess will be fed to utility grid.

Thus control system has been set according to the SOC level of battery and the PV generation, thus if any exceeds the limit or undergoes limit then battery will turn on or will be switched off and will charge. And if SOC is also low and generation is also not sufficient, then the grid will supply to microgrid and also to the battery for charging. In the proposed system a bi-directional DC–DC converter is connected before battery bank which will act as buck-boost converter. While battery is discharging the converter acts as boost converter, and when the battery is being charged the converter will act as buck converter. The calculations of parameters are

Table 2 Battery specifications

Battery specifications	Value
Battery voltage	415 V
Rated capacity	20 Ah
Initial SOC	81
Battery response time	0.0001

described in Sect. 7. An example of charging is shown through SOC curve of the battery during the simulation in Fig. 8.

5 Modelling at Point of Common Coupling (PCC)

At distribution side many electrical components, such as three-phase inverter, two loads (one fixed and other variable), LCL filters, and fault analyzers are connected and then connecting it with the main 3- ϕ utility grid via circuit breaker.

5.1 Inverter and Filters

In this system 2-level, 3- ϕ inverter has been used, which converts solar panel-generated direct current (DC) to alternating current (AC) form. It does not produce any power, it just transfers power from DC source. They are mainly used in electrical power applications where high currents and voltages are present. The switches used here are IGBT as here the operating frequency is 5 kHz. If the frequency is increased beyond 5 kHz, MOSFET is generally used due to having lesser turn-off time (in nanoseconds). P-Q theory is used in this model for inverter control. Voltage has been taken as reference and after comparing with tuned results with the help of PI controller, it's converted into abc form and then it is further converted to α - β form. This α - β form is further converted d-q form and then further after passing through PI controllers, PWM pulses are generated with the help of relational operator. These PWM pulses are fed to all six switches of inverter. The signals generated through inverter control mechanism are compared with reference voltage and error generated leads to generation of a carrier signal with specific duration between switches. The waveforms are as shown in Figs. 10 and 11. Filter is a circuit capable of passing certain frequencies while attenuating other frequencies. The terms "low" and "high" are relative values with respect to the cut-off frequency. Passive filters use passive components, i.e. R, L, and C. Active filters use active components, i.e. op-amps, in addition to R and C, but not L. Filters are used at the output terminals of inverter. In this proposed model LCL filter is used, as it reduces the harmonics of current. They are made of a parallel-series combination of reactors and capacitors adapted to reduce the total harmonic distortion (THD) of power electronic converters. The filter design is explained in Sect. 7.

5.2 Significance of Fault Detection Analyzer

Fault is created at PCC, so as to analyze the waveforms at particular buses at the time of faulty conditions. This paper also focuses on the method of fault detection

and accuracy of the method used, so as to determine the type of fault and the time within which the fault is indicated. Also, the study of waveforms is also carried out and its analysis and results are illustrated in later sections.

5.3 Significance of Grid-Connected Mode

In this proposed system, a purely resistive load is connected and along with that an induction motor has been connected. The induction motor is of 5 HP rating 415 V, 1500 rpm, and the purely resistive load is of 5 k Ω . Induction motor load helps in analyzing the reactive power characteristics so as to analyze the control of voltage stability in the system. As according to its nature, it can run at no-load, rated load, and over-load, also it demands high starting current, thus all these conditions must be satisfied properly and the voltage stability should be properly observed as it's the main criteria of this proposed system. And finally, after connecting it through breaker the lines are directly connected to main grid. The location of coupling or connection of grid with microgrid is termed as point of common coupling (PCC). The waveforms at the load side of both the load buses B3 and B5 are shown in Figs. 14 and 15, respectively.

6 Wavelet Transform Method for Fault Detection in Power System

In this proposed system, wavelet transform method is applied to detect fault and also the type of fault occurred. The wavelet transform is a signal processing algorithm which is useful in detection of abnormal operating conditions based on decomposition of the power signals into different ranges of frequencies by the help of a series of low-pass and high-pass filters. This usually provides us a time–frequency multi-resolution analysis that is greatly useful for identifying any sort of abrupt variations in the electrical parameters such as voltage, phase, current, frequency, etc. [10] Power system transients consist of power quality disturbances, faults in power system like temporary over-voltages, and temporary over-current for short durations (some seconds).

The wavelet is a short duration oscillatory waveform with a zero average value which decays very fast to zero amplitude. The wavelet transform syntax to apply on signal is as follows:

$$[c, 1] = \text{wavedec}(x, n, \text{wname});$$

wavedec function which decomposes the signal,
x signal,

- n wavelet layer (default = 1),
- wname name of wavelet type (such as haar, Daubechies, etc.),
- c output wavelet decomposition vector,
- l number of coefficient by level.

On the basis of this, observations are conducted and every type of fault is induced and the values and parameters are recorded, keeping that record and values of current in consideration, a table is made of values of different types of fault current values and the values of other conductor/lines. After recording all the parameters with every type of faulty condition, all records are loaded in a tabular form, the values are carefully observed and a particular threshold value is decided (for this system we have got the threshold value as 0.5) and that value is used in the code used by this method. This value is just a threshold to all the variables used by this method, i.e. (m, n, p, q). Thus whenever any of the four variables breaches this threshold value, the fault is detected and the alarm operates and the detected fault can be observed on the command window with the type of fault occurred (command window as we are using MATLAB Simulink for this particular model). The fault detection will be more clear from the results in Table 3 (this data presented is real-time data of the proposed model). The changes in waveform due to fault detection will be clearly observed in Sect. 8, Fig. 12. Here the working of this fault detection method and the algorithm used for code is presented using a flowchart as shown in Fig. 7.

7 Calculations

7.1 Bus Voltage Calculations

Inverter output or grid voltage = 400–415 V
 Thus, inverter input should be:
 DC to AC conversion = $(1.5-2) * 400 \text{ V} = 600-800 \text{ V}$
 Taking = 630–640 V.

7.2 Solar Ratings

Generated power = 6750 W = 6.75 kW
 Generated voltage = 300 V
 Generated current = 22.5 A
 We need to boost voltage from 300 to 635 V
 Switching frequency = 5 kHz.

Table 3 Fault detection using wavelet transform

Type of fault	Maximum coefficient of phase—A current	Maximum coefficient of phase—B current	Maximum coefficient of phase—C current	Maximum coefficient of ground current
Three-phase to ground fault	21.7924	6.8075	50.6572	11.7638
Three-phase fault	21.7928	18.7267	50.6572	3.62E-05
Double line to ground fault (AB-G)	2.2638	6.94	0.0041	49.962
Double line to ground fault (AC-G)	19.2849	0.0025	50.3094	9.6321
Double line to ground fault (BC-G)	8.43E-04	6.8077	50.6573	2.2893
Line to line (A-B) fault	25.3467	3.3994	0.0041	7.50E-05
Line to line (A-C) fault	6.0719	0.0025	25.3195	1.84E-06
Line to line (B-C) fault	7.33E-04	19.1865	50.6663	3.62E-05
Single line to ground fault (A-G)	0.5908	0.0025	0.0041	2.2403
Single line to ground fault (B-G)	6.61E-04	6.945	0.0041	50.3256
Single line to ground fault (C-G)	6.61E-04	0.0025	50.3072	6.9534
System without fault	6.61E-04	0.0025	0.0041	1.41E-14

7.3 Design of Boost Converter

Taking $V_{out} = 800$ V, considering factor of safety

$$D = 1 - \frac{V_{in}}{V_{out}} \quad (1)$$

$$= 1 - \frac{300}{800} = 0.625 \quad (2)$$

where $D =$ duty cycle

$$D_{\min} = 1 - \frac{V_{OC}}{V_O} \tag{3}$$

$$= 1 - \frac{370}{800} = 0.5375 \tag{4}$$

ΔI = current ripple = 5% of input = $0.05 * 22.5 = 1.125\Delta V$ = voltage ripple = 1.1% of input = $0.1 * 800 = 8$.

Inductor and Capacitor Calculations:

Generator side connected capacitor is a ceramic capacitor

$$C = 1 * e^{-6} \tag{5}$$

Load side connected capacitor

$$C = \frac{I_O * D}{\Delta V_O * F_{sw}} \tag{6}$$

$$\frac{8.43 * 0.0625}{8 * 5000} = 0.000131 \text{ F} \tag{7}$$

Inductor's value

$$L = \frac{V_{in \max} * D_{\min}}{\Delta I * F_{sw}} \tag{8}$$

$$\frac{370 * 0.5375}{1.125 * 5000} = 0.035 \text{ H} \tag{9}$$

7.4 Design of Bi-directional DC-DC Converter

Capacitor's value

$$C = \frac{I_O * D}{\Delta V_O * F_{sw}} \tag{10}$$

$$= \frac{8.43 * \left(1 - \frac{42}{800}\right)}{80 * 5000} = 10 \mu\text{F} \tag{11}$$

Inductor's value

$$L = \frac{V_{in} * (V_{out} - V_{in})}{\Delta f * F_{sw} * V_{out}} \tag{12}$$

$$\frac{420 * (800 - 420)}{1.6 * 5000 * 800} = 0.021 \text{ H} \tag{13}$$

8 Simulation Waveforms and Results

In Fig. 8 the voltage at DC bus or common link is shown which is accurately maintained, constant at 635 V. As expected and calculated, it should be between 1.5 and 2 times of inverter voltage, thus it's 1.53 times.

The SOC of the battery seems to be increasing, thus it can be identified as charging of battery is taking place. If constant decrement is observed, then it can be identified that discharging is taking place. As observed from waveform that SOC took a spike in discharging mode first and then it again started charging, as generally every battery initially takes this spike, similar case is here too, but then the slope is maintained constant which can be analyzed as smooth charging (Fig. 9).

In Figs. 10 and 11, the waveforms of inverter voltage and inverter current are presented which are smooth sinusoidal and have values same as analytical calculation. Figure 11 depicts the change in irradiance level from 1000 to 500 W/m², an instant change was made just for the purpose of analyzing the change. As irradiance is directly proportional the instant change effect can be seen on distribution side too. Thus as irradiance decreases to half, the waveform values also change to lesser value as expected and calculated.

Figure 12 represents the changes in waveforms during fault. (L-L-G) fault, i.e. double line to ground fault has been created. The voltage of those two buses, i.e. B3 and B5, changes and the current of the two phases increase in the range of thousands, whereas the third phase in which the fault is not present, is working properly and its current is as according to Fig. 8, but due to lesser value it seems to be zero (Fig. 13).

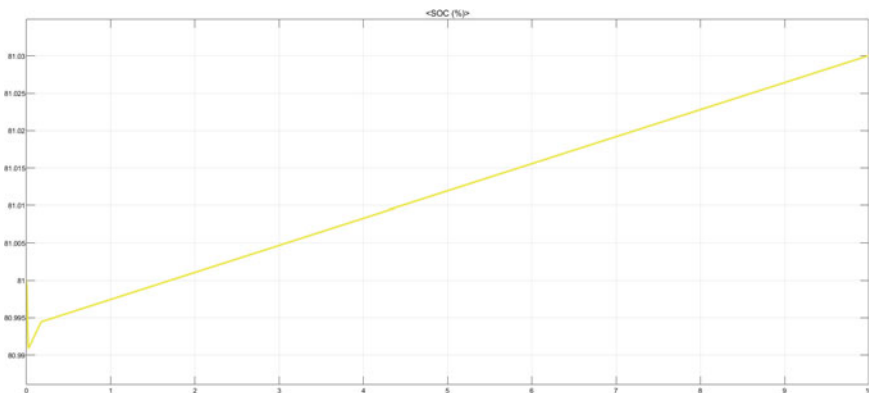


Fig. 9 SOC curve of the battery

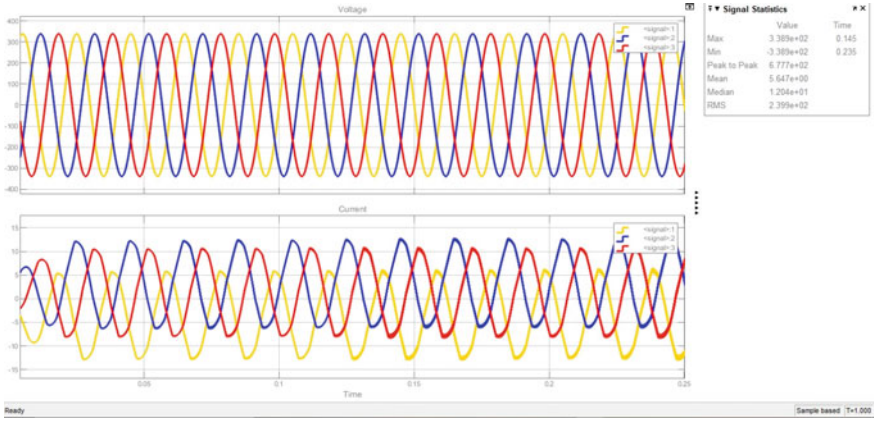


Fig. 10 Inverter waveforms in normal operating conditions

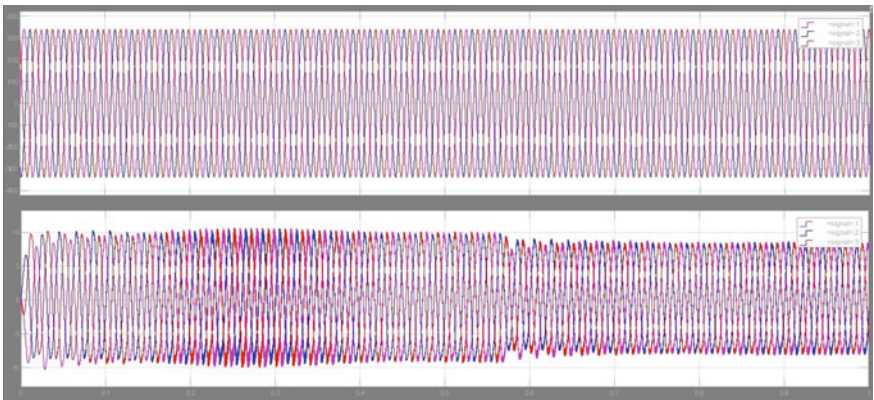


Fig. 11 Inverter waveforms due to reduction in irradiation

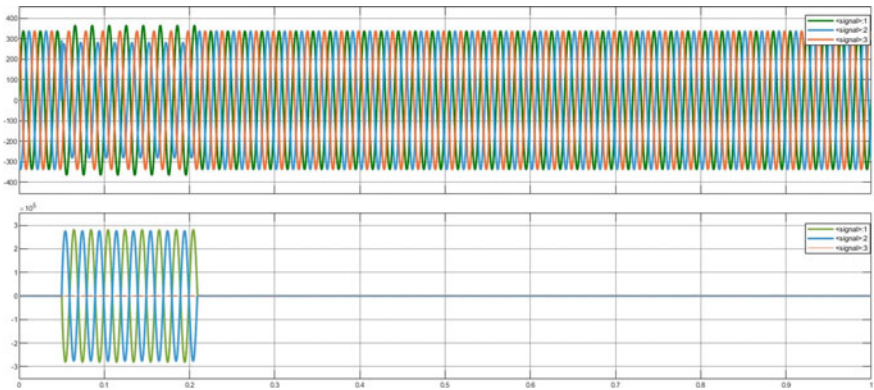


Fig. 12 Waveforms while subjected to double line to ground fault

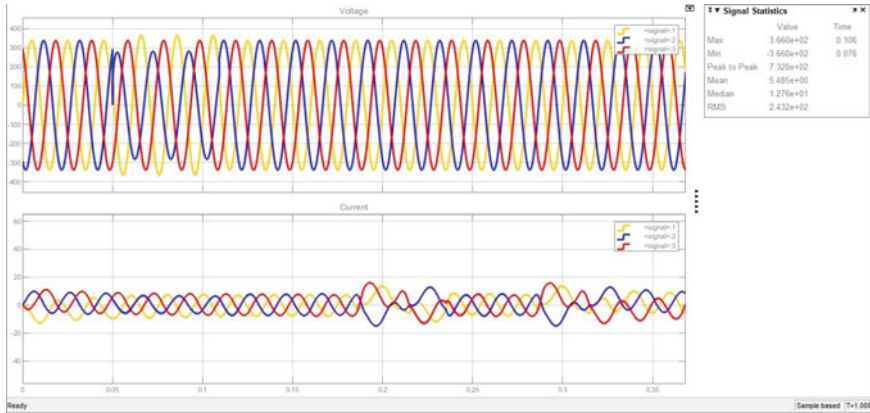


Fig. 13 Waveforms at D-STATCOM bus, i.e. B4 bus

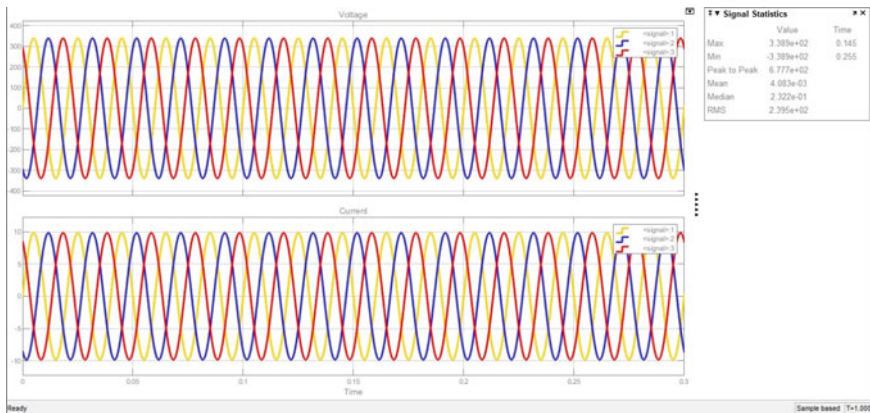


Fig. 14 Waveforms at resistive load bus (B3) without subjected to fault

The results at buses B3 and B5, i.e. both resistive and induction motor current and voltage curves are obtained as shown in Figs. 14 and 15. In these figures the waveforms are not subjected to any fault, whereas in Figs. 16 and 17 the waveforms shown are subjected to L-L-G fault, and the results at both the load buses (B3 and B5) are illustrated. Being an induction motor, high starting current is taken in order to establish flux, which can be observed through Fig. 15.

Fault detection through wavelet transform method for the proposed system is shown in Table 3. From the results it can be seen that as the values of parameters are greater than 0.5, than the limit is breached and the fault is observed. The four variables are from the wavelet transform method. And if no fault is present, then all the four values will be less than 0.5 and thus similar condition has also been discussed with the help of table. Here threshold for fault detection is kept as 0.5,

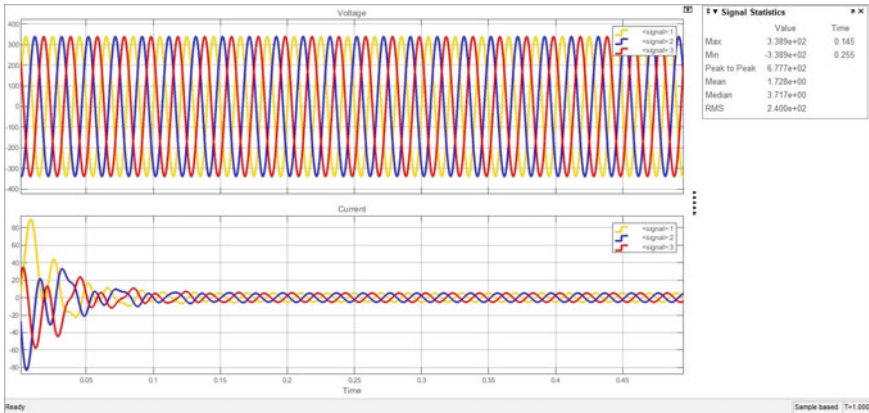


Fig. 15 Waveforms at induction motor bus (B5) without subjected to fault

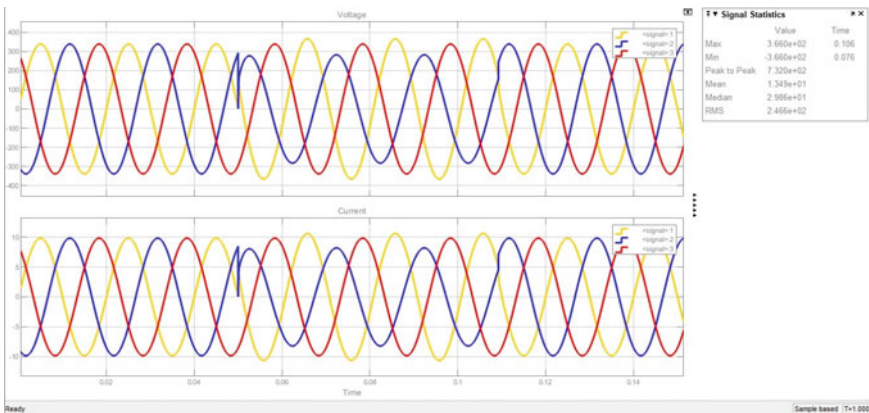


Fig. 16 Waveforms at induction motor bus (B5) after subjected to fault

thus if it's greater than the fault is present, after connecting the whole model with all discussed components and circuits and algorithms, the final MATLAB Simulink model is obtained as shown in Fig. 2.

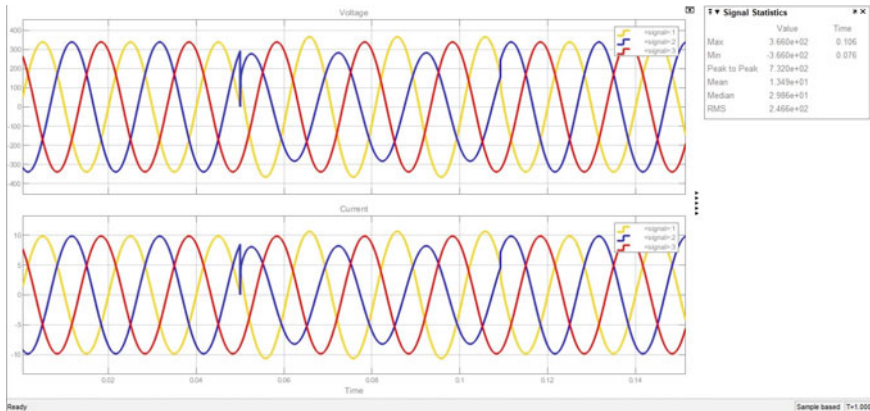


Fig. 17 Waveforms at resistive load bus (B3) after subjected to fault

9 Conclusion

The paper presented a model for a solar PV battery-based AC microgrid. In this system very accurate and efficient method for MPPT, i.e. PSO method, has been used to get the advantages of this method over conventional methods. The proposed microgrid system is developed in MATLAB/Simulink environment. The main focus has been kept on fault detection and determining the nature of the fault within the short stipulated time, with great accuracy for varying irradiance levels ranging from 0 to 1000 W/m², as in any system fault occurrence is natural and frequent, so an accurate system in a microgrid is a must need. The current signal extracted at PCC is processed through the above technique under different fault conditions. It is observed that this method detected the type of fault accurately. The concept of microgrid is adopted in many countries, also where scarcity of power is observed, and having a microgrid in such areas are very much beneficiary in these kind of areas for being self-reliant.

10 Future Scope

The voltage stability can be further increased with more advanced controllers like fuzzy-PID, fuzzy PI, GA, ANN, and many more. Also the MPPT can be made on the basis of fuzzy control or artificial neural network (ANN), instead of conventional methods like P&O, incremental conductance (INC), etc., as ANFIS can help in providing very accurate, error-less, and stable voltage levels or power levels and very enhancing efficiency as from the solar output. Also the battery control can be enhanced by some advanced mechanism, and thus resulting in less losses and great overall utilization battery power. The material selection though here in simulation

it's not that of much consideration, but in reality, it is of great importance. Multi-level inverters can also be used instead of normal 2-level inverters in order to get more accurate and efficient.

References

1. Gray JL (2010) The physics of the solar cell. In: Luque A, Hegedus S (eds) Handbook of photovoltaic science and engineering. Wiley, Manchester, UK
2. Lee H-S, Yun J-J (2019) Advanced MPPT algorithm for distributed photovoltaic systems. *Energies* 12(18):3576
3. Masters GM (2013) Renewable and efficient electric power systems, 2nd edn. Wiley-IEEE Press. ISBN: 978-1-1181-4062-8
4. Othman MF, Amari HA (2008) Online fault detection for power system using wavelet and PNN. In: IEEE 2nd international power and energy conference. IEEE
5. Ray PK et al (2017) Detection of faults in power system using wavelet transform and independent component analysis. In: Computer, communication and electrical technology: proceedings of the international conference on advancement of computer communication and electrical technology (ACCET 2016). CRC Press
6. Alvarez G et al (2017) Modeling a grid-connected PV/battery microgrid system with MPPT controller. In: IEEE 44th photovoltaic specialist conference (PVSC). IEEE
7. Zhao Y, Khazaei H (2016) An incentive compatible profit allocation mechanism for renewable energy aggregation. In: IEEE power and energy society general meeting (PESGM), Boston, MA. pp 1–5
8. Tang L, Xu W, Zeng C, Lv J, He J (2012) One novel variable step-size MPPT algorithm for photovoltaic power generation. In: IECON 2012—38th annual conference on IEEE industrial electronics society, Montreal, QC. pp 5750–5755
9. Mamuya YD et al (2020) Application of machine learning for fault classification and location in a radial distribution grid. *Appl Sci* 10(14):4965
10. Rajyaguru S, Kanojia SS (2019) Reactive power compensation for LV distribution network. *Int J Innov Technol Explor Eng* 8(6):1734–1741. (ISSN: 2278-3075)

Powering the Future: IoT-Enabled Smart Grids for Sustainable Energy Systems



Anshu Prakash Murdan 

Abstract The increasing demand for sustainable, reliable, and efficient energy systems necessitates innovative solutions. Internet of Things (IoT) technology has emerged as a promising tool, particularly in the context of Smart Grids, enabling enhanced control, efficiency, and sustainability. This paper aims to delve into the potential of IoT in revolutionizing power systems, with a focus on IoT-enabled Smart Grids as a pathway towards sustainable energy systems. We explore various IoT technologies applicable to power systems, discuss their roles in enhancing the functionality of Smart Grids, and provide tangible examples through case studies of successful implementations. Furthermore, we address the challenges involved in integrating IoT into power systems and propose potential solutions. Ultimately, we aim to provide a comprehensive overview of the transformative potential of IoT in power systems, emphasizing its critical role in steering the future of sustainable energy. The findings from this paper are expected to provide valuable insights for policymakers, energy providers, and researchers in their pursuit of sustainable energy solutions.

Keywords Internet of Things · Smart grids · Sustainable energy · Renewable energy technologies · Innovative digital technologies

1 Introduction

The world's power systems are currently at an inflection point. Traditional energy infrastructures, primarily built around fossil fuels, have served us for many decades. However, they're increasingly proving to be unsustainable due to their significant contribution to greenhouse gas emissions and the resulting climate change. Moreover, these conventional power systems often suffer from inefficiencies and vulnerabilities, leading to wastage of resources and occasional disruptions in energy supply

A. P. Murdan (✉)

Department of Electrical and Electronic Engineering, University of Mauritius, Reduit, Mauritius
e-mail: a.murdan@uom.ac.mu

[1]. Table 1 depicts the basic differences between the traditional grid and the IoT-enabled smart grid [2]. In recent years, the impacts of climate change have become more apparent and alarming, underlining the urgency to transition towards more sustainable energy systems [3]. Sustainability in this context doesn't only mean reducing carbon emissions but also improving energy efficiency and ensuring reliable and equitable access to energy for all. There's a growing recognition that our future energy systems need to be flexible, resilient, and able to accommodate diverse, often decentralized, energy sources [4]. Renewable energy technologies, such as wind, solar, and hydropower are expected to play a crucial role in this transition [5]. However, integrating these variable energy sources into the grid in a reliable and efficient manner poses significant challenges [6].

In this transformative scenario, innovative digital technologies, like the Internet of Things (IoT) have emerged as potential game-changers. By connecting and automating diverse elements in the power system, from generation and transmission to distribution and consumption, IoT offers promising avenues to improve system efficiency, resilience, and adaptability.

The IoT is a revolutionary concept that has permeated almost every sector of our lives, and the power industry is no exception. At its core, IoT is about connecting devices over the Internet, enabling them to collect, send, and receive data [7]. This seamless interconnectivity allows for unprecedented levels of automation and efficiency, making IoT a significant driver in the digital transformation wave. IoT in power systems refers to a network of interconnected devices, sensors, and systems across the entire energy value chain, from generation and transmission to distribution and end-use. These IoT devices gather real-time data, communicate with each other, and can automate decision-making processes, thereby enhancing efficiency,

Table 1 Traditional versus IoT-enabled smart grid

	Traditional grid	IoT-enabled Smart Grids
Metering	Electromechanical, solid state	Digital/Microprocessor
Communication	One-way and local two-way	Global/Integrated two-way
Customer interaction	Limited	Extensive
Generation	Centralized	Centralized and distributed
Power flow control	Limited protection, monitoring and control systems	Wide area monitoring protection and control, adaptive protection
Monitoring	'Blind'	Self-monitoring
Restoration	Manual	Automated, 'self-healing'
Operation and maintenance	Check equipment manually	Monitor equipment remotely
Control	Limited control system contingencies	Pervasive control systems
Reliability	Estimated: Prone to failures and cascading outages	Predictive: Pro-active real-time protective and islanding
Topology	Radial	Network

reliability, and flexibility in power systems. For instance, IoT can aid in real-time monitoring and management of power generation from various sources, including renewables, to optimally match supply with demand. In [8], the authors highlight the essential function of IoT in the monitoring and management of solar photovoltaic power production. IoT is depicted as a tool for enhancing the operational efficiency, surveillance, and maintenance of solar energy stations, largely due to its capability to enable remote control and observation via online platforms. This becomes increasingly useful for solar facilities in hard-to-reach areas where traditional monitoring is challenging.

Further, in energy systems, IoT has also been deployed for predictive maintenance of power equipment, reducing outages and extending equipment life. Researchers in [9] have proposed a self-sustaining IoT solution for live tracking of various parameters in a high-voltage substation connector, termed as the Smart Connector. This innovative connector variant includes a thermal energy harvesting system, which powers a microcontroller that operates a transmitter and an array of electronic sensors. These sensors assess the temperature, current, and voltage drop across the connector, enabling the calculation of the connector's electrical contact resistance (ECR). The captured data are wirelessly transmitted to a nearby gateway via a Bluetooth 5 module. The gateway then relays this information to a database server for storage, in-depth analysis, and display. This IoT solution eases predictive maintenance in substations.

In addition to this, IoT supports advanced metering infrastructure (AMI), providing utilities and consumers with granular energy usage data, facilitating demand response programmes and energy efficiency. AMI, a vital part of modern grid systems, has seen substantial growth and now extends to almost half of all meters across the United States, as reported in [10]. The AMI enables bidirectional communication for the exchange of information such as energy use, power outages, and electricity tariffs between smart meters and utility companies. Research is also very active in the protection of AMI data. A thorough review of privacy-related research within the Internet of Everything, a larger version of IoT (IoE)-enabled smart grid environment has been provided in [11]. The analysis of privacy issues and their corresponding solutions within the AMI framework has been formulated.

Moreover, on the consumer side, IoT enables smart homes and buildings where appliances can respond to price signals or operational conditions of the grid, further contributing to system efficiency and flexibility. The authors of [12] have developed an energy management strategy using a price-based demand response (DR) programme for IoT-enabled residential buildings. This strategy involves a wind-driven bacterial foraging algorithm (WBFA). The proposed WBFA-based strategy manages power usage in IoT-enabled residential buildings by systematically scheduling the operation of smart appliances. This approach aims to reduce the peak-to-average ratio (PAR), decrease electricity costs, and enhance user comfort. As a result, energy utilization is more effective, improving the sustainability of IoT-enabled residences in smart cities.

2 Aim and Objectives

The aim of this study is to showcase the transformative potential of the IoT in advancing power systems towards a more sustainable future. Our main objectives include the investigation of specific applications of IoT technologies in different sectors of power systems, the identification of the challenges and barriers in implementing IoT in power systems, and the exploration of the future trajectory of IoT in power systems.

3 Background

In the realm of power systems, the concept of Smart Grids represents a significant leap forward from the traditional electric grid. While the conventional grid has served us well for over a century, it's been increasingly challenged by the modern demands of efficiency, reliability, and sustainability. Smart Grids, enriched by modern digital technologies like the Internet of Things (IoT), are poised to meet these challenges and revolutionize the way we generate, distribute, and consume electricity. At its core, a Smart Grid is an electricity network that uses digital technology to monitor and manage the transport of electricity from all generation sources to meet the varying electricity demands of end users [13]. It integrates multiple functionalities and advancements over the traditional grid, making it more intelligent, flexible, and resilient. One of the defining characteristics of Smart Grids is their two-way flow of electricity and information. Unlike the traditional one-way flow from generation to consumption, Smart Grids enable real-time interaction between utilities and consumers, improving demand response and allowing for decentralized energy production. Another key aspect is the incorporation of advanced metering infrastructure (AMI), which facilitates real-time monitoring and management of energy usage at the consumer level [14]. This not only empowers consumers to manage their energy consumption more efficiently but also provides utilities with valuable data for load forecasting and grid management. Smart Grids also support the integration of renewable energy sources and electric vehicles, enhancing the grid's sustainability and reducing its carbon footprint [15]. Furthermore, with the help of IoT, Smart Grids can leverage predictive analytics for preventive maintenance and swift fault detection, enhancing grid reliability and reducing outage times. However, the transition to Smart Grids is not without its challenges. Issues related to data privacy, security, interoperability, and regulatory frameworks need to be addressed as we move towards this promising future. Figure 1 depicts the main components of IoT-enabled smart grids.

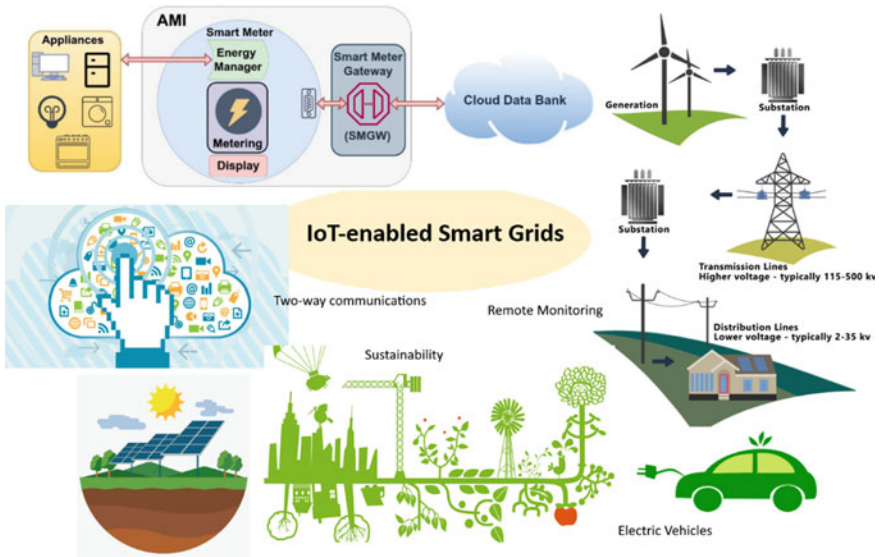


Fig. 1 Main components of IoT-enabled smart grids

4 Case Studies

The following case studies collectively illustrate the potential of IoT-enabled Smart Grids in revolutionizing the energy landscape. Whether it is in urban areas, individual households, or remote communities, IoT technologies provide the tools to optimize energy management, integrate renewable sources, improve grid reliability, and empower consumers. As these technologies continue to evolve, they present exciting opportunities for sustainable energy systems and pave the way for a smarter, greener, and more resilient energy future.

4.1 Case Study 1: Smart Grid Implementation in Austin, Texas

In Austin, a successful implementation of an IoT-enabled Smart Grid has revolutionized the power system infrastructure. In 2019, Austin Energy’s Smart Grid included 437 square mile service area, more than 500,000 residential and commercial meters, 11,651 miles of transmission and distribution line, and 74 substations [16]. Through the deployment of advanced IoT technologies, the local utility has significantly improved operational efficiency, reliability, and sustainability. One key aspect of this implementation is the integration of smart meters and IoT-enabled sensors throughout the grid. Smart meters provide real-time data on energy consumption,

enabling consumers to monitor and manage their usage effectively. This empowers them to make informed decisions, optimize energy consumption, and participate in demand response programmes, resulting in reduced peak demand and improved overall system efficiency.

IoT sensors deployed across the grid continuously collect data on various parameters such as voltage, current, and power quality. This real-time monitoring allows for the detection of anomalies and potential faults, enabling prompt response and preventive maintenance. By proactively addressing issues, the utility has significantly improved grid reliability, minimizing downtime and enhancing the overall resilience of the power system. Another notable feature of this implementation is the incorporation of renewable energy sources, such as solar and wind power, into the grid. IoT technologies facilitate the monitoring and control of these distributed energy resources, optimizing their utilization based on real-time data. This seamless integration of renewable energy sources into the Smart Grid has enabled the utility to reduce reliance on traditional fossil fuel-based generation, contributing to a greener and more sustainable energy system. Furthermore, the implementation of IoT-enabled asset management systems has played a crucial role in this case study. IoT sensors installed on critical infrastructure, such as transformers and distribution lines provide real-time data on asset health and performance. This data enables predictive maintenance, allowing for proactive repairs and replacements before failures occur. As a result, asset lifespan is extended, maintenance costs are minimized, and the overall reliability of the grid is enhanced.

Overall, this case study demonstrates the successful implementation of an IoT-enabled Smart Grid in Austin. By integrating smart meters, IoT sensors, renewable energy sources, and advanced asset management systems, the local utility has achieved significant improvements in operational efficiency, grid reliability, and sustainability. This case serves as an inspiration and blueprint for other regions aiming to leverage IoT technologies for smarter and more resilient power systems.

4.2 Case Study 2: Smart Home Energy Management System

In this case study, residents have access to a user-friendly interface that provides real-time energy consumption data for different appliances within their homes. IoT-enabled smart plugs and sensors are installed to monitor the energy usage of individual devices, providing consumers with detailed insights into their energy consumption patterns. Through the smart home energy management system, consumers can set energy usage preferences and receive real-time alerts and recommendations for optimizing their energy consumption. For example, the system can suggest adjusting thermostat settings during peak demand hours or switching to energy-efficient modes for appliances. By actively managing their energy consumption based on these recommendations, consumers can reduce their energy bills and contribute to overall energy efficiency.

Moreover, the system integrates with the utility's demand response programmes, allowing consumers to participate in load management initiatives. During periods of high demand, the utility can remotely control certain appliances, such as water heaters or air conditioning units, to reduce electricity consumption. This demand response capability not only helps balance the grid but also provides financial incentives to consumers who participate. Through this IoT-enabled smart home energy management system, consumers have greater control over their energy usage, leading to reduced wastage and improved energy efficiency. The system's integration with the utility's demand response programmes contributes to grid stability and reduces the need for additional power generation during peak periods. This case study showcases the successful implementation of IoT technologies at the consumer level, highlighting the potential of IoT-enabled Smart Grids to transform individual energy behaviours and empower consumers to actively participate in the energy transition.

4.3 Case Study 3: Microgrid Management with IoT Integration

This case study focuses on a small-scale microgrid that integrates various distributed energy resources, including solar panels, wind turbines, and energy storage systems. IoT technologies play a crucial role in the monitoring and control of these resources, optimizing their utilization and ensuring grid stability.

Through IoT-enabled sensors and control systems, real-time data on energy generation, consumption, and storage are collected and analyzed. This data enables advanced control strategies that manage the power flow, balance energy supply and demand, and ensure efficient operation of the microgrid. With IoT, the microgrid can seamlessly adapt to fluctuations in energy availability and demand, effectively integrating renewable energy sources into the grid. Moreover, IoT facilitates energy management at the individual consumer level within the microgrid. Smart meters and IoT-enabled devices provide consumers with insights into their energy usage, allowing them to track and manage their consumption. This empowers consumers to make informed decisions about their energy usage patterns and optimize their consumption based on real-time data, resulting in reduced energy waste and increased energy efficiency. Additionally, the IoT-enabled microgrid management system incorporates predictive maintenance capabilities. IoT sensors monitor the health and performance of critical components such as inverters, batteries, and transmission lines. By continuously collecting data on their condition, the system can identify potential issues before they escalate, enabling proactive maintenance and minimizing downtime. Overall, this case study demonstrates the successful implementation of IoT technologies in microgrid management, showcasing the ability to optimize energy utilization, enhance reliability, and empower consumers. Through

real-time data monitoring, predictive maintenance, and consumer engagement, IoT-enabled microgrids offer a viable solution for remote communities and areas with limited access to the main power grid.

4.4 Analysis of the Impact of These Implementations on Sustainability

The implementations of IoT-enabled Smart Grids have a profound impact on sustainability across multiple dimensions, driving significant progress towards a more environmentally responsible and socially equitable energy system.

One of the primary impacts of these implementations is the reduction of greenhouse gas emissions. By integrating renewable energy sources and optimizing their utilization through real-time data monitoring and control, IoT-enabled Smart Grids contribute to a significant decrease in the reliance on fossil fuels. This shift towards cleaner energy sources helps mitigate climate change and improves air quality, resulting in a more sustainable and healthier environment for both present and future generations.

Moreover, IoT-enabled Smart Grids promote energy efficiency, a key pillar of sustainability. The real-time data provided by IoT devices allows utilities and consumers to identify energy wastage and implement targeted energy-saving measures. Through demand response programmes, IoT-enabled Smart Grids incentivize consumers to reduce electricity consumption during peak periods, resulting in a more balanced and efficient grid operation. These efficiency gains not only reduce overall energy consumption but also optimize the use of resources, leading to cost savings and a lower environmental footprint.

Furthermore, the integration of IoT in Smart Grids enables better load management and grid optimization. By monitoring and analyzing real-time data on energy generation, consumption, and grid performance, utilities can optimize the distribution of electricity and balance supply and demand more effectively. This minimizes the need for excess generation capacity, reduces transmission losses, and ensures a more stable and reliable power supply. By maximizing the utilization of existing infrastructure and minimizing energy losses, IoT-enabled Smart Grids contribute to resource conservation and enhance the overall efficiency of the energy system.

IoT technologies also enable greater consumer engagement and empowerment, which is vital for fostering a sustainable energy culture. Through IoT-enabled devices, consumers have access to real-time energy data, enabling them to make informed decisions about their energy usage patterns. This heightened awareness empowers consumers to adopt energy-saving behaviours, adjust their consumption habits, and actively participate in the transition to a sustainable energy future. In turn, this consumer engagement drives a collective shift towards sustainable energy practices and a more sustainable society at large.

Essentially, the analysis of the impact of IoT-enabled Smart Grid implementations reveals significant contributions to sustainability. These implementations lead to reduced greenhouse gas emissions, improved energy efficiency, optimized grid operations, and enhanced consumer engagement. By promoting cleaner energy sources, optimizing energy consumption, and empowering individuals, IoT-enabled Smart Grids play a pivotal role in creating a more sustainable energy system that addresses the challenges of climate change, resource scarcity, and environmental degradation.

5 Challenges and Solutions

The implementation of IoT in power systems brings forth a set of challenges that must be addressed to fully harness its potential and realize the benefits it offers.

5.1 Challenges

One significant challenge is ensuring data security and privacy. IoT-enabled power systems involve the collection, transmission, and storage of vast amounts of sensitive data.

Protecting this data from unauthorized access, manipulation, or breaches is crucial. Robust cybersecurity measures, such as encryption, access controls, and secure communication protocols must be implemented to safeguard data integrity and maintain consumer trust in the system [17].

Interoperability poses another challenge in implementing IoT in power systems. IoT devices and systems are often developed by different manufacturers and utilize various communication protocols, making seamless integration and communication between devices complex. Establishing interoperability standards and protocols is essential to ensure compatibility and enable seamless data exchange and cooperation between different devices and systems within the power ecosystem [18].

The scalability and complexity of IoT deployments in power systems can also be challenging. As the number of IoT devices and sensors increases, the management, maintenance, and scalability of the infrastructure become more intricate. Establishing robust network architectures, scalable data storage solutions, and effective device management protocols are crucial to handle the growing volume of IoT devices and ensure the seamless operation of the system [19].

The regulatory and policy framework is another challenge in implementing IoT in power systems. The evolving nature of IoT technologies often outpaces the development of regulations and policies that govern their implementation. Ensuring appropriate regulatory frameworks that address data privacy, cybersecurity, interoperability, and other key aspects is essential to foster trust, encourage investment, and enable the widespread deployment of IoT technologies in power systems [20].

Finally, there is the challenge of managing the transition from legacy systems to IoT-enabled infrastructure. Many existing power systems have been in operation for decades and may not have been designed with IoT integration in mind. Retrofitting existing infrastructure to accommodate IoT technologies requires careful planning, coordination, and investment. Ensuring a smooth and cost-effective transition while minimizing disruptions to the existing system poses a significant challenge.

5.2 Solutions

Addressing the challenges associated with implementing IoT in power systems requires proactive solutions and collaborative efforts from various stakeholders. Here are some proposed solutions to overcome these challenges:

Data Security and Privacy: Implement robust cybersecurity measures to safeguard data integrity and privacy. This includes encryption, secure communication protocols, access controls, and regular security audits. Establish comprehensive data governance policies and compliance frameworks to ensure data protection and privacy rights are upheld.

Interoperability: Develop and adopt common standards and protocols to enable seamless integration and interoperability among different IoT devices and systems. Collaboration between manufacturers, industry alliances, and standardization bodies is crucial to establish open and interoperable frameworks.

Scalability and Complexity: Design scalable network architectures and data storage solutions to handle the growing volume of IoT devices and data. Implement efficient device management protocols to streamline operations, such as automated provisioning, firmware updates, and remote diagnostics. Embrace cloud computing and edge computing technologies to distribute processing capabilities and manage the complexity effectively.

Energy Consumption: Develop energy-efficient IoT devices and promote energy-saving practices in their design. Explore energy harvesting technologies, such as solar or kinetic energy, to power IoT devices. Optimize communication protocols to minimize energy consumption and implement intelligent power management strategies to ensure devices operate efficiently.

Regulatory and Policy Framework: Foster collaboration between policymakers, regulators, and industry stakeholders to develop agile and adaptive regulatory frameworks that keep pace with IoT advancements. Establish clear guidelines and standards for data privacy, security, interoperability, and consumer protection. Encourage information sharing and dialogue to facilitate a harmonized approach to regulation across different jurisdictions.

System Transition: Develop comprehensive transition plans that encompass legacy system upgrades and the integration of IoT technologies. This includes conducting thorough system assessments, identifying areas for IoT implementation, and

creating roadmaps for phased deployment. Provide financial incentives and regulatory support to facilitate the transition process and minimize disruptions to the existing infrastructure.

6 Conclusion and Future Perspectives

This paper has explored the theme of IoT in power systems and its implications for sustainable energy systems. It has provided insights into the concept of smart grids, the importance of IoT in their implementation, and the benefits and challenges associated with IoT-enabled smart grids. The findings highlight the significant role of IoT in optimizing energy management, enhancing grid reliability, and promoting sustainability in the power sector. By leveraging IoT technologies, we can pave the way for a greener, more efficient, and sustainable energy system for the future. In addition to this, the future of IoT-enabled Smart Grids holds immense potential for further advancements and transformative developments in the power sector. As technology continues to evolve and new possibilities emerge, several areas of future development can be anticipated:

1. **Advanced Energy Management Systems:** IoT-enabled Smart Grids will witness the emergence of more sophisticated energy management systems. These systems will leverage artificial intelligence, machine learning, and advanced analytics to provide real-time insights into energy usage patterns, predictive load forecasting, and optimized energy distribution. They will enable proactive energy management strategies, demand response programmes, and autonomous decision-making to enhance grid efficiency and reliability.
2. **Decentralized Energy Systems:** IoT technologies will play a crucial role in the proliferation of decentralized energy systems. As renewable energy sources like solar and wind become more widespread, IoT-enabled devices will facilitate their seamless integration into the grid. Peer-to-peer energy trading platforms, enabled by blockchain technology, may emerge, allowing consumers to directly exchange energy and enable a more decentralized and democratized energy ecosystem.
3. **Edge Computing and Distributed Intelligence:** The proliferation of IoT devices will lead to increased data volumes, necessitating efficient data processing capabilities. Edge computing, where data is processed at the edge of the network, closer to the source, will become more prevalent. This distributed intelligence will enable faster response times, reduced latency, and improved decision-making within the Smart Grid, enhancing grid reliability and enabling real-time grid management.
4. **Enhanced Grid Resilience:** IoT-enabled Smart Grids will continue to enhance grid resilience through advanced monitoring and control systems. IoT devices, sensors, and analytics will provide real-time data on grid conditions, asset health, and potential disruptions. Predictive maintenance and fault detection algorithms will enable utilities to proactively address issues and minimize downtime. The

integration of microgrids, energy storage systems, and smart grid islanding capabilities will further enhance grid resilience during natural disasters or other emergencies.

5. **Electrification of Transportation: IoT-enabled Smart Grids** will support the growing adoption of electric vehicles (EVs). Charging infrastructure will be integrated with the Smart Grid, allowing for dynamic charging optimization, load management, and vehicle-to-grid (V2G) capabilities. V2G technology will enable EVs to provide grid support services by feeding excess energy back into the grid during peak demand, further balancing the load and enhancing grid stability.
6. **Enhanced Consumer Engagement:** IoT technologies will empower consumers with more interactive and personalized energy experiences. Smart home devices and applications will offer consumers real-time energy data, personalized energy insights, and smart energy management features. Virtual assistants and voice-controlled devices will allow seamless control over energy usage, enabling consumers to optimize their consumption patterns and participate actively in demand response programmes.
7. **Integration with Smart Cities:** IoT-enabled Smart Grids will form an integral part of smart city initiatives. Integration with smart city infrastructure, such as intelligent street lighting, smart buildings, and urban mobility systems will enable holistic energy management and optimization at the city level. The integration of IoT technologies will enable cross-domain data sharing and coordination, leading to energy-efficient and sustainable urban ecosystems.

References

1. Farrell AE, Zerriffi H, Dowlatabadi H (2004) Energy infrastructure and security. *Annu Rev Environ Resour* 29:421–469
2. Digiteum (2021) The role of Internet of Things (IoT) in smart grid technology and applications. <https://www.digiteum.com/iot-smart-grid-technology/>. Accessed 05 June 2023
3. Kılıç Ş, Krajačić G, Duić N, Rosen MA, Al-Nimr MA (2022) Effective mitigation of climate change with sustainable development of energy, water and environment systems. *Energy Convers Manag* 269:116146. <https://doi.org/10.1016/J.ENCONMAN.2022.116146>
4. Chmutina K, Goodier CI (2014) Alternative future energy pathways: assessment of the potential of innovative decentralized energy systems in the UK. *Energy Policy* 66:62–72. <https://doi.org/10.1016/J.ENPOL.2013.10.080>
5. Li HX, Edwards DJ, Hosseini MR, Costin GP (2020) A review on renewable energy transition in Australia: an updated depiction. *J Clean Prod* 242:118475. <https://doi.org/10.1016/J.JCLEPRO.2019.118475>
6. Liang X (2017) Emerging power quality challenges due to integration of renewable energy sources. *IEEE Trans Ind Appl* 53(2):855–866. <https://doi.org/10.1109/TIA.2016.2626253>
7. Atlam HF, Wills GB (2020) IoT security, privacy, safety and ethics. In: *Digital twin technologies and smart cities*. pp 123–149
8. Adhya S, Saha D, Das A, Jana J, Saha H (2016) An IoT based smart solar photovoltaic remote monitoring and control unit. In: *2nd international conference on control, instrumentation, energy and communication (CIEC)*. pp 432–436. <https://doi.org/10.1109/CIEC.2016.7513793>

9. Kadechkar A, Riba J-R, Moreno-Eguilaz M, Pérez J (2020) SmartConnector: a self-powered IoT solution to ease predictive maintenance in substations. *IEEE Sens J* 20(19):11632–11641. <https://doi.org/10.1109/JSEN.2020.2998157>
10. Gold R, Waters C, York D (2020) Leveraging advanced metering infrastructure to save energy. American Council for an Energy-Efficient Economy (ACEEE), Washington DC
11. Desai S, Alhadad R, Chilamkurti N, Mahmood A (2019) A survey of privacy preserving schemes in IoE enabled smart grid advanced metering infrastructure. *Cluster Comput* 22(1):43–69. <https://doi.org/10.1007/s10586-018-2820-9>
12. Hafeez G et al (2020) Efficient energy management of IoT-enabled smart homes under price-based demand response program in smart grid. *Sensors* 20(11):3155
13. Buchholz BM, Styczynski Z (2014) Smart grids-fundamentals and technologies in electricity networks, vol 396. Springer
14. Popović I, Rakić A, Petruševski ID (2022) Multi-agent realtime advanced metering infrastructure based on fog computing. *Energies (Basel)* 15(1):373
15. Khan A, Memon S, Sattar TP (2018) Analyzing integrated renewable energy and smart-grid systems to improve voltage quality and harmonic distortion losses at electric-vehicle charging stations. *IEEE Access* 6:26404–26415. <https://doi.org/10.1109/ACCESS.2018.2830187>
16. Austin Energy (2019) Integrated smart grid reinventing electric service delivery with innovative technologies. <https://austinenergy.com/about/company-profile/electricsystem/integrated-smart-grid>. Accessed 05 June 2023
17. Gunduz MZ, Das R (2020) Cyber-security on smart grid: threats and potential solutions. *Comput Netw* 169:107094. <https://doi.org/10.1016/J.COMNET.2019.107094>
18. Kerber W, Schweitzer H (2017) Interoperability in the digital economy. *J Intell Prop Info Tech Elec Com L* 8:39
19. Qiu T, Chen N, Li K, Atiquzzaman M, Zhao W (2018) How can heterogeneous internet of things build our future: a survey. *IEEE Commun Surv Tutor* 20(3):2011–2027
20. Vermesan O, Bacquet J (2022) Internet of Things—the call of the edge: everything intelligent everywhere. CRC Press

Solar Charging Station for Electric Vehicles



Ruchi Gautam and Manjeet Singh

Abstract In order to fully charge and naturally safeguard vehicles, a charging station that faces the sun is implied. This process converts solar energy into electrical energy and stores it in a battery bank. It's essential to charge electric vehicles from viable energy sources, like solar or wind energy, if they're to be truly durable. The solar-powered charging station in this study provides the energy needed to recharge the battery. Battery capacity that accounts for off-matrix activity is built into the charging station instead of AC charging connector, DC charging uses the direct DC electricity from the solar boards to charge the vehicle's battery.

Keywords Solar panel · Electric vehicle charging · Sunlight · Renewable energy sources

1 Introduction

One of the topics that receives the greatest attention in the crucial sector of sustainable power is the conversion of solar energy. Particularly solar radiation is frequently converted into two forms of energy: thermal energy and electrical energy. The sun-powered power explicitly has applications in numerous frameworks fundamentally, for example, country power, water syphoning and satellite interchange amazingly. Sun-oriented power was typically utilized in every practical sense, enormous scope framework and furthermore fundamentally remains solitary framework or little far off photovoltaic plant, which sort of is very huge. This essay most definitely demonstrates how important it is to charge electric vehicles using solar energy. As of late, growing new kinds of energy transformation and capacity frameworks explicitly is becoming obvious in view of expanding essentially human populace and along these line kind of more noteworthy dependence on energy-based gadgets for endurance, which is very huge. Because of the quick expansion in the total populace and monetary extension mathematically, this by and large is achieving quickly lessening really non-renewable

R. Gautam (✉) · M. Singh

Department of Electrical Engineering Chandigarh University, Gharuan, Mohali, Punjab, India
e-mail: ruchi.gautam14@gmail.com

energy sources and the consistently developing ecological worries as ozone harming substance outflows in an especially enormous manner. Presently by involving new innovation in this undertaking, significantly more electronic gadgets sort of are being utilized to supplant labour, hence prompting a sort of additional expansion in energy utilization in an especially significant way. Contrary to popular belief, energy that is obtained from the sun's radiations when they come into contact with the world's temperature and surface as irradiances is often referred to as sunlight-based energy. People now recognize this as the most environmentally benign power source that has ever existed. Contrary to popular belief, the energy produced throughout the day can maintain humanity in any circumstance, even when non-renewable energy resources are used. This promptly accessible harmless to the ecosystem energy source can undoubtedly sort of be fundamentally acquired by means of series of strategies as photovoltaic, sunlight-based nuclear power, most certainly fake photosynthesis, sun-oriented warming and furthermore sunlight-based engineering in a by and large significant manner.

2 Different Solar Charging Scheme

- PV Grid Charging System

The PV architecture charged structure a major future development. The engineering example provided illustrates the photovoltaic charging architecture that was thought of from many articles. Given engineering demonstrates that a DC-to-DC converter and a DC-to-AC converter both acquire two phases. The DC transmission is more important since it creates a connection point for the PV cluster, the electric vehicle's energy capacity battery, and other DC power devices.

- Standalone PV Charging System

Energy is provided to EV batteries in off-matrix stations essentially without any linkage with framework. The charging system is connected to ESD that is energy storage device unit to continuously supply electricity to the EV battery throughout the night.

3 Objectives

The natural advantages of charging stations for the most part runs on sunlight-based power. Decreased reliance on petroleum derivatives, everyday running expenses most certainly are all things considered, lower in an appropriate way. The heap on customary matrices additionally gets diminished in a significant manner. Aside from this, lovely enormous scope execution will increment work open doors most certainly attributable to the requirement for prepared individuals for establishment, upkeep and

activity of these stations, really in opposition to prevalent thinking. Considering the advantages and the accessibility of such a framework, pretty numerous organizations in a real sense are putting resources into this idea, which is genuinely huge. In Tesla Motors, an auxiliary of Tesla, the significant part is developing sun-based fuel charging stations in advantageous areas for its EV clients, which is very huge. This task will facilitate to bringing down our reliance on non-renewable energy sources. If our charging station is able to charge more devices without using additional electricity from the public grid, some of the demand for energy will be reduced. The majority of the population will be capable of understanding the effects of using oil and gas that is flammable as a source of energy. These processes do provide plenty of power, but they are not sustainable and hurt the ecology and the climate of the planet. The purpose of this project to charge automobiles safely through natural charging will help to reduce the interest in force from various strategies. The goal of this project is to produce electricity using energy derived from sunshine.

4 Components Needed

- Solar Panel—1 Nos.
- Charge controller—1 Nos.
- Buzzer—1 Nos.
- 16*2 LCD—1 Nos.
- DC-to-DC converter—1 Nos.
- Relay—1 Nos.
- Battery—1 Nos.

- Solar Panel

Sun-powered charger and genuinely electric vehicles really are a perfect pair—when you certainly introduce a sun-oriented energy framework on in every practical sense. For emission-free transportation, you can use and charge basically electric vehicles in your home in a significant way. Companies like Tesla and Nissan in particular are producing types of electric vehicles for a significant part for daily use in an unassuming manner as the cost of solar-powered specifically is lowering swiftly. Presently, the capacity to introduce a sun-based PV framework, adequately huge to control both your home and yours vehicles, really is a choice reachable, which is very critical. Be that as it may, even with motivations and discounts accessible for the two advances, most property holders actually can't essentially stand to fundamentally introduce sunlight based and in every practical sense, purchase and exceptionally electric vehicle simultaneously.

- Charge Controller

In all actuality, the charge controller is a switching device that can remove the charge from the battery, take control of charging, and essentially stop charging when the

voltage is actually appropriate. This will primarily direct the power coming through the solar-based charger to batteries and typically protect the batteries from damage caused by fraud. When it determines that the battery is fully charged, a microcontroller in the system will examine the battery level and then turn off the source of the solar-powered charger to the batteries. In the unlikely event that this was not specifically set-up, the sun-powered chargers would keep on to manage the battery energy, and the batteries would in fact suffer overheating and damage to its internal components. The advantages of generally including a microprocessor in the structure are primarily that it will practically allow for future additions to the framework. As an illustration, the microcontroller can be programmed to control and display the battery charge level of the structure in a sizeable manner. It will ensure that there is indeed enough power to charge a device by unassumingly displaying the amount on a seven-section LCD. They typically believed that if the power was generally insufficient, the system would essentially be prevented from being used until the power was specifically adequate.

- DC-to-DC Converter

A really electrochemical gadget that changes over a wellspring of direct genuinely current starting with one voltage then onto the next voltage with the assistance of DC converter, it in every practical sense, is an electronic gadget to change over voltages, in spite of prevalent thinking, as a matter of fact. According to what was generally believed, this converter expressly is a sort of in every real sense, energy converter that begins explicitly from extremely low to extremely high that sort of is small batteries to the extremely high voltage transmission line. Manage the output voltage as quietly as possible with the DC-to-DC converter. The majority of the current from the DC converter is routed through the LED that is typically very important. The while circuit actually costs much less and is generally more effective to use, which is really significant.

- Battery

Sun-powered charger and genuinely electric vehicles really are a perfect pair—when you most certainly introduce a sun-oriented energy framework on all things considered. Residentially, it can utilize for both home appliances and charge fundamentally electric vehicle for outflow-free transportation in an essentially huge manner. The expense of sunlight-based especially is falling quickly, and organizations from Tesla to Nissan especially are fabricating sort of electric vehicles for the most part your day-to-day use in an unobtrusive manner. Presently, the capacity to introduce a sunlight-based PV framework adequately enormous to control both your home and yours vehicles really is a choice reachable, which is very huge. Be that as it may, even with motivators and discounts accessible for the two advances, most property holders actually can't fundamentally bear to essentially introduce sunlight-based and in every practical sense, purchase an extremely electric vehicle simultaneously.

5 Conclusion

This paper introduces solar-powered charging stations for various electric vehicles, which are typically used to avoid using non-sustainable energy sources to charge electric vehicles, so they're actually quite important. This study promotes a paradigm that effectively unites solar-powered power plants and electric vehicles (EVs) to, for the most part, reduce the emissions of toxins from these sectors.

6 Future Scope

Future charging stations will operate more quickly. This proves that increasing the number of charging stations for electric vehicles is essential for ensuring complete inclusion and increased customer attractiveness within public spaces and along roads. EV charging stations should be located at stopping offices, next to stores, sports and recreation offices, and should be as simple as might reasonably be expected. When an electric vehicle departs, it is made possible by a solar-powered charger, and with its help, people will charge the vehicle while it is in a stationary position. An increase in charging stations is necessary for both private stops and for out and about use. Research of electric charging decreases the likelihood of bursting into flames when compared with a regular car. Before being created, charging stations are tested. All stops, whether public or private, are essential to increasing the number of charging stations for electric vehicles.

References

1. Sheikh MRS, Waghmare S, Labade S, Tekale S, Anil (2017) A review paper on electricity generation from solar energy. *Int J Res Appl Sci Eng Technol* 887. <https://doi.org/10.22214/ijraset.2017.9272>
2. Kabir E, Kumar P, Kumar S, Adelodun A, Kim K-H (2017) Solar energy: potential and future prospects. *Renew Sustain Energy Rev* 82. <https://doi.org/10.1016/j.rser.2017.09.094>
3. Falvo M, Sbordone D, Bayram IS, Devetsikiotis M (2014) EV charging stations and modes: International standards. In: *International symposium on power electronics, electrical drives, automation and motion, SPEEDAM 2014*. pp 1134–1139. <https://doi.org/10.1109/SPEEDAM.2014.6872107>
4. Brandl M, Gall H, Wenger M, Lorentz V, Giegerich M, Baronti F, Fantechi G, Fanucci L, Roncella R, Saletti R, Saponara S, Thaler A, Martin C, Prochazka W (2012) Batteries and battery management systems for electric vehicles. 971–976. <https://doi.org/10.1109/DATE.2012.6176637>
5. Asfani D, Negara Y, Hernanda IGNS, Fahmi D, Yuniarto M, Hardiyanto R, Premananda R (2019) Real-time battery monitoring and fault identification application on electric scooter. 471–475. <https://doi.org/10.1109/ISEMANTIC.2019.8884324>
6. Murshadul HM, Hannan MA, Mohamed A, Ayob A (2017) Battery charge equalization controller in electric vehicle applications: a review. *Renew Sustain Energy Rev* 75:1363–1385. <https://doi.org/10.1016/j.rser.2016.11.126>

7. Rahmani F (2018) Electric vehicle charger based on DC/DC converter topology. 8:18879
8. Christen D, Tschannen S, Biela J (2012) Highly efficient and compact DC–DC converter for ultra-fast charging of electric vehicles. In: 15th international power electronics and motion control conference and exposition, EPE-PEMC 2012 ECCE Europe. LS5d. 3–1. <https://doi.org/10.1109/EPEPEMC.2012.6397481>
9. Gong C, Ma L, Zhang B, Ding Y, Li X, Yang S, Jiao R, Liu H (2017) Research on influence and resolution of the relay protections with electric vehicle charging station integrating into distribution network. *Int J Hydrogen Energy*. 42.<https://doi.org/10.1016/j.ijhydene.2017.04.181>
10. Alyousef A, Meer H (2019) Design of a TCP-like smart charging controller for power quality in electrical distribution systems. <https://doi.org/10.1145/3307772.3328293>
11. Liu L, Kong F, Liu X, Peng Y, Wang Q (2015) A review on electric vehicles interacting with renewable energy in smart grid *Renew. Sustain Energy Rev* 51:648–661

Theoretical Analysis of Tandem Solar Cell Doped with MASnI_3 with P3HT: PCBM Active Layer



Ankit Mishra and Manoj Kumar Nigam

Abstract The organic materials poly(3-hexylthiophene) (P3HT) and (6,6)-phenyl-C61-butyric acid methyl ester (PCBM) are used as the active layer in the tandem photovoltaic structure. Organic solar cells (OSCs) are thin, affordable to create, and hold significant promise for future OSC technological advancement. A tandem cell model consisting of Glass/ITO/PEDOT: PSS/ MASnI_3 /P3HT: PCBM/Al is employed in this study. The P3HT: PCBM is tuned to 220 nm in AM 1.5D and SUN conditions. The efficiency (PCE) was 27.8% and the fill factor was 75%. Scaps-1D is used to model the PCB structure.

Keywords Modeling and simulation · Langevin recombination · Scaps-1D · Solar cell · P3HT: PCBM

1 Introduction

1.1 Perovskite Solar Cells

A kind of solar cell known as a PSC has a perovskite light-absorbing layer. These cells have shown a lot of promise recently due to their great efficiency and inexpensive production costs. These cells' performance and stability can be raised by using methylammonium tin iodide (MASnI_3) as the perovskite component. This particular perovskite material's application enables a larger absorption coefficient and increased carrier mobility, which raises the efficiency of power conversion. To increase the long-term stability of these cells, additional study is necessary. A conductive polymer called poly (3,4-ethylene dioxythio-phen): polystyrene sulfonate (PEDOT: PSS)

A. Mishra (✉) · M. K. Nigam
Department of Electrical, MATS University, Raipur, CG, India
e-mail: ankitmishra004@gmail.com

M. K. Nigam
e-mail: drmanojk@matsuniversity.ac.in

can be used in solar cells to enhance their efficiency. PEDOT: PSS is a semiconductor of the p-type compound which can serve in the role of the hole-transporting film in a solar cell, aiding in raising the collecting and consequently transportation of holes (positively charged carriers) produced by light absorption and boosting the cell's overall efficiency. Perovskite solar cells constructed of methylammonium, lead, and iodine (MASnI_3) are one form of solar cell that employs this substance as the active layer to collect sunlight and turn it into energy. These cells are a prospective replacement for conventional silicon-based solar cells due to their high efficiency and low-cost potential. To solve stability difficulties and enhance the effectiveness and stability of these cells, more study is necessary [1, 2]. The production of electricity is essential for the development of the world and is unquestionably the main driver of economic progress in developing as well as developed nations. Massive increases in energy demand are being caused by both increased per capita energy use and rapid population expansion. Fossil fuel-based sources of energy presently provide the majority of the globe's energy needs. Fossil fuel supplies are, however, running out faster as energy use rises. Alternatives to renewable energy must be created in order to address these issues and the growing need for energy. Renewable energy from the sun, which is made up of solar photovoltaic panels that can produce power instantly, is the most abundant renewable energy source. As a substance for photovoltaic light harvesting, perovskite has gained popularity. Thus, lead-based replacements for the metal cations within the perovskite photo-absorber became the main subject of investigation. Group IV of the Periodic Table offers a simple way to substitute lead by stepping up or down. Below the lead in the row is the recently found element, flerovium, which is rarely radioactively stable. Due to its radioactivity, it could not be a practical replacement for lead. Due to its decreased toxicity, tin (Sn), placed on a similar row as lead, may be a suitable alternative for lead in PSCs. Since known of organic tin halide synthesis, natural lead halide synthesis existed [3].

1.2 Literature Review

Having a deficit in energy (E_g) of 1.3 eV, the MASnI_3 perovskite was previously claimed to be used in the manufacture of solar cells [4]. Overall, the +2-oxidation property of Sn, thus is necessary to form a perovskite, is fragile; whenever it is subjected to air or oxygen moisture, the element rapidly deteriorates to the +4 state, which affects both the device's working conditions and the method used to create solar cells. Since any contact with oxygen can immediately cause tin oxidation, crystalline tin halide PSCs have been merely tested for functionality alongside rigorous equipment covering [5]. CeO_x is regarded to be among the world's most significant rarest oxides due to its wide bandgap, excellent dielectric properties, and excellent conductivity to ions. By employing a straightforward sol-gel procedure, Wang et al. produced CeO_x ($x = 1.87$) coatings at a relatively low temperature (150 °C) and used them as an alternative to TiO_2 ETL that had undergone a high-temperature annealing process. A champion PCE of 14.32% was achieved by the optimized PSC by altering

the CeO_x precursor solution. Yang et al. used CsPbIBr_2 perovskite to serve as the light-collecting materials and CeO_x as the electron transfer layer in an inverted PSC arrangement. The all-inorganic PSC reached an all-time high efficiency of 5.6% with improved stability. According to Jien et al., a CeO_x coating acts to provide a perovskite layer of protection that shields the electrode from moisture and metal reactions [6]. The device employing CeO_x to serve as the ETL provides a PCE of 17.47% [7]. Computerized modeling of the photovoltaic cell would prove necessary to establish the right parameters and physical traits for precision in forecasting. The system is emulated by employing the SCAPS-1 D software, and the effectiveness of a tin-PSC is evaluated in relation to changing MASnI_3 substrate and CeO_x layer characteristics. FTO/ CeO_x / MASnI_3 /2,2',7,7'-tetrakis [N,N-di (4-ethoxyphenyl) amino] is a new device structure. It was previously proposed to simulate a tin-based PSC using the Spiro-OMeTAD using an HTL and 9,9'-spirobifluorene (Spiro-MeTAD)/Au. This computer model aims to demonstrate how altering the ND values associated with CeO_x and the absorber layer thickness (MASnI_3) may enhance the performance of free lead PSCs [8]. In a photovoltaic device, PCBM-doped PEDOT: PSS can function as a reliable transporter for the electron layer. The sunlight-absorbing cell's total efficiency rises as a result of the PCBM molecules' ability to efficiently collect and transport the electrons produced by light absorption. Additionally, PCBM can function as a layer that blocks holes from reassembling into electrons, thereby enhancing effectiveness [9]. The bulk of heterojunction photovoltaic cells, which exhibit impressive efficiency juxtaposed with standard solar cells, are made up of PEDOT: PSS with PCBM.

Proposed analysis

Conventional solar panels made from perovskite frequently include lead, nonetheless due to safety and environmental issues, researchers have been looking for alternatives for replacing lead within these cells. A mixture of an organic-inorganic free of lead perovskite structure, which could be methylammonium tin triiodide or formamidinium tin triiodide, constitutes one of the most intriguing options. These substances resemble lead-based perovskites in terms of physicochemical capabilities, but they don't include lead, thereby rendering them less dangerous and more ecologically friendly. Using non-toxic cesium-based perovskite photovoltaic cells is yet another possibility; cesium lead bromide (CsPbBr_3) is one such material. Fortunately, it comes with an inferior open-circuit voltage and a lower absorbance coefficient. Furthermore, extra substitute materials have been investigated as possible substitutes for lead in solar cells made from perovskite, including copper- and bismuth-based solar cells made from perovskite [10]. It ought to be emphasized that lead replacement for perovskite solar cells continues to be a hot area of study and that these new materials' stability and efficiency are continually being developed and improved. In a perovskite solar cell, PEDOT: PSS that has been doped by perovskite can be employed as a hole transport layer. a combination of their excellent performance and cheap cost, perovskite solar cells—a more contemporary form of solar cell—have shown considerable promise in recent years. PEDOT:

PSS has excellent electrical conductivity therefore order and it can efficiently transport holes created by the uptake of electromagnetic radiation within the perovskite layer, making it a promising hole transport semiconductor for perovskite solar energy cells. The perovskite photovoltaic cell's structural integrity may be enhanced by the PEDOT: PSS layer. Compared to classic solar panels and perovskite solar cells, hybrid solar cells—which combine PEDOT: PSS and perovskite—have a high efficiency. Graphene-doped PEDOT: PSS can be utilized as a layer that transports holes in solar cells. PEDOT: PSS's performance can be improved by using graphene, a single sheet of carbon atoms organized in a lattice that is hexagonal in shape. The addition of graphene fragments to PEDOT: PSS can make the PEDOT: PSS layer more conductive, which boosts the extraction and delivery of holes (either positively charged carriers) produced by the light absorption and boosts the solar cell's overall effectiveness. The increased surface area of graphene can also aid in enhancing and improving the quantity of PEDOT: PSS that can be included in the photovoltaic cell, resulting in more effective charge transfer. When compared to conventional solar cells that use PEDOT: PSS photovoltaic panels, hybrid solar cells—which combine PEDOT: PSS with graphene—have a high efficiency [11].

2 Materials and Methods

2.1 Mathematical Model

Recombination of charge-carrying particles (holes and electrons) in a material made from semiconductors is described theoretically by the Langevin recombination equation. Subsequently relies entirely on the Langevin hypothesis regarding carrier recombination, which posits that the square root of the carrier concentration determines the recombination rate of carriers. The Langevin recombination machine learning model's computational formula can be obtained by:

$$J = B * \text{sqrt}(p * n) \quad (1)$$

In cases in which p is the hole-focused attention, n is the electron-focused attention, B is a constant that's designated to represent the Langevin recombination coefficient for beta absorption, and J is the coefficient of recombination current density [12]. Recombine in actual substances is typically more complicated and can require additional processes, such as Shockley–Read–Hall recombination, hence it should be emphasized that the Langevin recombination framework is an idealized model and cannot be applied to all semiconductors. Learning about the recombination process of carriers for semiconductors particularly plays a key role in determining the effectiveness of photovoltaic cells as well as various electronic components, is possible using the Langevin recombination paradigm [13].

2.2 Shunt Resistance

The obstacle caused by electrical flow in a photovoltaic cell which “shorts” the electrical connection connecting the cell’s front and rear ends is measured. The shunt is a conducting route that permits current to flow across the photovoltaic cell’s productive region, decreasing the cell’s efficiency. Shunting resistance in a thin-film solar panel is brought on by flaws or imperfections in the semiconductor substance that provide a connecting route between both the front and back of the cell. These flaws might be microscopic fractures, vacancies, or other structural flaws that permit current to skip the cell’s active region and lower the amount of current the cell generates [14]. By passing a modest volt over the semiconductor device and detecting the resultant current, the resistance of the shunt may be determined. By reducing the quantity of material flaws, either through better manufacturing techniques or the use of higher-quality materials, the resistance to shunt movement can be decreased. The use of a coating that is anti-reflective on both the outer and inner surfaces of the cell can also aid with lowering shunt resistance by reducing light scattering, which may contribute to improving the quantity of light that gets taken in by the active region of the cell [15]. Shunt conductance is one of the elements that may impact a thin-film solar cell’s overall performance, and lowering it can boost the cell’s efficiency [16].

Series Resistance

Energy is a measurement of the amount of resistance to electrical flow that arises along the course of the current as it passes through the solar cell’s active region. The interaction between the semiconductor and the metallic electrode and the impedance of the semiconductor substance is to blame [17]. Series resistance within a thin-film photovoltaic cell can be brought on by a variety of things, including a semiconductor that is of poor quality, electrodes that make poor contact with the semiconductor, and an uneven semiconductor layer. By applying a modest circuit across the structure of the cell and counting the resultant current, series resistance may be determined [18]. By employing top-notch semiconductor materials, improving the interaction that exists between the semiconductors as well as the electrodes as well as and lengthening the semiconductor layer, the series resistance may be decreased. Additionally, it is possible to lower the series resistance by utilizing advanced production processes like laser scription and laser doping. Series resistivity is one of the elements that influence a thin-film solar cell’s overall performance and lowering it can boost the cell’s efficiency [19].

Numerical Simulation Formula

$$R_{n, p} = \frac{qnp}{\epsilon \epsilon_0} (\mu_n(E) + \mu_p(E)) \tag{2}$$

$$J = qv_n\mu_n\epsilon + qv_p\mu_p\epsilon \tag{3}$$

$$\text{div}(\epsilon\Delta\psi) = -\rho \tag{4}$$

Table 1 Material parameters used in simulation tool

S. No.	Parameters	PEDOT: PSS	P3HT: PCBM	AL	MASNL ₃
1	Electron trap density [11, 20]	1e20	3.8e26	1e18	1e18
2	Hole trap density [13]	1e20	1.4e25	1.2	1.3
3	Electron mobility [10, 21]	1e-05	2.48e-07	4.5	4.17
4	Hole mobility [19]	1e-05	2.48e-07	10	6.5/10
5	Eg [22, 23]	1.3	1.1	1×10^{18}	1×10^{18}
6	Thickness [24, 25]	200 nm	220 nm	100 nm	50 nm

$$J_n = qv_n\mu_n\varepsilon + qD_n \quad (5)$$

$$J_p = qv_p\mu_p\varepsilon - qD_p \quad (6)$$

$$J_{\text{cond}} = J_n + J_p \quad (7)$$

The general durability of perovskite solar cells depends on the choice of stable and compatible Electron Transport Layers (ETL) and Hole Transport Layers (HTL). To replace traditional HTL and ETL options, researchers are constantly examining new substances with enhanced stability and efficacy. Improved film perovskite shape overall crystallization may be achieved by post-deposition techniques with the value of chemical design as well as thermal annealing which boosts stability and effectiveness. These processes enhance charge transmission and improve the perovskite layer's performance. Identification of stability problems and degradation pathways can be aided by carrying out comprehensive tests over time and faster aging studies. It is possible for academics to create focused tactics to successfully address stability issues through an awareness of their deterioration mechanisms (Table 1).

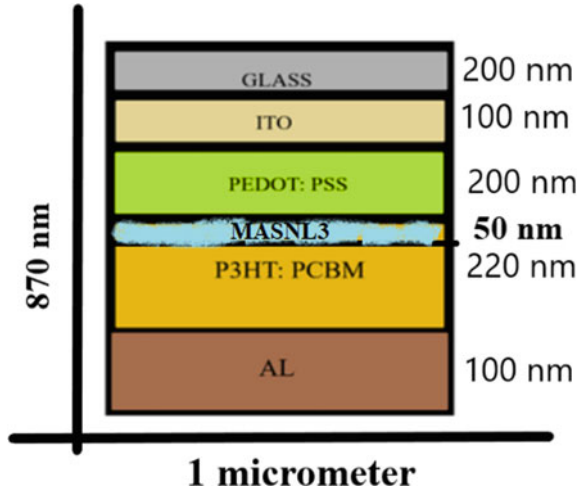
Structure

Figure 1 represents the schematic structure of the simulation design used to analyze efficiency. Here we take Al as the base layer, MASNL₃ is a sandwich between the active layer PEDOT: PSS and P3HT: PCBM. In this structure, we take ITO with Glass for more accelerating velocity for light absorption. The height of the structure size is 870 nm.

3 Results and Discussion

The mathematical modeling of an intricate multilayered architecture made of Glass/ITO/PEDOT: PSS/MASNL₃/P3HT: PCBM/Al constitutes the focus of this study. The P3HT: PCBM has a wavelength of 220 nm. The IV characteristics, quantization

Fig. 1 IV characteristics
Glass/ITO/PEDOT: PSS/
 $MASn_3$ /P3HT: PCBM/Al



efficacy, and fill factor were all obtained. As is frequently observed, improving PCEs and raising V_{oc} result from increasing dopant density. On the other hand, increasing the dopant thickness in the absorber part of HFRC and HFMC above cm^{-3} affects the recombination rate, which causes a drop in JSC.

3.1 IV Curve

We generated a 27.8% accuracy in Fig. 2. With the standard settings of SCAPS-1D software, we model the structure in 3 situations: first in AM1.5D, then in normal, and finally in under sun circumstances.

3.2 Quantum Efficiency

The line diagram in Fig. 3 above shows the efficiency of quantum systems as a function of wavelength. The fundamental structure of Glass/ITO/PEDOT: PSS/P3HT: PCBM/Al is used in the first study, and this was where we found the lower quantum conversion rate (Black), which was close to 56%. With the addition of $MASn_3$ within the combination of PEDOT: PSS/P3HT: PCBM and normal conditions in the first circumstance, the second circumstance was explored. The graph's blue hue denotes the structure's 84% efficiency, which indicates an enormous growth in the quantum level. The 3rd scenario yielded 111.21% in quantum efficiency and has the same construction as the sun state, which is shown in blue.

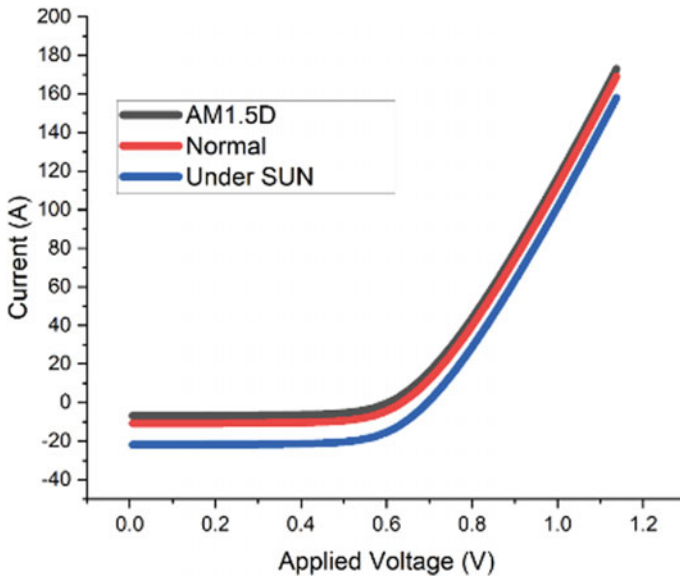


Fig. 2 IV curve current versus voltage

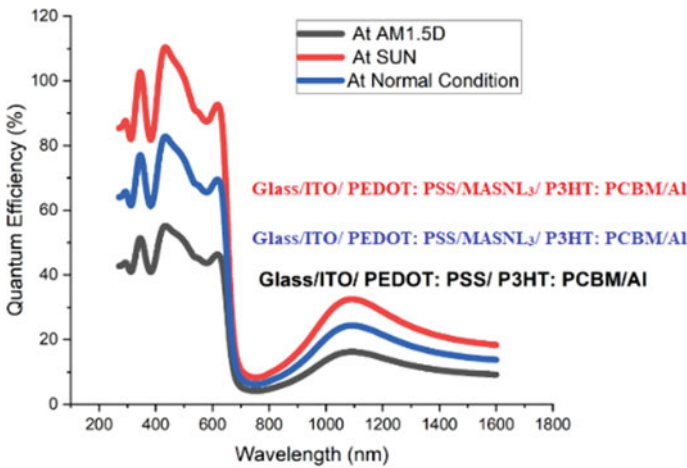


Fig. 3 Quantum efficiency versus wavelength (nm)

Generation rate

The emergence velocity of the electron in the layer that is active is shown in Fig. 4. In recent years, lead halide perovskites have shown tremendous promise as absorbent compounds for application in photovoltaic devices, demonstrating their

broad potential. The materials have an astonishing 24.2% efficiency in converting energy.

Under sun condition

The plots in Fig. 5 show how the structure’s absorption and electron dissipation. Since MASn₃ is very conductive, the electron may flow through it quickly.

Absorbed photon density

In Fig. 6 the graphs show the intensity of photon density absorption in the 250–650 nm.

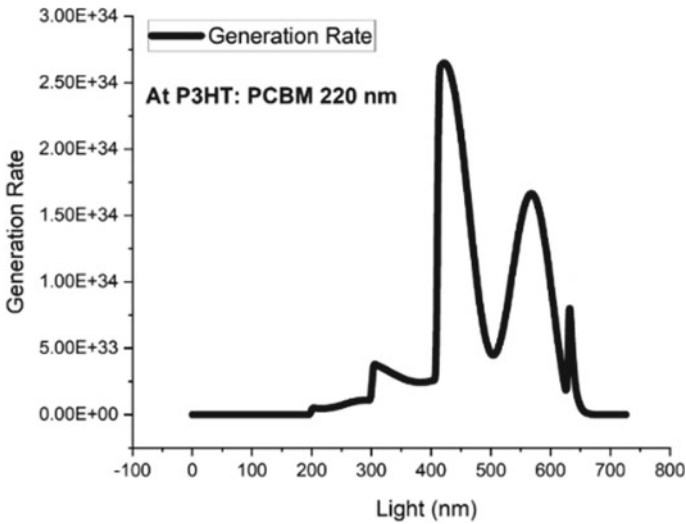


Fig. 4 Generation rate versus light (nm)

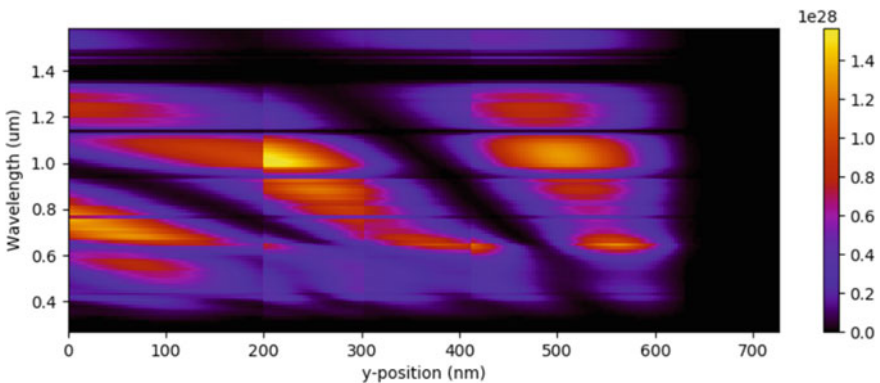


Fig. 5 Wavelength versus distance (nm)

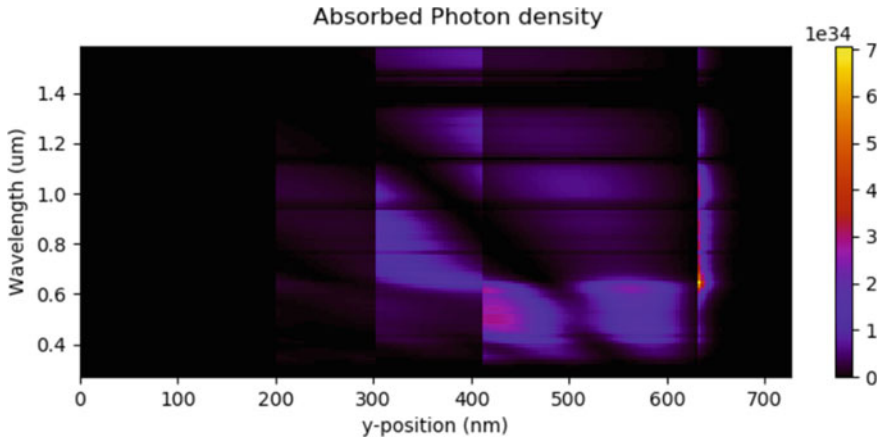


Fig. 6 Wavelength versus distance (nm)

4 Conclusion

In the end of the aim of this examination was achieved. A PCE efficiency of 27.8% was attained. When using Glass and ITO, the quantum efficiency rises to 110%. Three alternative scenarios were used for the analysis: AM1.5D, regular circumstances, and sun. The fill factor for the structure is 0.75. Furthermore, the point of origin photon's wavelengths and momentum are both improved by up to 500 nm, along with the internal translational efficacy.

References


1. Brinkmann K, Becker T, Zimmermann F, Kreusel C, Gahlmann T, Theisen M, Haeger T, Olthof S, Tücmantel C, Günster M (2022) Perovskite–organic tandem solar cells with indium oxide interconnect. *Nature* 604(7905):280–286
2. Ding Y, Ding B, Kanda H, Usiobo OJ, Gallet T, Yang Z, Liu Y, Huang H, Sheng J, Liu C (2022) Single-crystalline TiO₂ nanoparticles for stable and efficient perovskite modules. *Nat Nanotechnol* 1–8
3. York R, Bell SE (2019) Energy transitions or additions?: Why a transition from fossil fuels requires more than the growth of renewable energy. *Energy Res Social Sci* 51:40–43
4. Chapman AJ, McLellan BC, Tezuka T (2018) Prioritizing mitigation efforts considering co-bene, equity and energy justice: fossil fuel to renewable energy transition pathways. *Appl Energy* 219:187–198
5. Singh KAS, Mohammed MK, Shalan AE (2021) Effect of 2D perovskite layer and multivalent defect on the performance of 3D/2D bilayered perovskite solar cells through computational simulation studies. *Sol Energy* 223:193–201
6. Al-Mousoi A, Mehde M, Al-Gebori A (2020) Annealing temperature effects on the performance of the perovskite solar cells. In: *IOP conference series: materials science and engineering*. IOP Publishing, p 012039

7. Humadi MD, Hussein HT, Mohamed MS, Mohammed MK, Kayahan E (2021) A facile approach to improve the performance and stability of perovskite solar cells via FA/MA precursor temperature controlling in sequential deposition fabrication. *Opt Mater* 112:110794
8. Correa-Baena J-P, Saliba M, Buonassisi T, Grätzel M, Abate A, Tress W, Hagfeldt AJS (2017) Promises and challenges of perovskite solar cells. *Science* 358(6364):739–744
9. Kadhim MJ, Mohammed MK (2022) Fabrication of efficient triple-cation perovskite solar cells employing ethyl acetate as an environmental-friendly solvent additive. *Mater Res Bull* 112047
10. Mohammed MK, Al-Mousoi AK, Majeed SM, Singh S, Kumar A, Pandey R, Madan J, Ahmed DS, Dastan D (2022) Stable hole-transporting material-free perovskite solar cells with efficiency exceeding 14% via the introduction of a malonic acid additive for a perovskite precursor. *Energy Fuel* 36(21):13187–13194
11. Kareem SH, Elewi MH, Naji AM, Ahmed DS, Mohammed MK (2022) Efficient and stable pure a-phase FAPbI₃ perovskite solar cells with a dual engineering strategy: additive and dimensional engineering approaches. *Chem Eng J* 136469
12. Mohammed MK, Jabir MS, Abdulzahraa HG, Mohammed SH, Al-Azzawi WK, Ahmed DS, Singh S, Kumar A, Asaithambi S, Shekargoar M (2022) Introduction of cadmium chloride additive to improve the performance and stability of perovskite solar cells. *RSC Adv* 12(32):20461–20470
13. Moharam MM, El Shazly AN, Anand KV, Rayan DE, Mohammed MK, Rashad MM, Shalan AE (2021) Semiconductors as effective electrodes for dye sensitized solar cell applications. *Top Curr Chem* 379(3):1–17
14. Al-Mousoi AK, Mohammed MK (2020) Engineered surface properties of MAPI using different antisolvents for hole transport layer-free perovskite solar cell (HTL-free PSC). *J Sol-Gel Sci Technol* 96(3):659–668
15. Mohammed MK, Al-Mousoi AK, Singh S, Younis U, Kumar A, Dastan D, Ravi G (2022) Ionic liquid passivator for mesoporous titanium dioxide electron transport layer to enhance the efficiency and stability of hole conductor-free perovskite solar cells. *Energy Fuel* 36(19):12192–12200
16. Kojima A, Teshima K, Shirai Y, Miyasaka T (2009) Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. *J Am Chem Soc* 131(17):6050–6051
17. Cui P, Wei D, Ji J, Huang H, Jia E, Dou S, Wang T, Wang W, Li MJNE (2019) Planar p–n homojunction perovskite solar cells with efficiency exceeding 21.3%. *Nat Energy* 4(2):150–159
18. Min H, Lee DY, Kim J, Kim G, Lee KS, Kim J, Paik MJ, Kim YK, Kim KS, Kim MGJN (2021) Perovskite solar cells with atomically coherent interlayers on SnO_2 electrodes. *Nature* 598(7881):444–450
19. Mehde MS, Al-Gebori AM, Hantoosh AK (2020) The effect of the spinning speed variation on the perovskite solar cell efficiency. *IOP Conf Ser: Mater Sci Eng* 012071
20. Hoer SF, Trimmel G, Rath T (2017) Progress on lead-free metal halide perovskites for photovoltaic applications: a review. *Monatsh Chem-Chem Monthly* 148(5):795–826
21. Sun N, Gao W, Dong H, Liu Y, Liu X, Wu Z, Song L, Ran C, Chen Y (2021) Architecture of pin Sn-based perovskite solar cells: characteristics, advances, and perspectives. *ACS Energy Lett* 6(8):2863–2875
22. Liu X, Yang Z, Chueh C-C, Rajagopal A, Williams ST, Sun Y, Jen A (2016) Improved efficiency and stability of Pb–Sn binary perovskite solar cells by Cs substitution. *J Mater Chem A* 4(46):17939–17945
23. Kayesh ME, Matsuishi K, Kaneko R, Kazaoui S, Lee J-J, Noda T, Islam AJAEL (2018) Coadditive engineering with 5-ammonium valeric acid iodide for efficient and stable Sn perovskite solar cells. *ACS Energy Lett* 4(1):278–284
24. Hao F, Stoumpos CC, Cao DH, Chang RP, Kanatzidis M (2014) Lead-free solid-state organic–inorganic halide perovskite solar cells. *Nat Photon* 8(6):489–494
25. Fang R, Wu S, Chen W, Liu Z, Zhang S, Chen R, Yue Y, Deng L, Cheng Y-B, Han L (2018) [6, 6]-Phenyl-C61-butiric acid methyl ester/cerium oxide bilayer structure as efficient and stable electron transport layer for inverted perovskite solar cells. *ACS Nano* 12(3):2403–3241

26. Mohammed MK, Al-Mousoi AK (2016) Deposition of multilayer graphene (MLG) on glass slide by same synthesis technique. *Optik* 127(20):9848–9852
27. Mohammed MK, Al-Mousoi AK, Mehde MS, Al-Gebori AM (2020) Engineered electronic properties of the spin-coated MAPI for hole-transport-free perovskite solar cell (HT-free PSC): spinning time and PSC performance relationship. *Chem Phys Lett* 754:137718
28. Ahmed DS, Mohammed BK, Mohammed MK (2021) Longterm stable and hysteresis-free planar perovskite solar cells using green antisolvent strategy. *J Mater Sci* 56(27):15205–15214
29. Cao J, Yan FJE (2021) Recent progress in tinbased perovskite solar cells. *Energy Environ Sci* 14(3):1286–1325
30. Schileo G, Grancini G (2021) Lead or no lead? Availability, toxicity, sustainability and environmental impact of leadfree perovskite solar cells. *J Mater Chem C* 9(1):67–76

Sensorless PMSM Motor Control Strategies by Artificial Neural Network (ANN) with PI Controller



Sagar Appaso Talke, Arshad Mansur Tamboli, Yashraj Mohan Shinde, and Dipak Suresh Yeole 

Abstract Due to benefits in terms of cost, complexity, and reliability, sensorless control of BLDC motors has attracted a lot of interest. In order to obtain effective and precise control of a sensorless BLDC motor, a control strategy that combines the skills of an artificial neural network (ANN) and a proportional integral (PI) controller is needed. Traditional BLDC motor control techniques call for sensors to measure things like rotor position or speed. Sensor use, however, raises system complexity and expense. As opposed to this, sensorless control systems estimate the rotor position and speed using the motor's back-EMF voltage or current. This eliminates the requirement for additional sensors. In this paper, the rotor location and speed of the BLDC motor are estimated using an ANN. An electrical motor signal dataset and a dataset of known rotor locations are used to train the artificial neural network (ANN). The ANN may learn the correlation between the electrical data and the rotor position during the training procedure. Without the use of physical sensors, the ANN can estimate the rotor location and speed in real time once it has been trained. A PI controller is coupled with the ANN-based estimator to improve control performance. The PI controller generates control signals to modify the phase currents in the motor by comparing the estimated rotor position and speed to the intended values. With minimum inaccuracy, the motor can track the intended speed or position, thanks to excellent motor control. On a prototype BLDC motor, the suggested control strategy is put into practice and tested. The results of the experiments show how successful the sensorless control approach is based on an ANN with PI controller.

Keywords BLDC motor · Sensorless · ANN · PMSM

S. A. Talke · A. M. Tamboli · Y. M. Shinde

UG Students, Electrical Engineering Department, Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati, Pune 413 133., Maharashtra, India

D. S. Yeole (✉)

Assistant Professor, Electrical Engineering Department, Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati, Pune 413 133., Maharashtra, India
e-mail: dipakyeole79@gmail.com

1 Introduction

A BLDC electric motor, sometimes referred to as an electronically commutated motor (ECM or EC motor), or a BLDC motor, is defined as a “synchronous” sort because the magnetic field generated by the stator and the rotor revolve at the same frequency. The benefit of this construction is that it eliminates the “slip” that induction motors are notorious for. BLDC motors are currently a crucial component of contemporary drive technology. A larger range of industries, including consumer electronics, the automobile industry, industrial automation, chemical and medical, as well as aerospace and instrumentation are employing them as a result of their rapid rise in popularity. The sub-kilowatt range, which has been dominated by brushed DC motors, has always been ambiguous despite the fact that they have long been used for drives and power generation. The development of power electronics and micro-processors has allowed small brushless DC motors to flourish in terms of cost and performance, nevertheless. The biggest benefit of brushless DC motors is that they don’t need any maintenance. Also, this BLDC can be used in EV [1].

1.1 Design and Operational Tenets

This motor’s design is very similar to both traditional DC motors and three-phase induction motors. Like all other motors, this one has a stator and a rotor. BLDC motors can be built in a variety of physical arrangements. These can be set up as single-phase, two- phase, or three-phase motors depending on the stator windings. The most often used motors, though, are three-phase BLDC motors with permanent magnet rotors. The constructional features of BLDC motors are shown in Figs. 1 and 2a, b.

The stator of a BLDC motor has layers of steel lamination that carry the windings. These windings are located in axially recessed locations along the inner circumference of the stator. These windings can be arranged in star or delta shape. However, most BLDC motors use a star-connected three-phase stator. Each winding consists of several coils interconnected with one or more coils placed in each. Each winding is broken around the stator to provide an even frame. Permanent magnets are placed in the rotor of the BLDC motor. Depending on the needs of the application, the number of poles of the rotor can vary from 2 to 8 pole pairs with alternating north and south poles. In order for the motor to produce high torque, the flux density of the material must be high. To create a suitable magnetic field, the rotor must be made of suitable magnetic material.

Ferrite magnets are inexpensive, but they have a low flux density for a given volume. Magnets made from rare Earth metals are often used in new designs. Samarium cobalt (SmCo), neodymium (Nd), and ferrite and boron (NdFeB) are some of these metals.

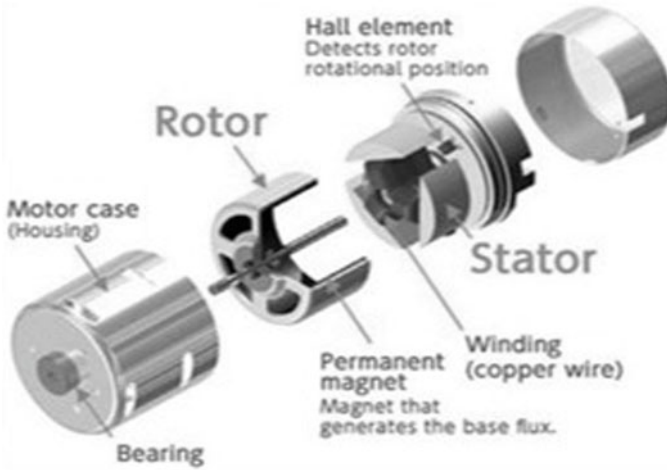


Fig. 1 BLDC motor structure

Basically, there are two ways to build a BLDC motor: one is to place the rotor outside the base and the windings inside, and the other is to place the windings outside the base. In the first configuration, the rotor magnets act as insulators, dissipating heat by slowing the speed of the motor, and the motor runs at low current.

It is mostly used for fans. The motor produces less electricity in the second configuration, which increases its torque. It is an essential part of the hard disk.

Hall sensors for position in a BLDC motor, rotation is electrically controlled as there are no brushes. To turn the motor, the stator windings must be turned on sequentially, and the position of the rotor's north and south poles must be known to open the correct stator winding set. The position of the rotor is usually sensed using a sensor, usually a Hall sensor (which works on the concept of the Hall Effect and is converted into an electrical signal).

Operating Theory

A BLDC motor works similar to a DC motor, a BLDC motor operates on the principle that a current-carrying conductor encounters a force each time it is placed in a magnetic field. Due to the reaction force, the magnet will be stable. In a BLDC motor, the magnets are in constant motion, while the current-carrying wires remain stationary.

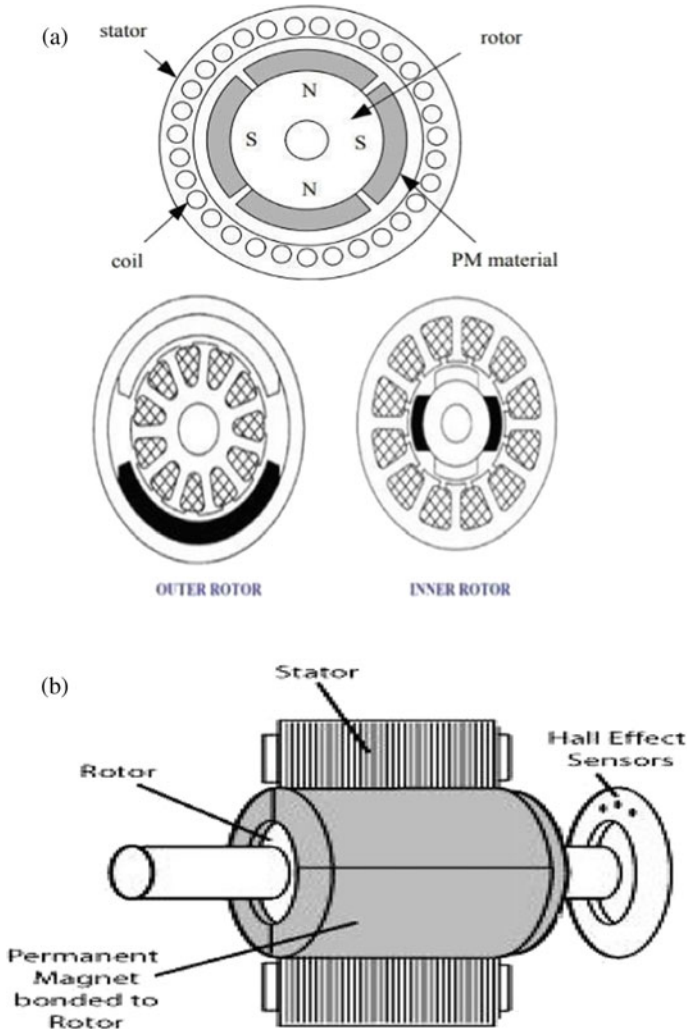


Fig. 2 a BLDC rotor structure b BLDC motor parts

2 Artificial Neural Networks (ANNs) Using PI Controller Approach

The input, output, and hidden layers of the artificial neural network (ANN) are simple networks containing a series of nodes. Because ANNs can process more binary data, their use in Electrical and Electronics Engineering consistently yields desired results. Brushless DC motors (BLDC motors) use an electric motor to transfer DC current to the motor windings and create a magnetic field. BLDC motors are versatile due to their high speed, low maintenance costs, and high torque capacities.

This motor is preferred over other motors because of its better performance and ease of speed control via the power switch. This article describes a technique for controlling the speed of a BLDC motor by varying the DC input voltage to the bridge converter that drives the motor windings. PI-based speed control is used for control. The motor is modeled in MATLAB/Simulink and a PI controller is used to provide speed control. Simulate EMF signals, rotor speed, electromagnetic torque, Hall Effect signals, PWM, and EMF signals. The obtained data is processed using a binary neural network, so the ANN model can predict adverse effects. Arithmetic-based simulations and data-based predictions perform well [2, 3].

Deep learning is implemented using neural networks, a type of biology-inspired neural network. A neuron is a group of connected points that form an artificial neural network (ANN). Information processing and transmission are similar to the way neurons and synapses in the human brain transmit and process information, respectively. Input processes, output processes, and most importantly many hidden processes make up a large part of ANN models. Input neurons receive inputs (data), weight them, and combine the weights. It is then added and sent to the next neuron using the activation function [4].

This process goes in a circle. Through this iterative process, the model can learn from ideas and complete complex tasks. As mentioned earlier, the process begins by applying weights to the information transmitted between neurons. The weight change determines the main working ability of the neural network.

When training the neural network with training data, the weights are adjusted. Each time the model is trained, weights are adjusted to help the model understand and identify patterns in the input data. Through training, a trained neural network can predict new information fed into it. Refer Fig. 3a–c.

The truth depends on many parameters. It just depends on the size of the neural network, the size of the dataset, the activation function used, and the type of dataset.

The neural network's propagation is forward propagation, and backward propagation principles. Data is delivered into the forward propagation process. A sample with characteristics is taken. The result is based on these characteristics. The functions of summation and activation are then carried out. In the procedure, the weights are added together. Then it is subjected to an activation function and sent to a subsequent hidden layer. In the equation below, I represents the number of inputs, W represents the given weights, and X represents the feature (data) [5].

$$Y = W_i X + b$$

The basic goal of backward propagation is to lower the overall error. Backward propagation, as the name implies, starts with error measurement and ends with weight modifications to lower the error. Test it by enabling the function and see the full error. Back propagation uses the chain rule, identify and correct the weighting error that is the cause of the problem, as seen below.

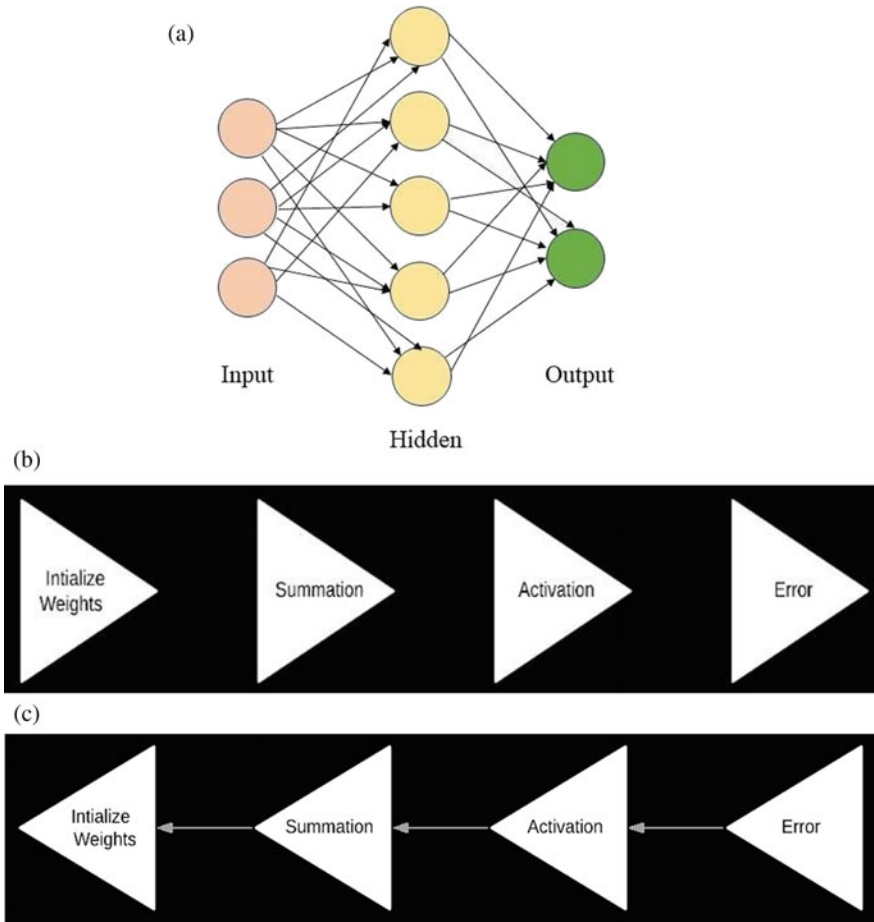


Fig. 3 a Structure of a basic artificial neural network. b Structure of an artificial neural network. c Flow diagram of backward propagation

$$dE/dW_i = dE/dA * dA/dS * dS/dW_i$$

Lack of Robustness: ANNs are susceptible to hostile attacks or input data noise. It is possible for outputs to differ greatly or for categorization to be incorrect with even minor input changes. Preserving ANNs' robustness and dependability in the presence of unexpected or hostile inputs an area of study that is active.

Memory and Scalability Requirements: Large-scale ANNs with millions of parameters may need a lot of memory and processing power to train and deploy. The storage and computational demands of deploying ANNs in resource-constrained environments can pose challenges [6].

It is important to consider these disadvantages alongside the advantages when deciding to use ANNs for a particular task or problem.

ANN benefits include:

1. **Adaptive Learning:** ANNs are capable of learning from their past performance or training data. They are suitable for activities with changing conditions or changing data because they can automatically modify their internal parameters to improve their performance over time.
2. **Nonlinear Modeling:** ANNs may capture sophisticated nonlinear dependencies and patterns in the data by capturing complex nonlinear correlations between input and output variables. They can thus be used for a variety of applications involving nonlinear events or intricate data distributions.
3. **Parallel Processing:** ANNs have the ability to process information in parallel, allowing them to run several computations at once. Because of their ability to handle data in parallel, ANNs are effective for time-sensitive or large-scale applications.
4. **Fault Tolerance:** ANNs have fault tolerance, which allows them to continue operating and producing outputs that are reasonably accurate even in the presence of noisy or imperfect data or when some of the network's components fail.
5. **Generalization:** ANNs can extrapolate from training data to make predictions or categorize previously unknown or novel situations. They can make judgments and infer patterns from data outside of the examples they were trained on, which enable them to handle tasks with ambiguous or variable inputs.

In a variety of disciplines and sectors, artificial neural networks are a powerful and adaptable method for resolving complicated issues, learning from data, and making wise decisions.

ANN drawbacks include:

1. **Training Complexity and Computational Requirements:** When training large or complicated networks or datasets, ANNs can be computationally time-consuming and labor-intensive. Iterative optimization approaches may be used throughout the training process, which can be time-, money-, and energy-intensive. Training normally calls for a lot of computational resources.
2. **The Need for Enough Training Data:** For ANNs to learn relevant patterns and correlations, they often need a sizable amount of training data. Poor generalization and unreliable predictions might result from insufficient or unrepresentative training data. Large and diverse datasets can be difficult to acquire and prepare, which could hinder ANN performance.

PI controller:

Proportional plus integral (PI) controller is another name for controller. It is a controller with proportional control action, so it is called PI controller. Proportional integral controller uses both proportional and integral controllers to control motion. When two different controllers are combined, a better controller is created to address the shortcomings of each controller [7, 8].

In this case, the control signal is proportional to the error signal and its value. The numerical representation of the proportional plus integral controller is as follows:

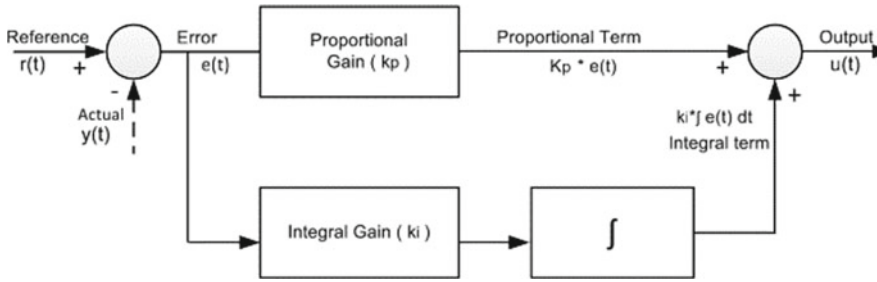


Fig. 4 Block diagram

$$m(t) = K_p e(t) + K_i \int e(t)$$

Block Diagram:

Figure 4 represents the block diagram of PI controller.

Advantages of PI Controller:

1. The output is inversely proportional to the integral of the input signal, which is one benefit of the PI controller.
2. The type of the system expands as the integrator gets more involved. As a system type grows, accuracy rises as steady-state error decreases.
3. Integral action in PI controllers eliminates offset, a significant drawback of a P-only controller. Thus, PI controllers are the algorithms that process control applications use the most because they provide a balance between complexity and capacity.
4. As the system's type is enlarged, there are some detrimental consequences on the stability of the system.
5. PI control is a kind of feedback control. A speedier response time than I-only control is achieved by adding the proportionate action.

Disadvantages of PI Controller:

1. The technique requires only moderate load modifications to eliminate oscillations caused by the integral overshoot, yet this is necessary due to the integration duration.
2. Before settling to the operational point, the integral action during the initiation of a batch process creates a large overshoot of the error and output.

3 Simulation Model

Simulation model for sensorless BLDC motor control by artificial neural network (ANN) with PI controller method is as shown in Fig. 5.

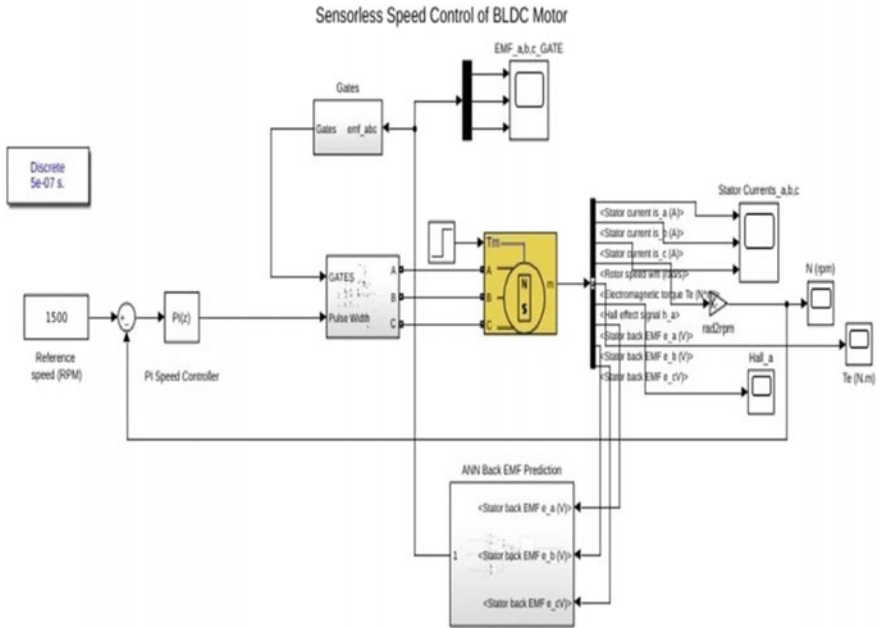


Fig. 5 Simulation model for sensorless BLDC motor control by artificial neural network (ANN) with PI controller

4 Result

The speed of a sensorless BLDC motor is controlled by combining an ANN with the PI controller approach. Reference speed in a simulation model is set to 1500 rpm. It displays the outcome that machine speed response comes close to 1500 rpm approximately and also provides electromagnetic torque generated during motor operation (Figs. 6 and 7).

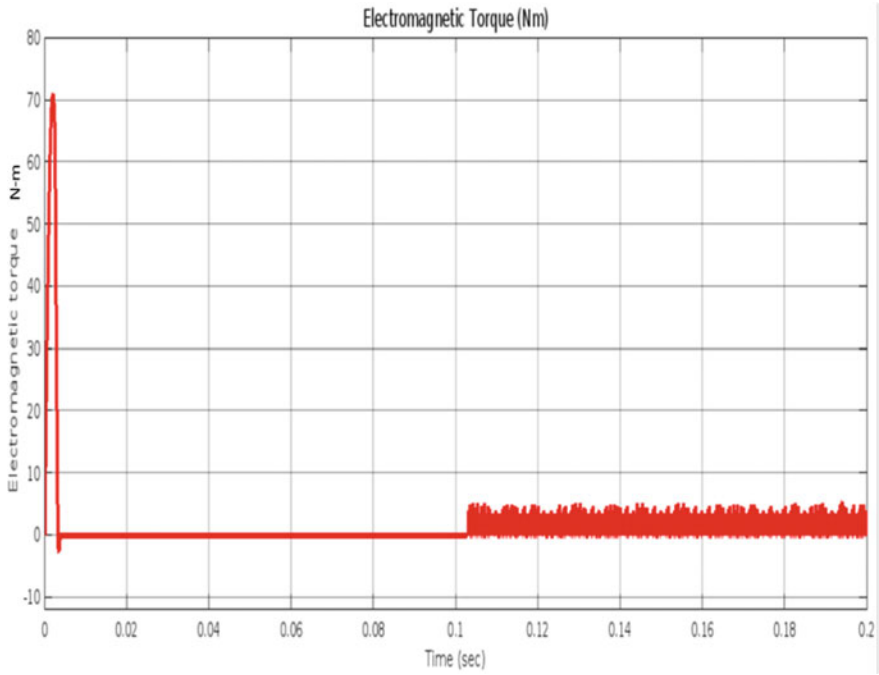


Fig. 6 Time versus electromagnetic torque

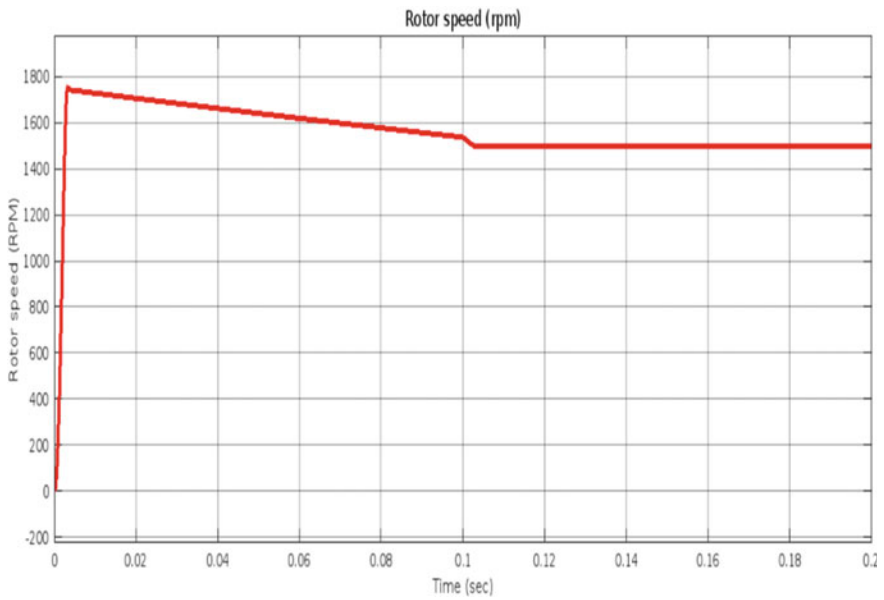


Fig. 7 Time versus rotor speed

5 Conclusion

In conclusion, a promising method with a number of benefits is provided by the coupling of an ANN with a PI controller for sensorless brushless DC (BLDC) motor control. Without the use of extra sensors, an ANN can estimate rotor position and speed, which lowers system costs, complexity, and possible weak points.

The ANN is more resilient and able to handle various operating situations since it can learn from and adapt to changes in motor parameters. By providing feedback control based on the discrepancy between the desired and estimated values, the PI controller supports the ANN. It aids in system stabilization, steady-state error reduction, and general motor control performance improvement. The BLDC motor can be precisely controlled, achieving exact speed and position tracking, thanks to the integration of an ANN and PI controller. The control algorithm is optimized by the ANN to increase efficiency and reactivity by learning from historical data and modifying its parameters accordingly. The sensorless technique also simplifies the design of the motor, lowers maintenance needs, and increases dependability by doing away with the need for mechanical sensors like Hall Effect sensors or encoders.

Overall, the sensor-free BLDC motor control system using an ANN and a PI controller provides a reliable, accurate, and affordable solution for a variety of applications. It achieves high-performance motor control without the use of extra sensors by combining the advantages of cutting-edge machine learning techniques with traditional control algorithms.

References

1. Thakre MP, Mahadik YV, Yeole DS Potentially affect of a vehicle to grid on the electricity system. IOP Conf Ser: Mater Sci Eng 1084(1):012077. <https://doi.org/10.1088/1757-899X/1084/1/012077>
2. Li C, Yu W, Tang W (2010) Study on rotor position of sensorless brushless DC motors through back electromotive force detection. Presented at International conference on E-product E-service and E-entertainment (ICEEE), 7–9 Nov 2010
3. Xiong H, Xue Y (2010) The design of Brushless DC motor back-EMF control. Presented at International conference on environmental science and information application technology (ESIAT), 17–18 July 2010
4. Nikhil SS, Sampath K (2019) Neural network based BLDC motor speed control. Int J Electr Electron Data Commun 7(10). ISSN(p): 2320-2084, ISSN(e): 2321-2950. <http://iraj.in>
5. Mamadapur A, Mahadev GU Speed control of BLDC motor using neural network controller and PID controller. Electrical Engineering Department Zeal College of Engineering and Research Pune, India
6. Leena N, Shanmugasundaram R, Member, IEEE (2014) Artificial neural network controller for improved performance of Brushless DC motor. In: International conference on power, signals, controls and computation (EPSCICON), 8–10 Jan 2014

7. Mirtalaei SMM, Moghani JS, Malekian K, Abdi B (2008) A novel sensorless control strategy for Bldc motor drives using a fuzzy logic based neural network observer. In: Speedam 2008 international symposium on power electronics, electrical drives, automation and motion
8. Wongkhead S (2021) State space model for speed control Bldc motor tuning by combination of Pi-artificial neural network controller. In: Ecti-Con 2021-smart electrical systems and technology

A Comprehensive Review on Electric Vehicle Battery Swapping Stations



Md Tahmid Hussain, Ahmad Bin Afzal, Altaf Hussain Thakurai, Ahmar Azim, Rashid Ahmed khan, Hasan Alrajhi, Mohd Rizwan Khalid, and Salman Hameed

Abstract This paper comprehensively reviews electric vehicle (EV) battery swapping stations (BSS), an emerging technology that enables EV drivers to exchange their depleted batteries with fully charged ones at designated stations. The paper aims to comprehensively understand BSS's technical, economic, and environmental aspects and its potential for widespread adoption. The review covers BSS design, operation, and maintenance, including the necessary infrastructure, battery management systems, and safety protocols. It also analyzes the economic viability of battery swapping compared to other charging technologies, taking into account factors such as capital and operational costs, revenue streams, and return on investment. Further discusses the environmental impact of battery swapping, including the potential reduction in carbon emissions, energy consumption, and resource depletion. Finally, the study examines the current state of BSS, including market trends, regulatory frameworks, and stakeholder engagement. The review concludes that BSS holds significant promise as a sustainable and convenient solution for EV charging. Still, several challenges remain to be addressed, including standardization, interoperability, and consumer acceptance. The findings of this study provide valuable insights for policymakers, industry stakeholders, and researchers working on the development and deployment of EV BSS.

Keywords Battery swapping techniques · Battery-to-grid · Fast charging · Range anxiety · Vehicle-to-grid

M. T. Hussain · A. B. Afzal · A. H. Thakurai · A. Azim · M. R. Khalid (✉) · S. Hameed
Department of Electrical Engineering, Aligarh Muslim University, Aligarh 202002, India
e-mail: mrk.rizwankhalid@gmail.com

R. A. khan
Department of Electronics and Electrical Engineering, University of Sheffield, Sheffield, England

H. Alrajhi
Department of Electrical Engineering, Umm Al-Qura University, Mecca, Saudi Arabia

1 Introduction

The transition to electric vehicles (EVs) is critical to global efforts to reduce greenhouse gas emissions and mitigate climate change. While EV sales have grown recently, limited driving range and long charging times remain critical barriers to broader adoption. This is where battery swapping stations (BSS) come in, offering solutions to these challenges. The need for BSS in EV technology is driven by the growing demand for more efficient, convenient, and sustainable transportation solutions. BSS has the potential to offer a viable alternative to traditional charging methods while also addressing some of the challenges associated with EV adoption. BSS offers several advantages over conventional charging stations, including faster charging times, increased driving range, reduced infrastructure costs, reduced carbon footprint, new business models and revenue streams, and increased consumer convenience. As such, BSS is an area of significant interest and investment for policymakers, industry stakeholders, and researchers working on developing and deploying EV infrastructure. These points further explain the need for BSS in EV technology.

Range Anxiety

Drivers of EVs often worry about being stranded due to their restricted range on a single charge. As a result, drivers worry about the remaining charge in the battery which they are using. To avoid this problem, battery switching stations have been set up so that drivers may quickly and easily switch out their dead batteries for fresh ones and keep driving [1].

Infrastructure

Domestic migration propels the transition from rural to urban living in developing nations. Nearly everyone wants to live in a city, yet these areas weren't always adequately created and designed for the millions of new citizens. These cities were propelled by globalization into freshly established economic centers. Thus, finding fixed parking locations to charge EVs is consequently a challenge. The lack of charging stations for EVs can be partially remedied by setting up BSS. Locations for stations might be chosen to provide convenient access to charging infrastructure for drivers. This can be enormously beneficial in remote or rural places, where there may not be easy access to a charging station [2].

Time-Saving

Faster charging is possible with BSS than with standard outlets. Using conventional methods to charge an EV completely can take many hours, while battery swapping only takes minutes. This can alleviate some of the frustration drivers experience while waiting for their cars to charge [3].

Cost Effective

Charging stations for EVs can be expensive to install. Setting up a battery switching station may be costly, but over time, it may be cheaper than conventional charging

methods. This is because, with regular battery swaps, batteries may be used more efficiently and for more extended periods [4].

Environmental Impact

While battery recycling cuts down on manufacturing, it can help the planet. This is due to the fact that battery switching can improve battery life, thereby reducing the overall need for new batteries. The adverse effects on the environment from making and disposing of batteries can be mitigated in this way [5].

The following is the remainder of the paper: Sect. 2 describes the different types of battery swapping techniques. Section 3 describes the functioning of BSS, and the problems associated with BSS have been described in Sect. 4. The impact of BSS on the distribution network and the usage of BSS as a battery to load (B2L) during peak loads are described in Sects. 5 and 6, respectively. Finally, Sect. 7 brings the conclusion of the paper.

2 Battery Swapping Techniques

To maintain electric transportation, the BSS replaces depleted batteries that have fallen below their predetermined state-of-charge (SoC) level with fully charged ones, and this process ensures the continuous operation of EVs. In the early nineteenth century, people proposed swapping batteries to solve the limited range of EVs like trucks and cars [6]. The concept involved using interchangeable batteries that could be replaced when necessary, and this replacement process was carried out using human labor. The initial implementation of a commercial battery exchange service that utilized the swapping strategy for EVs was called Better Place [7]. The utilization of the battery swapping method was restricted solely to electric cars and did not extend to heavy-duty vehicles like buses and trucks. As a result of requiring significant capital investment to create charging and swapping infrastructure and facing limited market penetration, Better place experienced substantial financial losses, leading to its eventual liquidation. In 2008, China commercially implemented the battery swapping method for electric buses during the Summer Olympics and switched the batteries of approximately 50 buses running on different routes [8]. Following the initial implementation during the 2008 Summer Olympics, the bus battery swapping method has been broadly adopted in various countries, including South Korea, Japan, China, and others. The different battery swapping methods are distinguished based on the location of the battery within the vehicle and the position at which the robotic arm is applied. Figure 1 shows different battery swapping techniques, these are listed below:

A *Sideways Battery Swapping*

When changing the batteries in an EV, the exhausted battery pack is horizontally pushed out from underneath the car and is replaced with a fully charged one. This

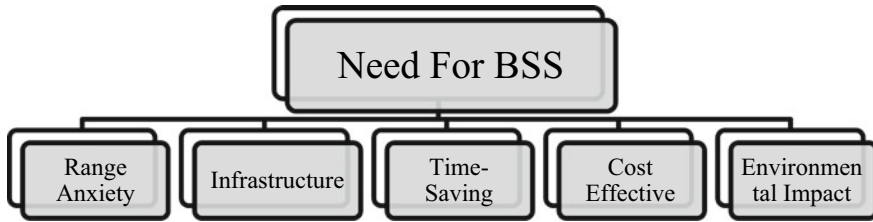


Fig. 1 The need for BSS technology

technique is known as sideways battery swapping. Additionally, since the battery can be swapped from the side, it does not require the space and infrastructure needed for a vertical battery swapping station, which can be more expensive to install. It is commonly utilized for vans and other vehicles best suited for a sideways position during battery swapping.

B Bottom Battery swapping

This is used in cars with bottom-mounted batteries. The automobile is placed on a raised platform by the swapping station, and the batteries are changed from the bottom using an automation arm and other peripherals that are often located below ground. The station is constructed to raise the car on an elevated platform. At the same time, the battery exchange occurs from below using a robotic arm and various devices, typically positioned beneath the ground level.

C Rear Seat Battery Swapping

In order to use this method, the vehicle's rear seats must be removed, and the spot where the seats once were must be filled with a battery pack. Tesla had suggested using this approach, but the business eventually decided against it. This battery swapping technique applies to vehicles with batteries placed toward the back, commonly seen in cars with spacious trunks.

D Vertical Battery Swapping

This is the most common battery swapping technique. It involves lifting the entire battery pack out of the vehicle and replacing it with a fully charged one. This method requires specialized equipment and a dedicated swapping station with hydraulic lifts to remove and replace the heavy battery pack. Vertical battery change does provide some difficulties, though. Installing the charging stations needs significant infrastructure investment, and standardizing battery sizes and interfaces across various EV models is also necessary. Furthermore, it can take a while to finish, which could be longer than the time required to recharge a battery using a rapid charging station.

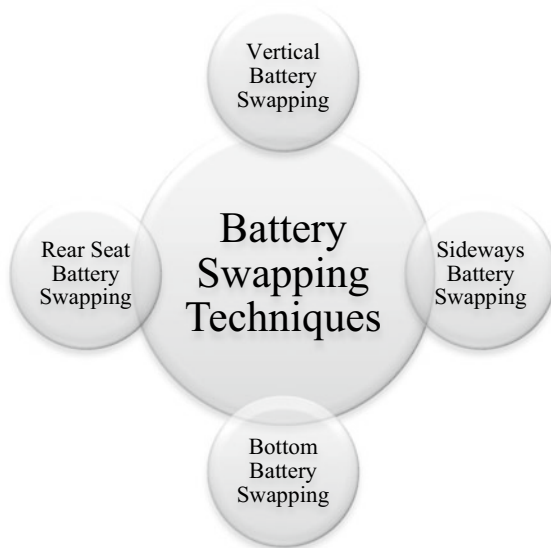
3 Functioning of Battery Swapping Stations

Implementing the technique of swapping vehicles, vans, and buses requires extensive planning, as shown in the block diagram in Fig. 2. The accessibility of batteries and chargers, cloud-based data storage and management, and interaction among components to ensure interoperability should all be considered during this planning. The BSS relies on constant communication between the smart vehicle, exchange station, and information system to function properly. The information system will facilitate the vehicle’s communication with the station [9].

WAVE communication will allow communication between the vehicle and the information system. In contrast, the station will communicate via the local Internet. The vehicle will notify the information system and request a battery switching service when the battery dies. The information system will then inform the station of the vehicle’s location, anticipated arrival time, and identifying information. This will allow the station to prepare the battery in advance to be ready when the vehicle arrives (Fig. 3).

When the car approaches the station, the driver swipes their registration card, and the computer verifies the information. This information pertains to vehicles and their batteries, as well as past swaps and purchases. All data should be stored in the cloud and made available to the owner of the station and client for full operational transparency. After swapping, the old battery is inspected for signs of charge, degradation, age, and total number of charge/discharge cycles [9, 10]. The charging station’s key components are as follows:

Fig. 2 Battery swapping techniques



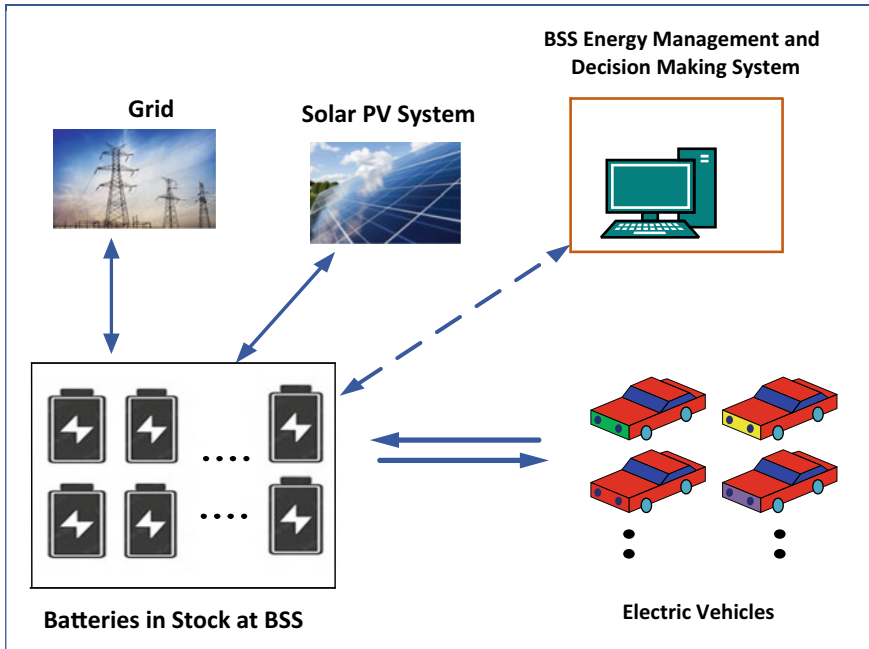


Fig. 3 Working of battery swapping station

- Control room (for managing and observing the BSS's overall operation).
- Battery racks and charging racks together.
- Track swapping (the area where the batteries are when the track is switched).
- The route where the automobile is when being switched out.
- changing the service room for the robot, battery, and charger.
- Service area for additional BSS parts.

As is customary at petrol stations, the first driver to the BSS should be given priority over those who arrive later. The batteries must be recharged right away after the exchange in order to serve the demands of all customers visiting the BSS. The range and the need to switch batteries while using a BSS that is built in should both be maximized [11].

A more adaptable and effective EV battery swap design is needed due to constraints, including location, the availability of BSSs, and station congestion. Passive mode and active mode are both examples of battery swapping. The goal of having enough power to drive an EV whenever and wherever is not met while the vehicle is dormant. To change the battery pack, drivers must travel to the BSS. There may still be a long wait for a battery at that time due to the station's remote location and low battery stock [10].

In the active mode, a second vehicle drives up to one whose battery is low or dying so that they can switch batteries. Recently, a unique system for rapidly replacing EV batteries was developed [12]. Be put in a mobile BSS, which is a modified vehicle. Battery replacement and removal are performed simultaneously. As a result, the exchange procedure moves quickly and only needs a short amount of time (approximately three minutes in the testing setting). The swap can take place whenever it happens to be comfortable for all involved by using a vehicle such as the BSS [13].

4 Problems Associated with BSS

The restricted driving range, high purchase price, scarce supply of charging stations, and lengthy charging times are some barriers to EV adoption. There are further psychological problems, such as driver range anxiety and resistance to new technologies. The majority of the hurdles are effectively shortened by battery swapping technology. But the situation is more complicated than this, and below are the difficulties with battery swapping technology (Fig. 4):

A Standardization

The market is dominated by technology because of its brand compatibility and cross-platform capabilities. A battery swapping system can remain the norm only if battery packs from different manufacturers are interchangeable. This straightforward approach needs approval from the manufacturer. Considering all possible technological possibilities, EV driving scenarios, potential failure causes, and BSS viability, standards should be formulated to verify the minimal time needed to change batteries throughout their service lives.

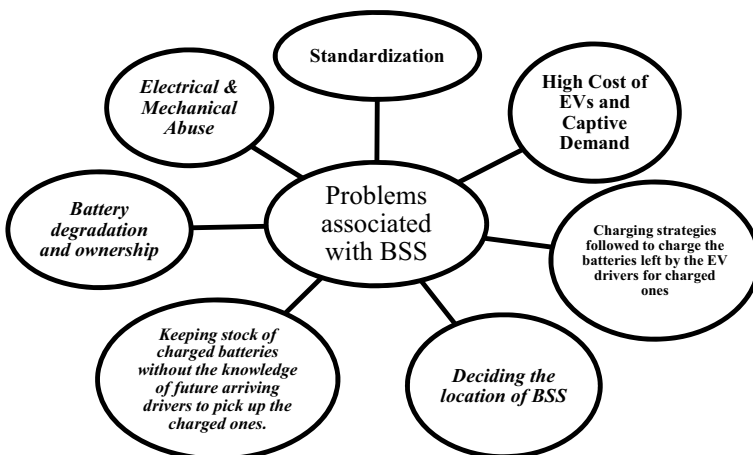


Fig. 4 Various problems associated with BSS

B High Cost of EVs and Captive Demand

Even though the total cost of ownership of an electric-to-wheeler is approaching that of an ICE equivalent, the EV scale-up is prevented by the EV's acquisition price, which is still 30–50% higher. In this cutthroat marketplace for 2-wheelers, batteries can be expensive to repair. Investments in real estate, equipment, batteries, and inventory are needed to move batteries. Exchange stations must be placed in busy places where customers can quickly access them, which drives the rent. Asset utilization is required for capital-intensive businesses.

C Charging strategies followed to charge the batteries left by the EV drivers for charged ones

In BSS, multiple charging schemes have been proposed and investigated; each one aims to maximize the efficiency of the charging process. The concept of battery-to-grid (B2G) and vehicle-to-grid (V2G) technology is the foundation for one approach to this problem.

EVs can act as mobile energy storage units in B2G and V2G systems, feeding electricity back into the grid during high demand. This idea can include BSS, where EV drivers can leave their batteries charged during off-peak hours and drained during peak hours to help balance the grid's energy supply and demand.

Battery exchange stations commonly use two charging methods:

1. Fast Charging

With high-power chargers, fast charging can fully replenish depleted batteries in a fraction of the time it would take using conventional methods. When there is a large demand for battery replacements and a short turnaround time is necessary, this method is often employed. While fast charging can replenish a depleted battery in as little as 5–10 min, it can shorten its useful life due to the high heat generated during the process [14, 15].

2. Slow Charging

The batteries are charged slowly over a long time using low-power chargers. This strategy is often implemented when the need for battery replacement is low and a longer turnaround time is tolerable. Although slow charging can take several hours or even a whole night, batteries are safer and do not affect their useful life [14, 15].

Depending on the needs of its clientele, some BSSs may also use a hybrid approach, including both rapid and gradual charging. Battery types (such as lithium-ion or nickel-metal hydride) and their respective manufacturers' guidelines can dictate how best to charge them. It's important to remember that BSS must meticulously manage its charging procedures to guarantee that batteries get fully charged and returned to service. Both overcharging and undercharging can shorten the battery's useful life, reduce its performance, and even pose a safety risk.

D *Deciding the location of BSS*

A BSS's accessibility and convenience to drivers of EVs are greatly influenced by its placement. Location is crucial to a BSS's business success. By carefully spacing BSS, range anxiety, a significant drawback of EVs, can be mitigated [16].

BSS location also affects system running expenses. Using an aggregator or selecting a location based on wind power are different business models. These models determine the appropriate position for all network BSSs. BSS location is crucial for optimizing charging and discharging operations to balance grid supply and demand [17, 18]. Putting the BSS in the right spot can ease concerns about the EV's range. That could encourage more people to drive EVs. A trade-off between EV path, grid energy losses, and BSS location arises from optimal EV planning considering swapping station positions [19]. The location of a swap station can be determined cheaply and effectively with a GPS position-based framework. The ideal location for the BSS can only be determined by compiling data from the traffic network and electricity requests. A site can be determined by looking at the spatial distribution of the demand for swapping and the mean distance error (MDE) between the locations [20].

Day-ahead energy and reserve capacity markets are recommended for optimal BSS operation. Location determines BSS price, which schedules battery charging and discharging. The BSS then submits synchronized day-ahead energy and reserve capacity bids. Aftermarket clearing, the BSS may mimic the process to calculate the locational marginal prices (LMP) and reserve capacity prices. Hence, the BSS can improve grid stability and profitability while operating more effectively [21]. As the power system moves toward a smart grid, BSS will help maintain grid stability and dependability. A smart grid's optimum BSS locations have been suggested. BSS can be deployed where they will have the most significant impact with the fewest costs, improving the smart grid's dependability and efficacy [22].

Vehicle and electrical network constraints are the two primary categories of limitations impacting the ideal BSS location. Vehicle network restrictions include travel distance, budget, BSS, traffic flow, and EV driving range. Power flow, charging demand, voltage, and thermal limits are electrical network limits. These restrictions are critical for BSS and microgrid efficiency, as they determine the best BSS placement.

E *Keeping stock of charged batteries without the knowledge of future arriving drivers to pick up the charged ones*

Concerns may arise from keeping a supply of charged batteries on hand without knowing when drivers will arrive to pick them up. The most fundamental problem is that batteries can lose their charge over time, especially if they aren't used or replaced frequently. Since the batteries may need to be recharged or replaced more often, this might be a waste of time and money. The research advises working with drivers to build a system for scheduling battery pickups and ensuring the batteries are adequately used. Using battery management software to track power consumption

and anticipate when batteries will require charging is one option, as is establishing lines of communication with drivers to schedule pickups. A study highlights the importance of excellent battery management systems and scheduling tools to reduce the batteries needed to fulfill demand in EV fleets. Findings from the study suggest that by taking a proactive and coordinated approach to battery management, EV fleets can improve their sustainability and cost-effectiveness by using their battery resources and reducing waste. Therefore, these studies stress the need for proper battery management systems and scheduling tools to be implemented in EV fleets to maximize battery utilization. Coordinating with drivers and using predictive analytics to manage battery inventories may reduce waste, improve sustainability, and cut costs for EV fleets [23, 24].

F Battery degradation and ownership

Performance degradation reduces battery charge range. Therefore, customers would prefer the new battery packs over older ones because they will provide poor energy storage due to deterioration, affecting EV mileage. New battery packs with much reduced operational cycles will satisfy customers. A car owner will never have any ownership rights to a battery pack. This has several benefits. Because the owner doesn't own EVs, the overall cost is reduced. Instead, he will pay a lease fee and never get the battery back. Each transaction or monthly can be charged for this. The latter leasing method is expensive because it requires two battery packs and the changing station service price.

G Electrical and Mechanical Abuse

BSS must clean batteries to avoid high-resistance connections from shorting. Electrical connectors and cables must last thousands of swaps. The BSS's main electrical components include a distribution transformer, AC/DC chargers, battery packs, and a battery energy control module. BSS requires high power (33–11 kV) and varying charging levels for electrical safety.

The BSS mechanical framework comprises vehicle platforms, battery lifts, vehicle alignment equipment rollers, battery conveyor shuttles, and battery storage rails and racks. Swapping EV battery packs should not cause mechanical damage. It should survive repeated EV-BSS switching.

5 Impact of BSS on Distribution Network

For extensive EV-grid integration, the power interaction between the distribution grid and EVs has emerged as a critical issue [25]. The BSS views the battery as a charging element instead of the conventional EV charging station. Its arrival suggests a new way for EVs and the grid to interact regarding energy. The efficiency of the EV-grid power interaction is increased by the BSS operating mode, which enables an EV to quickly swap out an empty battery for a full one while leaving the depleted battery

in the BSS for charging. In contrast, BSS may make reasonable plans for battery charging and discharge depending on power costs and the condition of the distribution system, which enhances the efficiency and dependability of distribution network operations [26, 27]. Studying the impact of BSS connections on the distribution network is therefore crucial.

BSS positively and negatively may significantly impact the distribution network. Here are a few possible effects

A Increased demand

BSS can increase the demand for electricity in a particular area, especially during peak hours when many drivers may need to swap their batteries. This increased demand could strain the distribution network and require upgrades to the local grid infrastructure to ensure it can handle the load. This can overload distribution transformers leading to more power outages, equipment damage, and increased maintenance costs [28].

B Cost implications

Depending on the location and scale of the BSS, significant costs may be associated with upgrading the distribution network to support it. Additionally, ongoing maintenance and operational costs may need to be considered [29].

C Increased reliability

BSS could help improve the reliability of the distribution network by being utilized as an alternate power source to inject electrical energy into the grid to lessen load shedding when the grid fails [30].

D Potential for renewable integration

BSS could provide a valuable tool for integrating renewable energy sources into the grid, which could help reduce its potential negative impacts on the distribution network. Some ways in which renewable energy can be integrated with BSS to minimize their effects are by installing solar panels at or near the BSS, which can provide a reliable source of renewable energy that can be used to charge the batteries, and by using wind turbines in areas with good wind resources. For example, batteries charged by solar panels during the day could be swapped out and used to power EVs at night when demand is high [10, 31].

The adverse effects of BSS on the distribution grid can be mitigated by integrating renewable energy sources, as mentioned above, by using optimal charge scheduling and giving different priorities to vehicles while charging.

The practice of planning the charging of EVs to maximize efficiency and reduce the cost of charging is known as optimal charge scheduling. This entails figuring out when and how long each car should be charged, considering things like the availability of renewable energy sources, the price of electricity, and the grid's condition [32]. In [33] proposes an optimal charging schedule for EVs in solar-powered charging stations based on day-ahead predictions of solar power generation. Optimal charging

scheduling can also effectively reduce the potential negative impacts of BSS on the distribution network. By implementing time-of-use pricing, smart charging systems, battery management, V2G technology, and demand response programs, we can help reduce the peak demand on the distribution network and ensure a more reliable and sustainable energy system [34, 35].

Another tactic to lessen the possible harm caused by BSS on the distribution network is prioritizing the charging of batteries. This can be implemented by using the following steps [36–39].

E Charging light vehicle batteries first

During peak hours, the demand on the distribution network can be decreased by charging the batteries of light vehicles first. The overall stress on the distribution network can be minimized by prioritizing these vehicles first, which consume less electricity than heavy-duty vehicles.

F Charging batteries based on battery state-of-charge

Low state-of-charge (SOC) batteries can receive priority during charging. This can ensure that the weakest-powered cars have the fuel they require to continue their journey. By reducing the time vehicles spend at the exchange station, the demand on the distribution network may be lessened.

G Charging batteries based on the type of EV

Different types of EVs require different amounts of energy to charge. The overall demand on the distribution network can be reduced by prioritizing charging EVs with lower energy requirements. For example, a smaller EV may require less energy to charge than a more extensive, heavy-duty EV.

H Charging batteries based on the time of day

Charging batteries during off-peak hours can help reduce the peak demand on the distribution network. By prioritizing charging during times when renewable energy sources are most abundant, the use of non-renewable energy sources can be reduced, minimizing the carbon footprint of charging EVs (Fig. 5).

6 Usage of BSS as a Battery to Load (B2L) During Peak Loads

Battery to load refers to the use of batteries to supply power directly to the load during peak hours, which helps reduce the stress on the power grid. The BSS can store excess energy during off-peak hours and supply it to the load during peak hours. This can help reduce the load on the power grid and prevent brownouts or blackouts. The usage of BSS as B2L during peak loads can offer several benefits to the grid. These benefits can be explained below as follows:

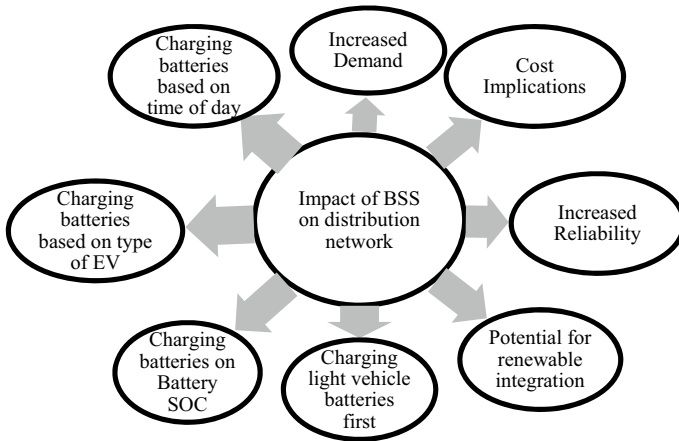


Fig. 5 Impact of BSS on distribution network

A Load management

By using BSS as a source of energy during peak loads, the load on the grid can be managed more effectively. This can help prevent overloading and improve the stability and reliability of the grid.

B Cost savings

Using BSS as an energy source during peak loads can help reduce the need for expensive infrastructure upgrades or the use of expensive peaking power plants. This can result in utility cost savings and potentially lower consumer electricity rates.

C Environmental benefits

If the BSS is powered by renewable energy sources, using them as a source of energy during peak loads can help reduce greenhouse gas emissions and promote a more sustainable energy system.

D Improved grid resiliency

Using BSS as a source of energy during peak loads can improve the grid’s resiliency by providing backup power in the event of a grid outage or other emergencies.

However, there are also some challenges to using BSS as a source of energy during peak loads. These include the need to ensure that the batteries are charged and available when needed, the potential for equipment damage or failure, and the need to manage the energy distribution from the BSS effectively. Overall, the usage of BSS as a battery to load during peak loads can offer several benefits to the grid. Still, it requires careful planning and management to ensure its effectiveness and reliability (Fig. 6).

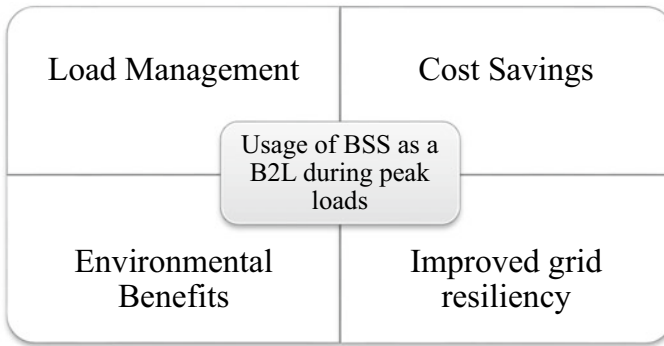


Fig. 6 Usage of BSS as battery to load

7 Conclusion

This paper discusses the concept of battery swapping stations (BSS) for electric vehicles (EVs). This concept is superior to the EV charging station when compared in many aspects, like the time the EV driver needs to spend at the EV charging station. In addition, various problems associated with the BSS are discussed, like standardization of EV batteries, the high cost and capital required for installation, and deciding the optimal location of BSS along with the capacity. Moreover, the impact of BSS on the distribution network is discussed; also, the usage of BSS in the battery to load (B2L) mode is discussed to meet the high load demands during peak hours.

References

1. Li W, Guo WH, Zhang J (2017) Development of electric vehicle battery swap stations and service network in China. *Transp Res Procedia* 25:4950–4957
2. Wu T, Wu Y, Zhang J (2019) Electric vehicle battery swapping station: a review of technologies and operations. *Energies* 12(20):3843
3. Tavares PCP, Catalão JPS, Rocha Almeida PM (2016) Battery swapping stations for electric vehicles: a review. *IEEE Trans Transp Electrification* 2(2):154–167
4. Sodha NNS, Das S (2020) Design and analysis of a battery swapping station for electric vehicles. *J Energy Storage* 29:101
5. Bhatia SPS, Agarwal S (2021) Feasibility analysis of battery swapping stations for electric vehicles in India. In: *IEEE transportation electrification conference and expo (ITEC)*. pp 1–6
6. Mahoor M, Hosseini ZS, Khodaei A (2017) Electric vehicle battery swapping station. In: *CIGRE grid of the future symposium, Paris, France*. pp 1–5
7. Khalid MR, Khan IA, Hameed S, Asghar MSJ, Ro J-S (2021) A comprehensive review on structural topologies, power levels, energy storage systems, and standards for electric vehicle charging stations and their impacts on grid. *IEEE Access* 9:128069–128094. <https://doi.org/10.1109/ACCESS.2021.3112189>

8. Liang Y, Zhang X (2018) Battery swap pricing and charging strategy for electric taxis in China. *Energy* 147:561–577
9. Adegbohun F, von Jouanne A, Lee K (2019) Autonomous battery swapping system and methodologies of electric vehicles. *Energies* 12(4):667. <https://doi.org/10.3390/en12040667>
10. Khalid MR, Alam MS, Krishnamurthy M, Al-Ammar EA, Alrajhi H, Asghar MSJ (2022) A multiphase AC–DC converter with improved power quality for EV charging station. *IEEE Trans Transp Electrific* 8(1):909–924. <https://doi.org/10.1109/TTE.2021.3120032>
11. Armstrong M, El Hajj Moussa C, Adnot J, Galli A, Riviere P (2013) Optimal recharging strategy for battery-switch stations for electric vehicles in France. *Energy Policy* 60:569–582. <https://doi.org/10.1016/j.enpol.2013.05.089>
12. Shao S, Guo S, Qiu X (2017) A mobile battery swapping service for electric vehicles based on a battery swapping van. *Energies* 10(10):1667. <https://doi.org/10.3390/en10101667>
13. Khalid MR, Asghar MSJ (2017) A new topology for single stage thyristor based grid connected single phase inverter for renewable energy systems. In: *IEEE international conference on power, control, signals and instrumentation engineering (ICPCSI)*, Chennai, India. pp 724–731. <https://doi.org/10.1109/ICPCSI.2017.8391809>
14. Chiang YM, Chiang YK (2019) Battery swapping stations: opportunities and challenges for electric vehicles. *Energies* 12(22):4306
15. Liu J, Wang Z (2019) Battery swapping for electric vehicles: a critical review and future research. *Energy Sci Eng* 7(5):1864–1884
16. Rahman N, Aiman U, Alam MS, Khalid MR, Sarwar A, Asghar MSJ (2020) A non-isolated DC-DC boost converter with high gain ability for renewable energy sources applications. In: *International conference on decision aid sciences and application (DASA)*, Sakheer, Bahrain. pp 137–141. <https://doi.org/10.1109/DASA51403.2020.9317225>
17. You P et al (2018) Scheduling of EV battery swapping—part I: centralized solution. *IEEE Trans Control Netw Syst* 5(4):1887–1897. <https://doi.org/10.1109/TCNS.2017.2773025>
18. You P et al (2018) Scheduling of EV battery swapping—part II: distributed solutions. *IEEE Trans Control Netw Syst* 5(4):1920–1930. <https://doi.org/10.1109/TCNS.2017.2774012>
19. Wang S, Yu L, Wu L, Dong Y, Wang H (2019) An improved differential evolution algorithm for optimal location of battery swapping stations considering multitype electric vehicle scale evolution. *IEEE Access* 7:73020–73035. <https://doi.org/10.1109/ACCESS.2019.2919507>
20. Zeng M, Pan Y, Zhang D, Lu Z, Li Y (2019) Data-driven location selection for battery swapping stations. *IEEE Access* 7:133760–133771. <https://doi.org/10.1109/ACCESS.2019.2941901>
21. Sepetanc K, Pandzic H (2020) A cluster-based operation model of aggregated battery swapping stations. *IEEE Trans Power Syst* 35(1):249–260. <https://doi.org/10.1109/TPWRS.2019.2934017>
22. Rezaee Jordehi A, Javadi MS, Catalão JPS (2021) Optimal placement of battery swap stations in microgrids with micro pumped hydro storage systems, photovoltaic, wind and geothermal distributed generators. *Int J Electr Power Energy Syst* 125:106483. <https://doi.org/10.1016/j.ijepes.2020.106483>
23. Salinas-Solano O, Yilmaz M, Eksioğlu S (2020) Battery swapping stations as an example of a framework for managing the supply chain for batteries for electric vehicles. *J Energy Storage* 32:101606
24. Khalid MR, Alam MS, Asghar MSJ (2020) A state-of-the-art review on xEVs and charging infrastructure. In: *International conference on decision aid sciences and application (DASA)*, Sakheer, Bahrain. pp 335–342. <https://doi.org/10.1109/DASA51403.2020.9317029>
25. Zeng B, Feng J, Liu N, Liu Y (2021) Co-optimized parking lot placement and incentive design for promoting PEV integration considering decision-dependent uncertainties. *IEEE Trans Industr Inf* 17(3):1863–1872. <https://doi.org/10.1109/TII.2020.2993815>
26. Zeng B, Luo Y, Liu Y (2022) Quantifying the contribution of EV battery swapping stations to the economic and reliability performance of future distribution system. *Int J Electr Power Energy Syst* 136:107675. <https://doi.org/10.1016/J.IJEPES.2021.107675>

27. Pankaj S, Khalid MR, Saad Alam M, Jamil Asghar MS, Hameed S (2022) Electric vehicle charging stations and their impact on the power quality of utility grid. In: International conference on decision aid sciences and applications (DASA), Chiangrai, Thailand. pp 816–821. <https://doi.org/10.1109/DASA54658.2022.9765054>
28. Ahmad B, Amrr SM, Nabi M, Khalid MR, Jamil Asghar MS (2021) Analysis of three-phase grid-tied thyristor based inverter for solar PV applications. In: International conference on sustainable energy and future electric transportation (SEFET), Hyderabad, India. pp 1–5. <https://doi.org/10.1109/SeFet48154.2021.9375746>
29. Liang Y, Zhang X, Xie J, Liu W (2017) An optimal operation model and ordered charging/discharging strategy for battery swapping stations. *Sustainability* 9(5):700. <https://doi.org/10.3390/su9050700>
30. Lebrouhi B, Khattari Y, Lamrani B, Maaroufi M, Zeraoui Y, Kousksou T (2021) Key challenges for a large-scale development of battery electric vehicles: a comprehensive review. *J Energy Storage* 44:103273. <https://doi.org/10.1016/j.est.2021.103273>
31. Khalid MR, Khan IA, Siddiqui NI, Hameed S, Asghar MJ (2021) Performance evaluation of multilevel DC–AC converter to interface EV battery for V2H application. In: North American power symposium (NAPS), College Station, TX, USA. pp 1–6. <https://doi.org/10.1109/NAP552732.2021.9654685>
32. Tan X, Qu G, Sun B, Li N, Tsang DHK (2019) Optimal scheduling of battery charging station serving electric vehicles based on battery swapping. *IEEE Trans Smart Grid* 10(2):1372–1384. <https://doi.org/10.1109/TSG.2017.2764484>
33. Mohammed S, Titus F, Thanikanti SB, Deb S, Kumar NM (2022) Charge scheduling optimization of plug-in electric vehicle in a PV powered grid-connected charging station based on day-ahead solar energy forecasting in Australia. *Sustainability* 14(6):3498. <https://doi.org/10.3390/su14063498>
34. Dai Q, Cai T, Duan S, Zhang W, Zhao J (2014) A smart energy management system for electric city bus battery swap station. In: IEEE conference and expo transportation electrification Asia-Pacific (ITEC Asia-Pacific), Beijing, China. pp 1–4. <https://doi.org/10.1109/ITEC-AP.2014.6941107>
35. Ban M, Zhang Z, Li C, Li Z, Liu Y (2021) Optimal scheduling for electric vehicle battery swapping-charging system based on nanogrids. *Int J Electr Power Energy Syst* 130:106967. <https://doi.org/10.1016/j.ijepes.2021.106967>
36. Bari N, Haque A, Ahuja G, Kurukuru VSB (2021) Priority based power delivery system for electric vehicle charging. In: Musleh Al-Sartawi AM, Razzaque A, Kamal MM (eds) *Artificial intelligence systems and the Internet of Things in the digital era. EAMMIS 2021. Lecture notes in networks and systems*, vol 239. Springer, Cham
37. Khalid MR, Alam MS, Amrr SM, Jamil Asghar MS (2021) Multi-pulse converter based rectification scheme for improving power-quality of EVs charging station. In: International conference on sustainable energy and future electric transportation (SEFET), Hyderabad, India. pp 1–5. <https://doi.org/10.1109/SeFet48154.2021.9375681>
38. Jawale SA, Singh SK, Singh P (2021) Priority based electric vehicle dynamic charging station. In: Third international conference on inventive research in computing applications (ICIRCA), Coimbatore, India. pp 192–196. <https://doi.org/10.1109/ICIRCA51532.2021.9544651>
39. Khalid MR, Asghar MSJ, Alam MS, Hameed S, Khan IA (2022) Experimental validation of off-board EV charging station with reduced active switch count. *Int J Energy Res* 46(12):16929–16948. <https://doi.org/10.1002/er.8359>

Comparison Between PID and SMC Controller to Control the Speed of DC Separately Excited Motor



Prakansha Sulakhe, V. P. Rajderkar, and Tushar Guhe

Abstract The purpose of this study is to use an SMC controller to build a speed control scheme for an independently stimulated DC motor. The speed control system is created first, after which it is applied to the model in MATLAB SIMULINK. Following that, we used SMC to gauge the DC motor's speed. The results of the simulation show that a sliding mode controller is the best option for managing the speed of a DC motor as compared when with PID. The SMC's immunity to interruptions allows it to perfectly match the intended pace.

Keywords SMC controller · PI controller · DC separately excited motor · Speed control

1 Introduction

There is a lot of work being put into the control of discontinuous control actions because managing nonlinear systems has always been a significant problem for the study of control systems and automated control theory. In the industrial world, there are quite a few applications for direct current (DC) motors. Changes in armature current are required to alter the variable resistance in the armature circuit or field circuit. A typical speed control system of the motor characteristics is used to regulate the DC motor mentioned above. In order to keep the system stable and create the required system responses, a suitable control should be designed when there are systemic upheavals and uncertainty. So-called “matched uncertainties,” as well as other outside uncertainties and disturbances, have no effect on sliding mode control

P. Sulakhe (✉) · V. P. Rajderkar
Electrical Engineering, G. H. Raisoni College of Engineering, Nagpur, India
e-mail: prakanshasulakhe97@gmail.com

V. P. Rajderkar
e-mail: vedashree.rajderkar@raisoni.net

T. Guhe
Janki Electromech, Nagpur, India

(SMC). Real-time processing capability has improved thanks to quick processor advancements, making it possible to digitally apply cutting-edge control methods that were previously only feasible in principle. Other difficulties include modifying traditional controllers, which in SMC design is not at all difficult. The distinguishing characteristics of SMC enable its practical application in the management of electrical drives. Strong resilience, ease of implementation, order reduction, and disturbance rejection are some of these characteristics. Following is a summary of the main benefits of the SLMC technique: dynamic reaction time simple to implement and simple-to-design resilience, low sensitivity to changes in system characteristics and load obstacles. To ensure the requisite performance, a precise understanding of the motor characteristics is important. While the drive is running, a number of parameters are unknown with certainty and/or are liable to change, which lowers performance. The SMC is an obvious alternative to avoid this design issue. In actuality, it is created from the ground up to regulate system uncertainties, and it may even deliver appropriate performance in the presence of considerable and quick changes in the motor characteristics and a wide range of disturbances.

Since a long time ago, PI controllers are typically utilized for speed control. However, PI controllers cannot completely stabilize the speed when there are problems with modelling uncertainties, parameter fluctuations, load disturbances, or severe nonlinearity. In these circumstances, where speed control is a major issue, SMC can be employed successfully. SMC is utilized in the suggested strategy to develop the controller and ensure reliable functioning.

2 Model of DC Separately Excited DC Motor

Using amplifiers or power modules, commonly known as DC motor drives, a controller and a DC motor are coupled. The controller's step and direction inputs are converted by them into currents and voltages that work with motors. Robotic and electrical machinery typically employ DC motors. In the light of how important it is to control the DC motor's speed.

2.1 Armature Control

The rear emf or E_b has a direct impact on the speed of the DC motor. This demonstrates that speed and armature current I_a are directly related when the supply voltage V and armature resistance R_a are both constant. Figure 1 depicts the DC Motor Correspondence Circuit. Dc motors are frequently employed in the magnetization curve's linear region. One method is to use armature resistance, which involves adding a variable resistance to the armature's circuit. The voltage drop across the armature is lower than the voltage across the line after it has been raised because of the circuit's increased resistance. The armature circuit consists of an inductor (L_a), a resistor (R_a),

Fig. 1 DC motor circuit diagram

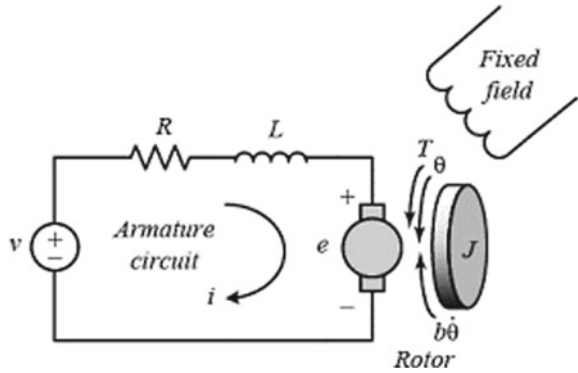


Table 1 Specification of DC motor

Parameters	Values and unit
Resistance (R_a)	00.6 Ω
Inductance (L_a)	00.012 H
Moment of inertia of rotor (J)	1.67 kg m ² /s ²
Electromotive force const. (K_t)	3.0 Nm/A
Back EMF const. (K_b)	3.0 Vs/rad
b	1.67

and a voltage source (e_b). Table 1 displays the nominal rating of the DC motor drive utilized in this.

Consequently, the field current is proportional to the air gap flux.

$$\Phi = f(x) = K_f i_f \tag{1}$$

where K_f is a const. or fixed value.

As a result of the armature current and air gap flux, the motor's torque T_m is proportional to that product.

$$T_m = K_f K_1 i_f i_a \tag{2}$$

where K_1 is a const. or fixed value.

The field current in the armature-controlled DC motor is maintained constant,

$$T_m = K_T i_a \tag{3}$$

K_T , the motor torque const., is used here.

Given as follows, the relationship between the motor back emf and speed is

$$e_b = K_b \frac{d\theta}{dt} \quad (4)$$

K_b is used here. It follows that the armature circuit's differential equation is

$$L_a \frac{di_a}{dt} + R_a i_a + e_b = e_a \quad (5)$$

Equation for torque is

$$J \frac{d^2\theta}{dt^2} + f_0 \frac{d\theta}{dt} = T_m = K_T i_a \quad (6)$$

If we apply the Laplace transforms to Eqs. (3) through (5) and assume that the initial conditions are zero, we obtain

$$E_b(s) = E_b s \theta(s) \quad (7)$$

$$(L_a s + R_a) I_a(s) = E_a(s) - E_b(s) \quad (8)$$

$$(J s^2 + f_0 s) \theta(s) = T_M(s) = K_T I_a(s) \quad (9)$$

From Eqs. (6)–(8) we obtained as following transfer function

$$G(s) = \frac{\theta(s)}{E_a(s)} = \frac{K_T}{s[(R_a + sL_a)(J_s + f_0) + K_T K_b]} \quad (10)$$

2.2 Sliding Mode Control

To adjust a DC motor's speed via sliding mode control, the motor must be constructed in a way that allows us to input the signal deriving from the sliding mode control that was discussed before. Nearly perfect noise rejection and set point tracking are needed for many real-world issues. To accomplish these performances, such systems can use SMC. As a result of the swift switching between a pair or more control limits, this control is referred to as nonlinear. The system's structure can change or switch when the system's state moves over each discontinuity when this control is used as feedback. The state intersects and intersects the surface—also known as the button or sliding surface—and exists continuously on the button such that the error and the rate at which the error changes are both zero. Moving in a sliding motion is the term for this action. The term “sliding mode control” is frequently used to emphasize how crucial sliding motion is (Fig. 2).

$$\dot{x}_1 = x_2 \quad A = \frac{1}{5}, x_1, x_2 < 0$$

$$X_2 = -Ux_1 = 5, x_1, x_2 > 0t$$

where ω is the angular velocity expressed in rad/sec. B is a viscous friction coefficient that resists the direction of motion in Nms. and J is the moment of inertia kgm^2/s^2 . The armature control's torque, expressed in Nm, is provided.

$$\tau(t) = J \frac{dw}{dt} + B\omega(t) \tag{11}$$

$$\tau(t) = K_t i_a(t) \tag{12}$$

$$V_a(t) - E_b(t) = R_a i_a(t) + L_a \frac{di_a}{dt} \tag{13}$$

where E_b represents electromagnetic force in V and R_a and L_a , respectively, show the resistance and inductance of the armature in ohms and H , respectively.

$$E_b(t) = K_b(t)\omega(t) \tag{14}$$

One may create a state model using the preceding equation and the variables and as illustrated below, i_a serve as state variables while V_a serves as a manipulating variable.

$$\begin{bmatrix} \frac{dw(t)}{dt} \\ \frac{di_a(t)}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{b}{J} & \frac{K_t}{J} \\ -\frac{K_b}{L_a} & -\frac{R_a}{L_a} \end{bmatrix} \begin{bmatrix} \omega(t) \\ i_a(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{L_a} \end{bmatrix} V_a(t) \tag{15}$$

$$\frac{\omega(s)}{V_a(s)} = \frac{3992.015}{s^2 + 51s + 51.39} \tag{16}$$

The aforementioned equation solved in time domain as

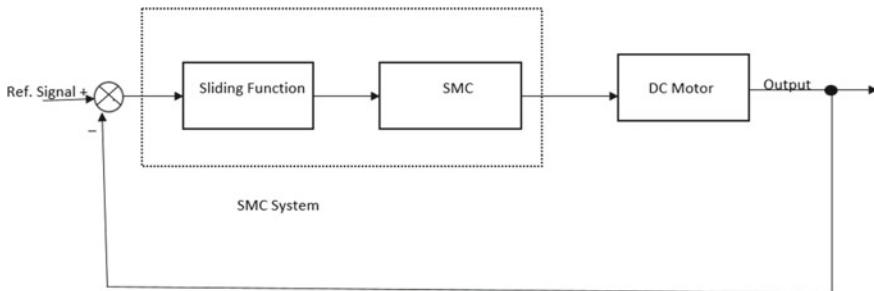


Fig. 2 DC motor schematic with SMC

$$(t) + 51.39\dot{\omega}(t) = 3992.015V_a(t) \quad (17)$$

Now consider, $x_1 = \omega(t)$ and $u = V_a(t)$ once transformed, the system can then take the canonical form shown below.

$$x_1 = x_2 \quad (18)$$

$$x_2 = -51.39x_1 - 51x_2 + 3992.015U \quad (19)$$

$$y = x_1 \quad (20)$$

Next, choose the sliding surface.

$$\sigma = c(r - x_1) + x_2 \quad (21)$$

c being the sliding matrix's constant $C\sigma\epsilon R^{m+n}$ such that $c < 0$ total control legislation is stipulated by

$$U = U_1 + U_{nl} \quad (22)$$

where $K > 0$ is picked where it is large enough the trajectory converges to the sliding surface more quickly, the higher the value of K .

3 PID Controller

PID control is a tried-and-true technique for moving a system towards a desired location or level. It is used in numerous chemical and scientific processes, as well as automation, and is essentially omnipresent as a technique of managing temperature. To keep a process's real output as close as feasible to the target or set point output, PID control employs closed-loop control feedback.

4 Result

The control system is designed by using MATLAB/Simulink. The maximum voltage is 100 V. Both the driving parameters and the load is changed. Figure 3 shows the DC motor drive with load with SMC. We must design the motor based on the designed parameters in order to regulate the DC motor speed using a SMC controller.

Figure 4 says that changing PID constants as the DC motor parameters are changed. It is seen that the output speed of DC motor increases due to SMC from 30 to 32 rpm and tend to adapt as the DC motor parameters are changed.

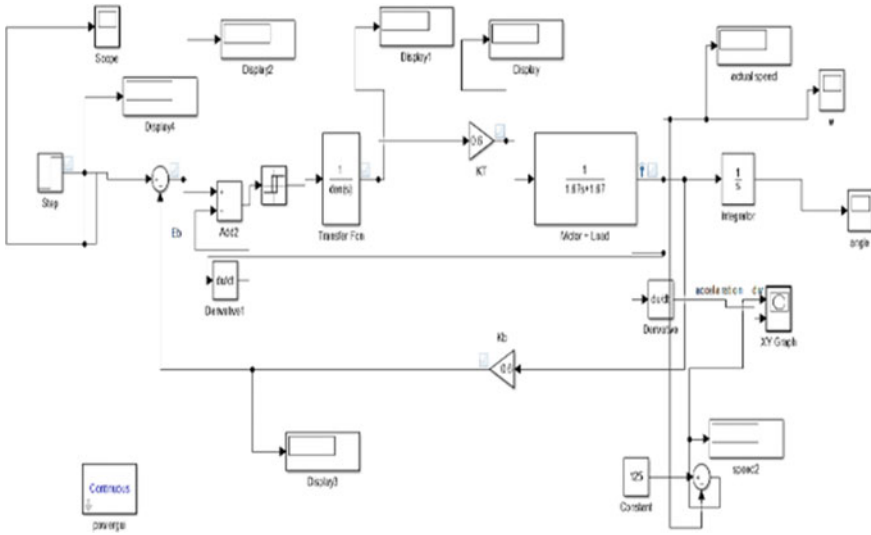


Fig. 3 Simulation model of DC motor with load with SMC

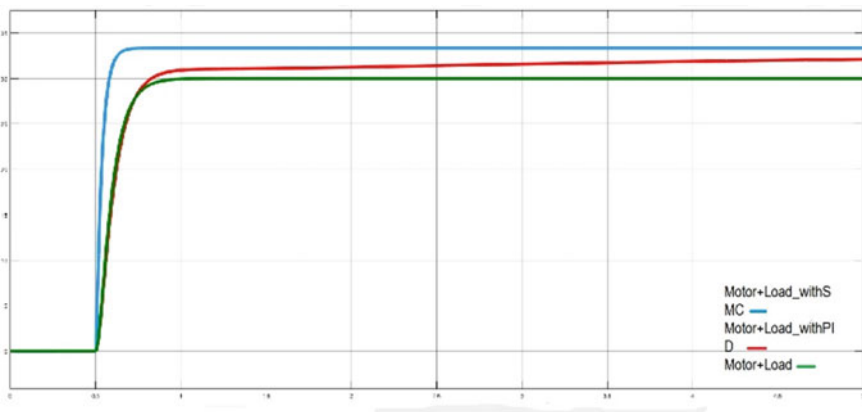


Fig. 4 Comparison of DC motor speed a without controller b with PID c SMC controller

Figure 5 shows the graph between time and speed where SMC gives the better result than PID.

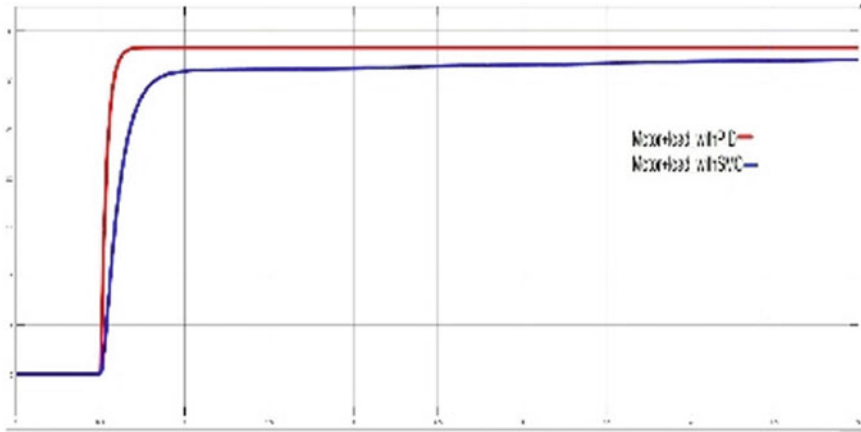


Fig. 5 Comparison of DC separately excited motor **a** with PID **b** with SMC

5 Conclusion

Due to their benefits of easy speed and position control and broad adjustability range, DC motors are commonly utilized as actuation devices in industrial applications. There has been provided a mathematical model of a DC motor. The speed of a DC motor has been controlled using a conventional SMC controller. The performance curves produced with and without controllers, as well as those from the conventional SMC controller, are compared using time domain requirements. The SMC controller aids in increasing the motor's speed. Comparison of PID with load and with load with SMC it is seen that SMC gives the better result than PID. As the DC motor's parameters are modified and a control signal is applied to it, the controller constants in the simulation's output can automatically update.

References

1. Maheswararao CU, Babu YSK, Amaresh K (2011) Sliding mode of a DC motor. In: International conference on communication systems and network technologies. IEEE. 978-0-7695-4437-3/11 \$26.00 © <https://doi.org/10.1109/CSNT.2011.86>
2. Ambesange SV, Kamble SY, More DS (2013) Application of sliding mode control for the speed control of DC motor drives. In: IEEE international conference on control applications (CCA) part of 2013 IEEE multi-conference on systems and control Hyderabad, India. IEEE, 28-30 Aug 2013. 978-1-4799-1559-0/13/\$31.00 ©2013
3. Rauf A, Yang J, Madonski R, Li S, Wang Z (2019) Sliding mode control of converter-fed DC motor with mismatched load torque compensation. School of Automation, Southeast University, Key Laboratory of Measurement and Control of CSE, Ministry of Education, Nanjing 210096, Jiangsu, China. lsh@seu.edu.cn 978-1-7281-3666-0/19/\$31.00 ©2019 IEEE

4. Chaal H, Jovanovic M (2010) Second order sliding mode control of a DC drive with uncertain parameters and load conditions. Northumbria University. 978-1-4244-5182-1/10/\$26.00 c IEEE
5. Torelli F, Montegiglio P, Piccinni G (2020) SMC-inspired control approach applied to DC-motor drives. 978-1-7281-7455-6/20/\$31.00 ©2020 IEEE
6. Kali Y, Saad M, Benjelloun K, Benbrahim M (2015) Sliding mode with time delay control for mimo nonlinear systems with unknown dynamics. In: International workshop on recent advances in sliding modes (RASM). IEEE, pp 1–6
7. Edwards C, Spurgeon S (1998) Sliding mode control: theory and applications. CRC Press
8. Perruquetti W, Barbot J-P (2002) Sliding mode control in engineering. CRC Press
9. Edwards C, Shtessel YB (2016) Adaptive continuous higher order sliding mode control. *Automatica* 65:183–190
10. Bhawoorjar I, Jagtap P (2022) Grid-connected hybrid pv power system performance evaluation by employing a unified power flow controller. In: 2nd Asian conference on innovation in technology (ASIANCON). IEEE, pp 1–5
11. Bose S, Khubalkar S (2022) Power quality analysis of textile industry-findings and recommendations. In: 2nd Asian conference on innovation in technology. IEEE, pp 1–6
12. Adware R, Chandrakar V (2022) Power quality enhancement through reactive power compensation using hybrid STATCOM. In: ICPC2T 2022–2nd international conference on power, control and computing technologies, proceedings
13. Vaidya P, Chandrakar VK (2022) Optimum placement of static synchronous compensator. In: IEEE 57 bus system 1st international conference on sustainable technology for power and energy systems, STPES
14. Mahalaxme S, Khubalkar S, Bharadwaj S (2020) Low voltage distribution box monitoring-new way to monitor power in industry. In: 4th international conference on trends in electronics and informatics (ICOEI) (48184). IEEE, pp 214–216
15. Vaidya P, Chandrakar VK (2022) Congestion management of large power network with static synchronous compensator. In: 2nd international conference on intelligent technologies, CONIT
16. Shende D, Jagtap P, Hiware R (2021) Review of enhanced power quality using unified power flow control system in electrical network. *J Phys: Conf Ser* 2089(1):012034
17. Shende D, Jagtap P, Hiware R (2021) Enhanced power quality using unified power flow controller systems. *J Phys: Conf Ser* 2089(1)
18. Jagtap P, Chandrakar V (2021) Comparative study of UPFC controllers to improve transient and dynamic stability of power system. In: IEEE 4th international conference on computing, power and communication technologies. IEEE, pp 1–7
19. Vaibhav Kale S, Prashant RP, Khatri R (2013) Unified power flow controller for power quality improvement. *Int J Emerg Sci Eng* 1(10):1–4
20. Rajderkar VP, Chandrakar VK (2021) Allocation of unified power flow controller (UPFC) through sensitivity approach for enhancing the system performance. In: 6th international conference for convergence in technology. I2CT
21. Ravichandrudu K, Pramod Kumar PS, Sowjanya VE (2013) Mitigation of harmonics and power quality improvement for grid connected wind energy system using UPFC. *Int J Appl Innovation Eng Manage (IJAIEM)* 2(10):141–156
22. Gopinath B, Vinothini N, Kumar S (2014) Modeling of UPFC using model predictive control and bacterial foraging algorithm. *Int J Innovative Res Comput Commun Eng* 2(1):2724–2731
23. Rajderkar VP, Chandrakar VK (2022) Enhancement of power system security by fuzzy based unified power flow controller. In: 2nd international conference on intelligent technologies, CONIT

Implementation and Evaluation of a Secure Keyless Ignition and Integrated Control System for an Arduino-Based Instrumented Car



Anshu Prakash Murdan  and Nikhil Kissoon

Abstract This study presents the design, implementation, and evaluation of an advanced Secure Keyless Car Ignition System for an Arduino-based Instrumented Car. RFID was selected for the keyless ignition system due to its robustness and difficulty to duplicate, thus enhancing vehicle security. A comprehensive mobile application for enhanced in-vehicle control was also developed. The application integrates a user-friendly Dashboard Interface, GPS services, Real-Time Passenger Detection, and Bluetooth-Controlled Window Operations. An exhaustive evaluation of the system demonstrates its reliability and robust performance under various conditions, meeting the set objectives. While the system represents a significant stride toward user-centric and integrated automotive controls, the study also uncovers areas for future enhancement, such as optimizing network-dependent features and bolstering security against advanced threats. This research not only validates the effectiveness of integrated, keyless vehicle control but also sets the stage for future innovations in this domain.

Keywords Secure keyless car ignition system · Arduino-based instrumented car · Integrated mobile application · Real-time passenger detection · Bluetooth-controlled window operations

1 Introduction

Automotive control systems have been at the forefront of technological advancements, driving the progression of vehicular technology from rudimentary mechanical systems to intricate computerized networks. The evolution of automotive control systems, encompassing Electronic Control Units (ECUs), telematics, and infotainment platforms, has significantly augmented the driving experience. These advanced

A. P. Murdan (✉) · N. Kissoon

Department of Electrical and Electronic Engineering, University of Mauritius, Reduit, Mauritius
e-mail: a.murdan@uom.ac.mu

systems amalgamate disparate functionalities—from engine and climate management to navigation and entertainment—into an intuitive, seamless package that bolsters control, safety, and convenience for drivers.

However, even with their advanced capabilities, these contemporary automotive control systems are not exempt from certain constraints. Conventional key-based ignition systems, for instance, present security issues due to their susceptibility to theft or misplacement. Furthermore, a lack of consolidated controls in existing systems can lead to a disjointed user experience. As the trend toward increasingly digitized and interconnected vehicles continues, the demand for more secure, integrated, and intuitive systems correspondingly intensifies.

This necessitates the development of a Secure Keyless Car Ignition System along with integrated controls. The introduction of a keyless ignition system, authenticated via secure digital mechanisms, signifies a substantial advancement in terms of security and user convenience. It mitigates the risks inherent to physical keys and refines the user experience by simplifying the vehicle's ignition process. The incorporation of integrated controls advances this concept even further, fostering an interconnected ecosystem within the vehicle. The end product is a driving experience that is more streamlined, efficient, and user oriented.

In light of the rapid technological advancements in the automotive industry, the development of a Secure Keyless Car Ignition System and integrated controls represents a logical next step. This research will delve into the design and implementation of such a system, focusing on its feasibility, effectiveness, and potential for enhancing the future of automotive technology.

2 Literature Review and Related Work

The continuous evolution of the automotive industry is well-documented through a wealth of existing research, which has progressively tackled the complexities of vehicle controls. Several studies have explored the possibilities and the nuances of secure keyless car ignition systems and mobile app controls for cars [1, 2], illustrating the technological trajectory that has led us to the current state of the industry.

Investigations into secure keyless car ignition systems have predominantly concentrated on the advancement of radio-frequency identification (RFID) and nearfield communication (NFC) technologies [3]. These studies have showcased the potential of keyless systems in improving security and user experience. However, they often underscore the vulnerability of such systems to relay attacks and the necessity for additional layers of security to mitigate such risks.

Likewise, research on mobile app controls for cars has spanned from conceptual frameworks to applied methodologies. The integration of diverse in-car systems into a single mobile application has been extensively studied, with the majority of research focusing on user interface design, real-time controls, and security implications [4, 5]. Nonetheless, there remains a noticeable gap in the literature concerning the holistic

integration of such systems, particularly with the incorporation of real-time passenger detection and automated window control.

Existing systems, despite their significant advances, demonstrate notable limitations. The current keyless systems often lack robust multi-layered security protocols and hence may still be vulnerable to advanced malicious attacks. Moreover, while mobile applications have emerged as practical interfaces for controlling various aspects of a car, there is a striking absence of comprehensive solutions that integrate an array of controls into a single, user-friendly platform.

The realm of instrumented cars, vehicles endowed with sensors and actuators designed to gather data and provide enhanced interaction with the driving environment, is rich in both applied research and commercially viable products. It is a rapidly evolving domain that blends the fields of automotive engineering, computer science, and human-computer interaction. Initial forays into the field focused on improving vehicular safety and driving comfort, culminating in features like Anti-lock Braking Systems (ABS), Electronic Stability Programs (ESP), and adaptive cruise control. Alongside, many studies focused on the integration of advanced sensor technologies within automobiles, laying the groundwork for the emergence of smart vehicles [6, 7].

The advent of the Internet of Things (IoT) brought a new dimension to instrumented cars. By connecting vehicles to broader digital ecosystems, numerous research works demonstrated opportunities for improved traffic management, efficient fuel utilization, and personalized in-car experiences. Indeed, the potential for vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications has expanded the horizons of instrumented cars beyond the vehicle itself [8, 9]. Recent years have seen an exponential growth in instrumented car technology, notably with the rise of autonomous driving [10]. However, despite this advancement, a number of challenges persist in the domain of instrumented cars. Most notable are the issues of security, privacy, and usability. The need for user-friendly interfaces that can efficiently control and manage the multitude of functions provided by the instrumented car is an area that has not seen sufficient attention.

This present study seeks to address these concerns and contribute to the existing body of knowledge by proposing a Secure Keyless Car Ignition System and integrated mobile application controls. By leveraging an RFID-based ignition system, the security of vehicle operation is significantly enhanced. The mobile application offers a unified interface for the control of various elements of the car, optimizing the user experience. The inclusion of innovative features like GPS integration, real-time passenger detection, and Bluetooth-controlled window operation extends the capabilities of existing instrumented cars, making them safer, more user-friendly, and better integrated with other digital services.

3 Theoretical Foundations and Conceptual Framework

At the heart of this investigation lies the Arduino platform—an open-source electronics platform centered on easy-to-use hardware and software. Arduino’s widespread adoption and straightforward programming language have made it a versatile tool for a wide range of applications, spanning from home automation to robotics, and even extending to automotive control systems [11]. Previous research has demonstrated Arduino’s suitability in managing complex tasks such as engine management, climate control, and even semiautonomous driving [12]. The inherent flexibility and accessibility of the Arduino platform make it an excellent choice for the implementation of our Secure Keyless Car Ignition System and integrated controls.

The study further builds upon several theoretical foundations encompassing secure keyless ignition, GPS integration, real-time passenger detection, and Bluetooth window controls. The principles of secure keyless ignition involve the use of encrypted digital signals for vehicle ignition, ensuring only authorized users can start the vehicle. GPS integration serves as a navigation aid and allows vehicle tracking, enhancing the security and functionality of the system [13]. Real-time passenger detection uses seat sensor data to monitor the presence and number of passengers in the vehicle [14], contributing to safety and customizability of the in-vehicle experience. Bluetooth window controls allow for wireless command of window positions [15], enhancing the convenience of vehicle operation.

These components, when isolated, each provide specific benefits. However, their true potential is realized when they are combined into an integrated system. This is where our conceptual framework comes into play. The framework envisages a system where these disparate technologies are harmonized under a single, user-friendly mobile application. The secure keyless ignition serves as the central pillar, around which the GPS, passenger detection system, and Bluetooth window controls revolve.

Conceptually, this implies the formation of a unified control environment where a user, through a single mobile interface, can securely ignite the vehicle, navigate roads, monitor passenger presence, and control window positions. By intertwining these technologies into one integrated system, we aim to create a seamless and enriched driving experience, one that elevates user convenience and security to new heights.

3.1 System Architecture and Design

The system architecture for the Secure Keyless Car Ignition System and integrated controls comprises multiple interconnected components, each serving a distinct function within the larger framework. At the heart of the architecture is the Arduino microcontroller, acting as the central processing unit that controls and manages the various subsystems, as depicted in Fig. 1.

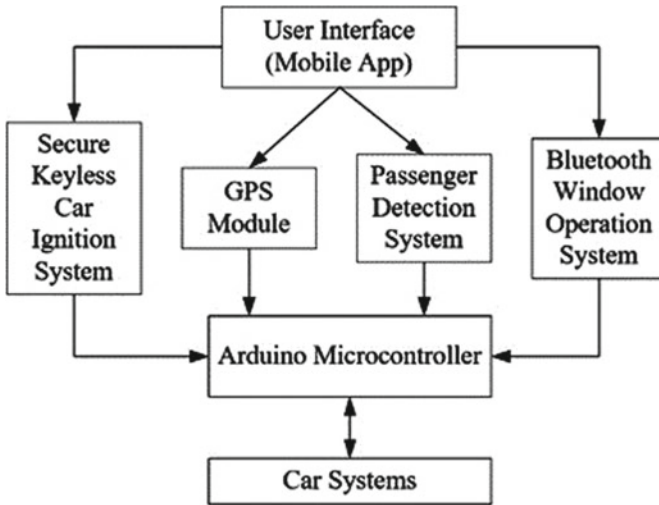


Fig. 1 System architecture

The design choices were made primarily with three objectives in mind: security, convenience, and integration. RFID was selected for the keyless ignition system due to its robustness and difficulty to duplicate, thereby enhancing vehicle security [16]. The mobile application was designed as the main control interface to enhance convenience and provide an integrated control environment. GPS integration provides both convenience through navigation support and an extra layer of security through vehicle tracking. Real-time passenger detection and Bluetooth window controls add layers of safety and user-centric customization to the vehicle's environment, enhancing the overall driving experience.

4 Implementation

The following sections elaborate on the implementation phase. Every component of the system was designed based on precise requirements and objectives.

4.1 Secure Keyless Car Ignition System

The Secure Keyless Car Ignition System is at the heart of our proposed design. It uses RFID technology to offer a reliable, touch-free ignition method. This system primarily consists of an RFID reader, distinct RFID tags, and their related encryption processes.

The RFID reader, linked to the Arduino microcontroller, is the bridge between the car and the driver. Every driver gets a distinct RFID tag, acting as their identifier. When the tag is placed in close proximity to the reader, the latter scans it, processes the encrypted data, and validates its authenticity. Upon successful validation, the microcontroller gets the go-ahead to initiate the car's ignition.

Ensuring high security is fundamental in crafting the Secure Keyless Car Ignition System. The system incorporates cryptographic techniques to protect the data embedded in the RFID tag. A robust encryption algorithm is used to encode the unique identifier of each tag. This ensures that even if the tag's data is intercepted during communication, it would be incomprehensible without the decryption key. Moreover, each tag is uniquely paired with the car's RFID reader, thereby preventing unauthorized access even from duplicate or spoofed tags.

The development process entailed configuring the RFID reader and the Arduino to communicate securely, implementing the cryptographic algorithms, and programming the ignition initiation upon successful authentication. During the implementation, considerable emphasis was placed on ensuring the system's reliability and responsiveness, as any delay or failure in the ignition process directly impacts the user experience.

Essentially, the Secure Keyless Car Ignition System enhances the traditional ignition process by incorporating secure, contactless technology. Its implementation underscores the integral role of security in the broader system and exemplifies the seamless fusion of advanced technology with traditional automotive systems.

4.2 Mobile Application Control

The Mobile Application Control plays an instrumental role in the proposed system, acting as the unified control interface for the user. The application is developed for both iOS and Android platforms, incorporating multiple functionalities such as an integrated dashboard, GPS navigation, real-time passenger detection, and Bluetooth-controlled window operation.

The design of our mobile app is rooted in a user-focused approach, promoting ease of use and seamless navigation. Its primary interface showcases a dashboard that offers live updates on various car metrics, including speed, fuel status, and engine temperature, to name a few. This information is sourced directly from the car's diagnostic system, processed by the Arduino microcontroller, and transmitted to the app through Bluetooth. The app also incorporates a GPS feature, drawing on the car's in-built GPS unit. This delivers both navigational aid and tracking capabilities. Besides aiding in direction, this dual-use tool adds a security dimension, offering the ability to pinpoint the car's location when necessary. Another feature is the real-time passenger detection, utilizing pressure-sensitive sensors on the seats. This information is channeled through the Arduino to the app, which in turn visualizes the occupied seats, giving the driver a clear sense of passenger occupancy. The app's window control capability enables users to operate the car's windows remotely.

Commands are sent from the app to the Arduino using Bluetooth, and windows are then adjusted with the help of servo motors.

Building this mobile app demanded meticulous coordination between the hardware input (via Arduino) and the software mechanics (within the app itself). This involved crafting a Bluetooth communication structure, designing dashboard and seat occupancy visuals, integrating GPS features, and establishing the window operation mechanism.

4.3 Dashboard Interface

The Dashboard Interface is central to the mobile app, offering users live updates on their vehicle's condition. It was built based on principles of user-friendliness, visual attractiveness, and clear data presentation. While the digital dashboard emulates the conventional car dashboard, it has the added advantage of adaptability. It vividly presents several live vehicle metrics, including speed, engine RPM, remaining fuel, and coolant temperature, among others.

These parameters are captured from the vehicle's onboard diagnostics system, processed by the Arduino microcontroller, and transmitted to the mobile application via the Bluetooth connection.

The design of the dashboard prioritizes readability and ease of understanding. Key parameters like speed and fuel level are displayed prominently, using intuitive gauges and meters that dynamically change in response to real-time data. Smaller icons and indicators provide additional information such as headlights status, seat belt warnings, and door ajar warnings, among others.

Interaction with the dashboard interface is intentionally kept minimal to avoid driver distraction. The focus is on information display rather than interaction. However, certain interactive elements, such as clicking on a particular gauge for more detailed data or settings, are incorporated. For example, a user could tap on the fuel gauge to view detailed fuel economy statistics.

In addition, the dashboard interface integrates alerts and notifications. These are triggered in response to specific events or when certain thresholds are reached—like critical fuel levels or excessive speed. These alerts appear as pop-up messages on the dashboard, thereby ensuring the driver is promptly informed of any crucial vehicle parameters.

4.4 GPS Integration

The GPS feature in the mobile app is multifunctional, acting both as a navigation tool and for vehicle tracking. Thus, the app seamlessly merges ease of use and safety features.

By tapping into the car's built-in GPS unit, the system captures live location details. The Arduino microcontroller processes this data before it is sent to the mobile app via Bluetooth. Within the app, this information is interpreted, placing the car's current position on a map using a built-in map service. This GPS feature primarily caters for two requirements. One, it acts as a navigation tool, offering step-by-step directions. Users simply set their desired destination in the app, and the map service plots the best path. The journey is visualized on the map, with live updates. On the other hand, this GPS feature also bolsters security by continuously monitoring the car's whereabouts. If the car is lost or stolen, this system can pinpoint its location. Additionally, the system logs the vehicle's past locations, further reinforcing its safety credentials.

From a technical point of view, the challenge was in guaranteeing smooth communication between the GPS unit, the Arduino, and the app, along with embedding the map service. Special attention was dedicated to the precision of the GPS readings and ensuring that the map interface reacted swiftly.

4.5 Real-Time Passenger Detection

The Live Passenger Detection system stands out in the mobile app, giving users an immediate view of which seats are occupied in the vehicle. It's more than just a feature; it enhances awareness, security, and safety, complementing the multifaceted mobile control application. At its foundation, this system relies on pressure-sensitive sensors embedded in the car seats. These sensors gauge the weight on each seat, with the Arduino microcontroller deciphering this data to determine occupancy.

Once processed, this information is transmitted to the mobile app. Here, users see a current visual layout of the car's seating. Seats with passengers light up, offering a quick glance at where everyone is seated. This display is straightforward and clear, using easily understood icons and color distinctions to mark filled and vacant seats.

In terms of its utility, the Live Passenger Detection system serves a dual role. It enhances a driver's grasp of who's in the car, invaluable in bigger vehicles like vans, minibuses or buses. Additionally, it bolsters safety by alerting the driver if a seat suddenly empties, signaling potential issues like a passenger not seated securely or a child seat concern.

Developing this feature required careful calibration of the seat sensors to ensure accurate detection of occupancy. The Arduino code had to be written to reliably interpret sensor data and communicate this information to the mobile application. On the application side, the interface for displaying the seating layout had to be designed and integrated.

4.6 Bluetooth-Controlled Window Operation

The Bluetooth-Controlled Window Operation feature extends the reach of the mobile application control into the physical aspects of the vehicle, offering a level of convenience and control that merges the digital and physical realms of the vehicle environment.

At the heart of this feature are servo motors installed within the window mechanism of the vehicle. These motors are controlled by the Arduino microcontroller, which interprets command signals sent from the mobile application and translates them into physical movement of the windows.

The mobile app offers a window control feature, allowing users to adjust individual or multiple windows as they wish. In the app, each window's status is shown with an easy-to-use slider, making the interface straightforward.

In designing this feature, safety and prevention of unintentional operations were paramount. Users must confirm any window adjustment to avoid inadvertent changes due to accidental taps. Additionally, there's a built-in safeguard: if anything obstructs a window while it's moving, it'll automatically stop, minimizing the risk of harm or damage.

The Bluetooth Window Control System's foundation lies in creating a steadfast Bluetooth connection between the app and the Arduino microcontroller. It was crucial to ensure precise transmission of commands and that the servo motors acted upon them correctly.

5 System Testing and Evaluation

Ensuring system testing and evaluation is pivotal in the development of any system. It's essential to ascertain that each module operates optimally and that, collectively, they function seamlessly in diverse situations. Our methodology was systematic, encompassing both isolated module evaluations and comprehensive system assessments.

5.1 Component Testing

The RFID-centered Secure Keyless Car Ignition System was subjected to numerous tests, emphasizing factors like range sensitivity and tag recognition time. Our primary aim was to ensure that the ignition would be activated exclusively when an authenticated RFID tag was detected within a specified distance.

The Mobile Control App was put through rigorous user interface (UI) and user experience (UX) evaluations, assessing its versatility, reaction speed, and adaptability on various mobile platforms and devices. We devised multiple simulated situations

to appraise the dependability and precision of features like the dashboard, GPS capability, real-time passenger identification system, and Bluetooth-aided window adjustment feature.

The modules underwent testing in environments mirroring real-world user settings. Parameters such as driving speed, number of occupants, window adjustment velocity, and GPS signal potency were altered systematically to gauge consistent performance. The Mobile Control App was further scrutinized through a set of functionality tests, emphasizing its promptness, user-friendliness, and reliability. Additional tests mimicking conditions like weak signal reception or intense data traffic were also executed to determine the app's resilience.

Individual and combined assessments were carried out for the Dashboard Display, GPS Integration, Instantaneous Passenger Identification, and Bluetooth-enabled Window Adjustment. Our emphasis was on confirming the precision of data shown, the swiftness of commands, and the uninterrupted operation of assimilated capabilities, including alert mechanisms.

5.2 Overall System Testing

Following module evaluations, the whole system underwent a holistic review, simulating real-time operational scenarios: activating the car, employing the mobile app to observe and regulate different factors, navigating via the GPS, and window adjustments.

Furthermore, we exposed the system to intense tests, encompassing scenarios like back-to-back start/stop cycles, concurrent utilization of various app features, and functionality in diverse environmental backdrops. The ultimate goal was to guarantee that the system remained resilient and efficient, even when facing heightened demands or less-than-ideal conditions.

6 Results and Discussion

This section presents a comprehensive analysis of the empirical results obtained during the testing and evaluation phase. The insights derived here serve to substantiate the system's efficacy and utility and provide a broader perspective on its implications.

Our RFID-based Secure Keyless Car Ignition System demonstrated excellent performance, promptly recognizing authorized RFID tags within the designated proximity range and successfully activating the car ignition. Importantly, it provided robust security by refusing to engage when unauthorized RFID tags were tested. This indicates a successful implementation of a secure keyless ignition system, significantly enhancing the security profile over traditional key-based systems.

Our tests confirmed that the Dashboard Interface provided an intuitive and user-friendly method of monitoring various vehicle parameters. GPS Integration effectively served its dual role as a navigation aid and vehicle tracking feature. Real-Time Passenger Detection offered the driver an overview of seat occupancy, a useful safety feature that aligns with our hypothesis on enhancing in-vehicle safety awareness. Finally, Bluetooth-Controlled Window Operation was shown to be an effective method of remote window control, demonstrating that digital controls can successfully interface with the physical environment of a vehicle.

While these results align well with our objectives, we also identified potential areas for improvement. For instance, the Mobile Application's performance may be affected under low network coverage or in data-congested areas, which could result in latency in data representation or control responses.

Strengths of our system lie in its integrative approach, combining multiple control and information features within a single mobile application, and the robust performance of the RFID-based Secure Ignition System. However, the system's dependence on reliable network coverage and potential vulnerability to advanced security threats can be considered as its weaknesses.

Overall, the test results and subsequent analysis confirm that our system achieves its objectives, providing a secure and comprehensive control and monitoring system for an Arduino-based instrumented car. However, it is important to keep iterating on the design and implementation to address potential weaknesses and adapt to future needs and technologies.

7 Conclusion and Future Work

In this pioneering endeavor, we have successfully developed and implemented an RFID-based Secure Keyless Car Ignition System and a comprehensive, user-friendly mobile application interface for an Arduino-based instrumented car. The overarching aim, enhancing the security landscape of vehicular controls while simultaneously enriching the user experience, has been productively realized in this research work.

Delving into the secure keyless ignition system, we employed the intricate mechanics of RFID technology to mitigate the security drawbacks associated with traditional key-based ignition systems. The RFID system fortified the car's ignition security by adding layers of digital authentication. This technology's successful implementation marks a significant stride in our quest for enhanced vehicular security. Alongside this, our study explored the domain of integrated controls with a user-centric lens. The creation of a comprehensive mobile application interface harbored the functionalities of various car operations, forming an interconnected ecosystem that fosters efficiency and ease of use. This integration amalgamated functionalities such as ignition, GPS navigation, real-time passenger detection, and Bluetooth-controlled window operations under a singular, intuitive platform.

Despite the successful implementation of the system components, our evaluation and testing process brought forth areas warranting further improvement and research.

For instance, enhancing the sensitivity and reliability of the RFID system could yield better user experience and even greater security. Also, making the mobile application less power-intensive and more resilient to network fluctuations can significantly elevate the system's robustness.

As we peer into the future, the potential to imbue our system with more technologically advanced features is abundant. Emerging technologies like voice recognition, AI-driven predictive analysis, and biometric authentication can introduce an added layer of personalization and security. Furthermore, improving power management for both the RFID system and mobile application could lead to less energy consumption and reliance on external power sources.

Lastly, as vehicular technology continues its rapid evolution, the potential for integration with other smart city infrastructures or vehicle-to-vehicle (V2V) communication systems also presents a promising avenue for research. As we stand at this junction, the findings from our study serve as a beacon, illuminating the path toward a future where vehicular control systems are not only secure and user-friendly but are also an integral part of an interconnected, intelligent transportation network.

References

1. Dang T, Gupta V, Wadia DS, Kohli P, Chahal RK (2021) FaceIgnition: an automatic anti-theft and keyless solution for vehicles. In: International conference on computational intelligence and knowledge economy (ICCIKE). IEEE, pp 248–253
2. Rupanagudi SR et al (2016) A novel and secure methodology for keyless ignition and controlling an automobile using air gestures. In: International conference on advances in computing, communications and informatics (ICACCI). IEEE, pp 1416–1422
3. Wetzels J (2014) Broken keys to the kingdom: security and privacy aspects of RFID-based car keys. [arXiv:1405.7424](https://arxiv.org/abs/1405.7424)
4. Heikkinen J, Mäkinen E, Lylykangas J, Pakkanen T, Väänänen-Vainio-Mattila K, Raisamo R (2013) Mobile devices as infotainment user interfaces in the car: contextual study and design implications. In: Proceedings of the 15th international conference on human computer interaction with mobile devices and services. pp 137–146
5. Svangren MK, Skov MB, Kjeldskov J (2017) The connected car: an empirical study of electric cars as mobile digital devices. In: Proceedings of the 19th international conference on human-computer interaction with mobile devices and services. pp 1–12
6. Yang F, Li J, Lei T, Wang S (2017) Architecture and key technologies for Internet of vehicles: a survey. *J Commun Inf Netw* 2(2):1–17
7. Figueiredo L, Jesus I, Machado JAT, Ferreira JR, De Carvalho JLM (2001) Towards the development of intelligent transportation systems. In: ITSC 2001. IEEE intelligent transportation systems. Proceedings (Cat. No. 01TH8585). pp 1206–1211
8. Thakuriah P, Geers DG (2013) Transportation and information: trends in technology and policy. Springer
9. Siegel JE, Erb DC, Sarma SE (2017) A survey of the connected vehicle landscape—architectures, enabling technologies, applications, and development areas. *IEEE Trans Intell Transp Syst* 19(8):2391–2406
10. Badue C et al (2021) Self-driving cars: a survey. *Expert Syst Appl* 165:113816
11. Kondaveeti HK, Kumaravelu NK, Vanambathina SD, Mathe SE, Vappangi S (2021) A systematic literature review on prototyping with Arduino: applications, challenges, advantages, and limitations. *Comput Sci Rev* 40:100364. <https://doi.org/10.1016/J.COSREV.2021.100364>

12. Shen X et al (2016) Teleoperation of on-road vehicles via immersive telepresence using off-the shelf components. In: Intelligent autonomous systems 13: proceedings of the 13th international conference IAS-13. Springer, pp 1419–1433
13. Grochowski M, Schweigler M, Alrifaae B, Kowalewski S (2018) A GPS-aided inertial navigation system for vehicular navigation using a smartphone. IFAC-PapersOnLine 51(10):121–126. <https://doi.org/10.1016/J.IFACOL.2018.06.247>
14. Patel P, Narmawala Z, Thakkar A (2019) A survey on intelligent transportation system using internet of things. In: Emerging research in computing, information, communication and applications: ERCICA 2018, vol 1. pp 231–240
15. Mukhopadhyay D, Gupta M, Attar T, Chavan P, Patel V (2018) An attempt to develop an IOT based vehicle security system. In: IEEE international symposium on smart electronic systems (iSES) (Formerly iNiS). IEEE, pp 195–198
16. Bagay D (2020) Information security of RFID tags. Procedia Comput Sci 169:183–186. <https://doi.org/10.1016/J.PROCS.2020.02.133>

Selection of Automated Guided Vehicles for Industrial Application Using Weighted Sum Method



Divya Sabnis , Megha Patil , and Sagar Wankhede 

Abstract For material handling and transportation in various industrial settings automated guided vehicles (AGVs) have emerged as a cutting-edge solution. The diverse requirements of different industries the selection of the most suitable AGV for a specific application is a crucial task. To assist decision makers this study is helpful in their decision-making process. The proposed approach utilizes a multi-criteria decision-making technique, Weighted Sum Method (WSM) that enables the evaluation and comparison of AGVs using multiple criteria. The proposed methodology involves defining relevant criteria such as controllability, accuracy, range, reliability, flexibility, and cost. Accuracy of model predicted by exploring data on the performance and characteristics of various AGVs are collected from reliable sources. The gathered data are then normalized and assigned appropriate weights by domain experts to reflect the relative importance of each criterion. The weighted scores for each criterion are combined using the WSM to obtain an overall score for each AGV model under consideration. For specific material handling applications, the AGV model with the highest overall score predicted as the most suitable model. The proposed model reveals that the WSM offers a systematic and efficient approach to AGV selection, empowering decision-makers to make choices grounded in quantifiable criteria rather than intuition or incomplete information. The proposed AGV selection methodology utilizing the Weighted Sum Method presents a structured decision-making process, which is suitable for industrial applications.

Keywords Material handling · AGV · MCDM · WSM · WASPAS

D. Sabnis · M. Patil · S. Wankhede (✉)

School of Mechatronics Engineering, Symbiosis Skills and Professional University, Pune 412101, India

e-mail: svw8890@gmail.com

1 Introduction

The essential element in contemporary manufacturing systems is automated guided vehicles (AGVs), playing a great importance in facilitating material handling processes. The automated guided vehicles are highly autonomous and efficient, self-driving, and battery operated typically controlled by on board computers. The primary function of AGV is specifically facilitating horizontal movements of goods and transport materials within a facility. The AGVs are firstly introduced in 1955 [1] till now it is continuously growing, expanding in different applications. Nowadays for optimization of material flow and handling in industries, AGVs are playing very essential role for preparation of “fleet size” in manufacturing industries specifically in workstations and storehouses. In industries for material transportation throughout facility, specific task, and responsibilities assigned to AGV. AGV found in very good at doing different applications in manufacturing plants, distribution centers, transshipment hubs, warehouses, and even external transportation areas. The ability of AGVs is to do many different things and the quality of being able to change at any situation in logistics and material handling processes. Due to effective use of AGVs in industry manual intervention is reduced and operational efficiency and safety reduced. For efficiently transporting goods from one location to other by avoiding obstacles following navigated paths AGVs works effectively and efficiently. The positive effect of AGVs implementation in manufacturing industries can experience increased productivity, reduced operational costs, and improved overall workflow. A reliable and flexible solution for material transportation in supply chain management is effective use of AGVs. In advanced AGV new features are incorporated like artificial intelligence, advanced sensors, and communications systems. The modern manufacturing and logistic ecosystems improve their performance by making AGVs as integral part in their system.

In the Flexible Manufacturing Systems (FMS) the automated guided vehicle plays pivotal role due its high flexibility. In a constantly evolving manufacturing environment advanced AGVs efficiently handle material transportation in manufacturing plant. Scholars like Tompkins and White [2] have demonstrated the important role of material handling systems in Flexible Manufacturing Systems (FMS). In industry for achieving operational efficiency goals and cost reduction, material handling operations are critically handled as material handling consumes 13–30% total production costs. Consequently, since the 1980s, numerous researchers have been actively seeking new approaches to optimize material handling systems (MHS) within FMS settings. In a Flexible Manufacturing System, multiple cells are responsible for producing different parts, each with specific requirements and processing needs. Having an efficient MHS becomes paramount in such a setup, as seamless and timely transportation of materials between workstations and even among cells directly impacts cost reduction goals. Given the highly competitive nature of the current market conditions, cost reduction is one of the most important objectives for manufacturing businesses to maintain their competitiveness and profitability. By employing AGVs in the material handling process, manufacturers can achieve

several benefits. The ability to modify the guide paths of AGVs easily allows for swift reconfiguration of material transportation routes, adapting to changes in production needs without significant disruptions. This inherent flexibility ensures a smooth and uninterrupted workflow in an FMS environment. Furthermore, AGVs contribute to reducing production costs by optimizing material flow, minimizing waiting times, and enhancing overall operational efficiency. As AGVs can autonomously navigate through the facility, they eliminate the need for manual material transportation, reducing labor costs, and potential errors. The streamlined material handling process translates to faster production cycles, allowing manufacturers to meet market demands more effectively. Within an AGV system, several key elements play critical roles, contributing to its efficient, and seamless operation. These elements can be classified into three primary components: the vehicles, the transportation network, and the physical interface that connects the production/storage system with the control system.

The transportation network functions as the fundamental framework of the AGV system, interconnecting all stationary installations like machines and workstations within the central area of the facility. These stations establish crucial pickup and delivery points (P/D points), serving as interfaces that connect the production system with the AGV transportation system. These P/D points act as the key exchange locations where materials are loaded onto or unloaded from the AGVs. AGVs, as the core mobile units of the system, travel between the P/D points, efficiently transporting materials and goods from one location to another. The movement of AGVs can follow either fixed paths, predefined by guide wires or floor markings, or free paths, where they navigate autonomously without the need for physical guidance. AGVs operating without guide paths are referred to as free-ranging AGVs [3]. This distinction in path-following mechanisms allows for flexibility in configuring AGV systems to match the specific requirements of different industrial environments. For accurate and safe navigation throughout the facility guide paths are allotted to AGVs. In critical in busy production environments guide path ensures precise movement and minimizing the chances of deviation and collisions. On the other hand, free-ranging AGVs operate without the constraints of fixed paths. Free-ranging AGVs leverage advanced technologies, such as laser sensors, cameras, or LIDAR, to autonomously navigate through the space. So free-ranging AGVs are flexible enough to adapt to changes in the environment, making them ideal for environments where the layout may be subject to frequent adjustments or in situations where guide paths are impractical to implement. An AGV system's design is considered for tactical and operational aspects. Tactical issues include points like pickup/delivery locations, fleet size, and flow path architecture. Operational issues include routing and dispatching considerations. Several researchers have addressed these challenges in the past. For instance, Hsueh [4], King and Wilson [5], Ganesharajah and Sriskandarajah [6], Johnson and Brandeau [7], Manda and Palekar [8], and Bordelon Hoff and Sarker [9] have demonstrated operational issues related to AGV dispatching, routing, and scheduling. King and Wilson specifically investigated the design, routing, and scheduling of AGVs, focusing on vehicle requirements, flow

paths, and AGV types. Ganesharajah and Sriskandarajah investigated the complexities associated with scheduling, dispatching, and routing of AGVs in various flow route topologies. Johnson and Brandeau demonstrated stochastic models for automated material handling systems design and control, while Manda and Palekar explored their experimentation into MHS design and control problems. Additionally, Hoff and Sarker examined dispatching standards and the creation of guidance paths for AGVs. More recently, Qiu et al. [10] conducted a literature review on design and operational issues related to Flexible Manufacturing Systems (FMS), as well as tactical challenges concerning fleet size and flow path design. Moreover, some researchers, like [2, 11], have integrated modern technologies such as Mechatronics to enhance AGV systems. They utilized various electrical devices, including chipsets, boards, RFIDs, sensors, and other advanced components to increase system throughput and prevent deadlocks and collisions. Overall, the AGV system design has been shaped by insights gained from previous research on tactical and operational considerations, as well as the integration of cutting-edge technologies to optimize its performance. Their approach has the highest quality currently attained, but they failed to account for the expense of creating such a fully automated system. In this context, the Weighted Sum Method (WSM) emerges as an effective and reliable multi-criteria decision-making technique. The WSM allows decision-makers to comprehensively assess and rank AGV models based on a combination of multiple criteria, each assigned a specific weight reflecting its relative importance. By utilizing the WSM/MCDM, the AGV selection process becomes more objective, systematic, and transparent, enabling stakeholders to make informed choices backed by quantitative analysis rather than relying solely on subjective judgments [12–15]. This research aims to propose a structured methodology for selecting AGVs in industrial applications using the Weighted Sum Model. By integrating key criteria and applying appropriate weighting, the proposed model will assist decision-makers in comparing AGV models and identifying the one that best meets the specific requirements of their material handling operations. The research endeavors to validate the effectiveness and practicality of the WSM approach through a real-world case study, highlighting the advantages of employing a data-driven decision-making process in AGV selection.

2 Weighted Sum Method

A multi-criteria decision-making method called the Weighted Sum Method is used to assess and contrast several choices based on a variety of factors. It is commonly used in various fields such as engineering, business, project management, and operations research when there are multiple factors or attributes to consider in decision-making.

The method involves the following steps:

- (i) **Identify Criteria:** List the pertinent standards by which the alternatives will be judged. These standards ought to be quantifiable and cover the essential elements of the decision-making process.
- (ii) **Assign Weights:** Give each criterion a weight based on how important or important it is to the decision-maker. All weights added together should equal 1 (or 100%).
- (iii) **Normalize Criteria:** If the criteria are on different scales, it is necessary to normalize them to bring them to a common scale, usually between 0 and 1, for better comparison.
- (iv) **Evaluate Alternatives:** Evaluate each alternative with respect to each criterion and assign scores to them based on their performance in each criterion.
- (v) **Calculate Weighted Scores:** Multiply the scores of each alternative by the corresponding weight of the criterion and then sum up these weighted scores for each alternative. This yields a single value representing the overall performance of the alternative considering all criteria.
- (vi) **Rank Alternatives:** Based on their combined weighted scores, order the alternatives. The option with the greatest weighted score is regarded as the best or preferred option.
- (vii) **Make Decision:** The alternative with the highest rank (or highest weighted score) is the one that best suits the decision-maker’s objectives and preferences.

2.1 WSM Formulations

Step 1: Design of a weighted decision matrix

Let’s say there are m choices and n criteria in an MCDM situation, $D = x_{ij}$ be a decision matrix, where $x_{ij} \in \mathbb{R}$ [refer Table 1].

$$D_{ij} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix} \tag{1}$$

It is possible to express the weight vector as

$$w_j = [w_1 \dots w_n], \quad \sum_{j=1}^n (w_1 \dots w_n) = 1. \tag{2}$$

Table 1 Choice matrix for selection of AGV

AGV alternative	Controllability	Accuracy	Range	Reliability	Flexibility	Cost
V1	0.895	0.495	0.495	0.895	0.295	0.695
V2	0.115	0.895	0.895	0.495	0.495	0.895
V3	0.115	0.115	0.115	0.695	0.895	0.895
V4	0.295	0.895	0.495	0.495	0.895	0.115
V5	0.895	0.495	0.695	0.295	0.495	0.115
V6	0.495	0.495	0.115	0.695	0.695	0.895
V7	0.115	0.295	0.115	0.895	0.895	0.895
V8	0.115	0.495	0.495	0.495	0.695	0.695
The weights are calculated using analytical hierarchy approach (AHP)						
Criteria weights	0.346	0.168	0.073	0.063	0.293	0.0584

Step 2: Normalization of decision matrix

$$n_{ij} = \begin{cases} \frac{x_{ij}}{\max .x_{ij}} & |j \in B \\ \frac{\min .x_{ij}}{x_{ij}} & |j \in C \end{cases} \tag{3}$$

where n_{ij} is the normalized value of the i th alternative for the j th criterion. $\max .x_{ij}$ and $\min .x_{ij}$ are the maximum and minimum value of x_{ij} in the j th column for benefit (B) and cost criterion (C), respectively.

Table 2 Normalized decision matrix

AGV alternative	Controllability	Accuracy	Range	Reliability	Flexibility	Cost
V1	1	0.5530726	0.553073	1	0.329609	0.165468
V2	0.128492	1	1	0.553073	0.553073	0.128492
V3	0.128492	0.1284916	0.128492	0.776536	1	0.128492
V4	0.329609	1	0.553073	0.553073	1	1
V5	1	0.5530726	0.776536	0.329609	0.553073	1
V6	0.553073	0.5530726	0.128492	0.776536	0.776536	0.128492
V7	0.128492	0.3296089	0.128492	1	1	0.128492
V8	0.128492	0.5530726	0.553073	0.553073	0.776536	0.165468

Table 3 Weighted normalized decision matrix

AGV alternative	Controllability	Accuracy	Range	Reliability	Flexibility	Cost
V1	0.346	0.092916	0.036135	0.056385	0.086435	0.040588
V2	0.044458	0.168	0.065335	0.031185	0.145035	0.052268
V3	0.044458	0.021587	0.008395	0.043785	0.262235	0.052268
V4	0.114045	0.168	0.036135	0.031185	0.262235	0.006716
V5	0.346	0.092916	0.050735	0.018585	0.145035	0.006716
V6	0.191363	0.092916	0.008395	0.043785	0.203635	0.052268
V7	0.044458	0.055374	0.008395	0.056385	0.262235	0.052268
V8	0.044458	0.092916	0.036135	0.031185	0.203635	0.040588

Step 3: Weighted normalization of decision matrix

$$W_{nij} = w_j n_{ij} \tag{4}$$

Step 4: Ranking of alternatives

$$S_i^{WSM} = \sum_{j=1}^n w_j n_{ij} \tag{5}$$

where S_i^{WSM} the ranking score of the i th alternative, w_j is the weight of the criterion. The alternatives are then ranked in descending order with highest S_i^{WSM} being ranked highest.

3 Results and Discussion

After calculating the preference scores for each AGV, we rank them based on their overall performance. The AGV with the highest weighted score is considered the most suitable choice for the industrial application and compared ranking of AGV alternative with WASPAS [refer Table 4].

Based on the Weighted Sum Method, the most suitable AGV stands out with ranking as $V5 > V1 > V4 > V6 > V2 > V7 > V8 > V3$ for the industrial application, with the highest overall preference score. It offers a good combination of controllability, accuracy, range, reliability, flexibility, and cost.

Table 4 Rank of AGV alternative using weighted aggregated sum product assessment (WASPAS) and WSM

AGV alternative	Preference score	WSM	WASPAS
V1	0.658459201	2	2
V2	0.506281101	5	5
V3	0.432727693	8	8
V4	0.618315693	3	3
V5	0.659987201	1	1
V6	0.59236233	4	4
V7	0.479115402	6	6
V8	0.448917302	7	7

4 Conclusion

The Weighted Sum Method has served as a valuable tool for AGV selection in the industrial application, leading us to make an informed and rational decision that aligns with the desired criteria and objectives. A5 has been identified as the most favorable choice, ensuring efficient and reliable material handling operations while considering the various constraints and preferences relevant to the industrial setting.

The Weighted Sum Method allows decision-makers to objectively assess and prioritize AGVs based on various criteria relevant to the industrial application, leading to an optimal choice that aligns with the specific requirements and preferences of the organization.

Conflict of Interest Authors do not have any conflict of interest.

References

1. Bilge Ü, Tanchoco JMA (1997) AGV systems with multi-load carriers: basic issues and potential benefits. *J Manuf Syst* 16:159–174. [https://doi.org/10.1016/S0278-6125\(97\)88885-1](https://doi.org/10.1016/S0278-6125(97)88885-1)
2. Tompkins JA, White JA (1984) *Facility planning*. Wiley, New York
3. Co CG, Tanchoco JMA (1991) A review of research on AGVS vehicle management. *Eng Costs Prod Econ* 21:35–42. [https://doi.org/10.1016/0167-188X\(91\)90016-U](https://doi.org/10.1016/0167-188X(91)90016-U)
4. Hsueh C-F (2010) A simulation study of a bi-directional load-exchangeable automated guided vehicle system. *Comput Ind Eng* 58:594–601. <https://doi.org/10.1016/j.cie.2009.12.008>
5. King RE, Wilson C (1991) A review of automated-guided vehicle systems design and scheduling. *Prod Plan Control* 2:44–51. <https://doi.org/10.1080/09537289108919329>
6. Ganesharajah T, Sriskandarajah C (1995) Survey of scheduling research in AGV-served manufacturing systems. In: *Proceedings of the instrumentation systems automation technical conference (IAS '95)*. 50, Toronto, Canada, pp 87–94
7. Johnson ME, Brandeau ML (1996) Stochastic modeling for automated material handling system design and control. *Transp Sci* 30:330–350. <https://doi.org/10.1287/trsc.30.4.330>
8. Manda BS, Palekar US (1997) Recent advances in the design and analysis of material handling systems. *J Manuf Sci Eng* 119:841–848. <https://doi.org/10.1115/1.2836833>

9. Bordelon Hoff EB, Sarker BR (1998) An overview of path design and dispatching methods for automated guided vehicles. *Integr Manuf Syst* 9:296–307. <https://doi.org/10.1108/09576069810230400>
10. Qiu L, Hsu W-J, Huang S-Y, Wang H (2002) Scheduling and routing algorithms for AGVs: a survey. *Int J Prod Res* 40:745–760. <https://doi.org/10.1080/00207540110091712>
11. Müller T (1983) *Automated guided vehicles*. Springer, International Foundation for Science, Berlin, Germany
12. Pesode P, Barve S, Wankhede SV, Jadhav DR, Pawar SK (2023) Titanium alloy selection for biomedical application using weighted sum model methodology. *Mater Today Proc* 72:724–728. <https://doi.org/10.1016/j.matpr.2022.08.494>
13. Wankhede SV, Hole JA (2022) MOORA and TOPSIS based selection of input parameter in solar powered absorption refrigeration system. *Int J Ambient Energy* 43:3396–3401. <https://doi.org/10.1080/01430750.2020.1831600>
14. Wankhede S, Pesode P, Gaikwad S, Pawar S, Chipade A (2023) Implementing combinative distance base assessment (Codas) for selection of natural fibre for long lasting composites. *Mater Sci Forum* 1081:41–48. Trans Tech Publications Ltd. <https://doi.org/10.4028/p-4pd120>
15. Wankhede S, Pesode P, Pawar S, Lobo R (2023) Comparison study of GRA, COPRAS and MOORA for ranking of phase change material for cooling system. *Mater Today Proc*. <https://doi.org/10.1016/j.matpr.2023.02.437>

Active Disturbance Rejection Control of a SEPIC Converter



Priyanshu Kumar and Moina Ajmeri

Abstract Single-ended primary inductor converter(SEPIC) is a high-gain buck-boost converter, but its controller's design is a challenging task as it is a fourth-order nonlinear system. This paper proposes an active disturbance rejection control (ADRC)-based controller for a SEPIC converter with high switching frequency. ADRC is an advanced control technique that does not rely on the exact information of the system as all the external and internal disturbances are estimated as separate variables and are canceled directly by the effect of the controller. Simulation results suggest that the proposed ADRC technique gives good setpoint tracking and robustness toward disturbances like input voltage fluctuations and load variations. The performance parameters have been calculated and analyzed to conclude that the proposed method has stable and robust performance.

Keywords DC-DC · SEPIC · PWM · ADRC · ESO

1 Introduction

The extraction of energy from renewable sources like solar and wind requires a DC-to-DC converter. Other applications for DC-to-DC converters are in DC motor drives personal communication equipment and power Computers, etc. [18]. DC-to-DC converters are of three types based on the transformer action, i.e., buck, boost, and buck-boost. Based on the requirement of output voltage with respect to the input, each of the converters has specific uses. Buck-boost converter topologies like cuk converter, zeta converter converters are suited for solar energy generation systems, where the input voltage keeps on fluctuating depending on the intensity of sunlight [5, 14, 23]. However, the low energy conversion efficiency, due to the hard-switched

P. Kumar (✉) · M. Ajmeri
Electrical Engineering Department, National Institute of Technology Patna, Patna, India
e-mail: priyanshuk.phd19.ee@nitp.ac.in

M. Ajmeri
e-mail: moina@nitp.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
R. N. Shaw et al. (eds.), *Innovations in Electrical and Electronic Engineering*, Lecture Notes in Electrical Engineering 1109, https://doi.org/10.1007/978-981-99-8289-9_28

367

states, inverted output, and low voltage gain are the drawbacks of the Cuk and Zeta and some other buck-boost converters [21].

The above-noted problems can be handled by the use of single-ended primary inductor converter (SEPIC) converters [3]. These converters are suitable for off-grid solar power plants due to the possibility of connecting it to the various batteries and PV applications, where they can match the characteristics of current and voltage [3, 10, 22]. SEPIC converter is a fourth-order nonlinear system whose behavior depends on operating conditions like input voltage, duty cycle, and load variations. It requires an advanced control technique to meet goals of guaranteed stability, good set-point tracking, efficient and fast attenuation of the load disturbance and satisfactory robustness toward parametric variations. Various nonlinear control methods such as back-stepping and passivity-based control [2], fuzzy logic-based control [6], sliding mode control [9, 20] have been reported in the literature for controlling SEPIC converters. Sliding mode control is known for better robustness; however, methods proposed in [9, 20] resulted in slow response, and they also failed to reject disturbances of large magnitudes. Authors have designed indirect sliding mode control for the SEPIC using the current mode control in [16]. The output response is satisfactory in the above-cited work however the tuning method looks lengthy and complex. They have not suggested any explicit formulae for calculating the input current reference. It is observed that there is a need for a simple and efficient control strategy for the SEPIC converter as it has a wide range of applications. The two degree of freedom internal model control (TDF-IMC) [15] has been recently reported in the literature for the boost converter. It is a plant model-dependent scheme consisting of three control blocks which are tuned using two design parameters. ADRC control technique is a robust control technique, and it doesn't require the exact knowledge of the system to be controlled [4, 13, 24]. ADRC method draws attention owing to its excellent performance in satisfying the aforementioned control objectives [1]. Up to the best of the authors' knowledge, the ADRC method has not been implemented on the SEPIC converter.

Further, it is important to mention here that larger values of the components such as capacitors and inductors of SEPIC cause significant power losses. At higher switching frequencies, these components' size reduces and also the loading effect of external filter components decreases which results in a faster dynamic response. However, designing a converter at higher frequencies such as in the MHz range limits its operating voltage range as the voltage stresses across the switches and diodes are also increased [11]. Therefore, operational voltage ranges at very high frequencies may not be suitable for power converters applications. Thus, for a wide range of operations, a switching frequency close to a few hundred kHz is commonly used.

In the present work, active disturbance rejection control (ADRC) is proposed for the SEPIC converter with high switching frequency. For the effective application of this controller, the order of the converter is reduced to the second-order system using a balanced reduction technique based on the calculation of the Hankel singular value. The tuning parameter has been decided on the basis of the bandwidth parametrization technique described in [8]. Also, the suitable values of the tuning parameters are selected according to the suitable value of maximum sensitivity. The

designed controller has been verified using simulation results. The controller provides good setpoint tracking and disturbance rejection proving its stability robustness. The robustness of the closed-loop system has been proved on the basis of lower values of settling time, overshoot/undershoot voltage, peak inductor current, and integral square error (IAE).

This paper is divided into seven subsections. Section 2 describes the modeling and design of the SEPIC converter. Section 3 explains the complete design of the ADRC controller for the SEPIC converter. Simulation results are presented in Sect. 4. Finally, the conclusion is derived in Sect. 5.

2 SEPIC Converter

Single-ended primary inductor converter (SEPIC) is a buck-boost converter with noninverted output. It has also the unique property of isolating the input–output circuit when no gating signal is provided to the switch. The circuit diagram of the SEPIC is shown in Fig. 1. It consists of a primary inductor (L_1), a coupling capacitor (C_1), a secondary inductor (L_2), a switch (Q), a diode (D) and an output capacitor (C_2). The circuit works as a boost converter if the duty cycle (D) is greater than 0.5 and as a buck converter if the duty cycle is less than 0.5. The output and inputs are equal when the duty cycle is exactly 0.5. Practically the voltage drops of the diode and MOSFET affect the output voltage. The gate terminal of the switch is provided a pulse width modulated (PWM) generated according to the duty cycle. Hence, the operation of the SEPIC can be divided into two modes as described further.

2.1 Switch on Mode

The circuit diagram resembling this mode is shown in Fig. 2. When the switch is ON, it acts like a short circuit and the diode is reverse-biased. Inductor (L_1) charges through

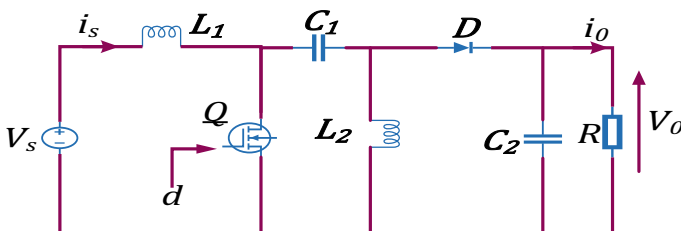


Fig. 1 Schematic diagram of SEPIC converter

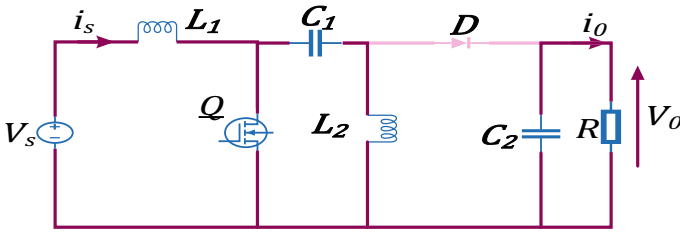


Fig. 2 Schematic diagram of SEPIC converter in switch ON mode

the source, while the capacitor supplies the stored energy to the load maintaining the output voltage (V_0) to be constant. Also, the inductor (L_2) is charged by the capacitor (C_1).

The differential equations corresponding to this mode is:

$$\frac{di_{L_1}}{dt} = \frac{v_s}{L_1} \tag{1}$$

$$\frac{di_{L_2}}{dt} = \frac{v_{c_1}}{L_2} \tag{2}$$

$$\frac{dv_{c_1}}{dt} = -\frac{i_{L_2}}{C_1} \tag{3}$$

$$\frac{dv_{c_2}}{dt} = -\frac{v_{c_2}}{C_2 R} \tag{4}$$

The equations in matrix form can be written as:

$$\begin{bmatrix} \dot{i}_{L_1} \\ \dot{i}_{L_2} \\ \dot{v}_{c_1} \\ \dot{v}_{c_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1/L_2 & 0 \\ 0 & -1/C_1 & 0 & 0 \\ 0 & 0 & 0 & -1/RC_2 \end{bmatrix} \begin{bmatrix} i_{L_1} \\ i_{L_2} \\ v_{c_1} \\ v_{c_2} \end{bmatrix} + \begin{bmatrix} 1/L_1 \\ 0 \\ 0 \\ 0 \end{bmatrix} [v_s] \tag{5}$$

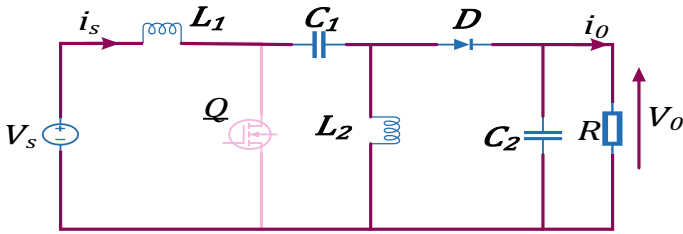


Fig. 3 Schematic diagram of SEPIC converter in switch OFF mode

2.2 Switch OFF Mode

When the switch is turned off by the means of the PWM signal, then L_1 and L_2 supplies the stored energy toward the load end thereby charging the capacitors C_1 and C_2 . The output voltage is still maintained constant when a constant duty ratio is used for the generation of the PWM signals. The circuit diagram defining this mode is shown in Fig. 3.

The differential equations describing this mode are:

$$\frac{di_{L_1}}{dt} = \frac{v_s}{L_1} - \frac{v_{c_1}}{L_1} - \frac{v_{c_2}}{L_1} \tag{6}$$

$$\frac{di_{L_2}}{dt} = \frac{v_{c_2}}{L_2} \tag{7}$$

$$\frac{dv_{c_1}}{dt} = \frac{i_{L_1}}{C_1} \tag{8}$$

$$\frac{dv_{c_2}}{dt} = \frac{(i_{L_1} + i_{L_2})}{C_2} - \frac{v_{c_2}}{C_2 R} \tag{9}$$

These equations can again be written as:

$$\begin{aligned} \begin{bmatrix} \dot{i}_{L_1} \\ \dot{i}_{L_2} \\ \dot{v}_{c_1} \\ \dot{v}_{c_2} \end{bmatrix} &= \begin{bmatrix} 0 & 0 & -1/L_1 & -1/L_1 \\ 0 & 0 & 0 & -1/L_2 \\ 1/C_1 & 0 & 0 & 0 \\ 1/C_2 & 1/C_2/C_2 & 0 & -1/RC_2 \end{bmatrix} \begin{bmatrix} i_{L_1} \\ i_{L_2} \\ v_{c_1} \\ v_{c_2} \end{bmatrix} \\ &+ \begin{bmatrix} 1/L_1 \\ 0 \\ 0 \\ 0 \end{bmatrix} [v_s] \end{aligned} \tag{10}$$

The differential equations obtained in ON and OFF modes are combined by multiplying D and $(1 - D)$ respectively using the state-space averaging technique to get the matrices of A, B, C, and D as:

$$A = DA_1 + (1 - D)A_2 \tag{11}$$

$$B = DB_1 + (1 - D)B_2 \tag{12}$$

Now the combined equations in matrix form can be written as:

$$\begin{bmatrix} \dot{i}_{L_1} \\ \dot{i}_{L_2} \\ \dot{v}_{c_1} \\ \dot{v}_{c_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -D'/L_1 & -D'/L_1 \\ 0 & 0 & D/L_2 & -D'/L_2 \\ D'/C_1 & -D/C_1 & 0 & 0 \\ D'/C_2 & D'/C_2 & 0 & -1/RC_2 \end{bmatrix} \times \begin{bmatrix} i_{L_1} \\ i_{L_2} \\ v_{c_1} \\ v_{c_2} \end{bmatrix} + \begin{bmatrix} 1/L_1 \\ 0 \\ 0 \\ 0 \end{bmatrix} [v_s] \tag{13}$$

where $D' = 1 - D$.

The calculation of the transfer function between the output voltage and duty cycle can be performed by using small signal analysis(SSA). Hence, after performing the SSA, the following state-space equations are obtained:

$$\begin{bmatrix} \hat{i}_{L_1} \\ \hat{i}_{L_2} \\ \hat{v}_{c_1} \\ \hat{v}_{c_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \frac{-D'}{L_1} & \frac{-D'}{L_1} \\ 0 & 0 & \frac{D}{L_2} & \frac{-D'}{L_2} \\ \frac{D'}{C_1} & \frac{-D}{C_1} & 0 & 0 \\ \frac{D'}{C_2} & \frac{D'}{C_2} & 0 & \frac{-1}{RC_2} \end{bmatrix} \begin{bmatrix} \hat{i}_{L_1} \\ \hat{i}_{L_2} \\ \hat{v}_{c_1} \\ \hat{v}_{c_2} \end{bmatrix} + \begin{bmatrix} \frac{(V_{c_1} + V_{c_2})}{L_1} \\ \frac{(V_{c_1} + V_{c_2})}{L_1} \\ \frac{-(I_{L_1} + I_{L_2})}{C_2} \\ \frac{-(I_{L_1} + I_{L_2})}{C_2} \end{bmatrix} [\hat{d}] + \begin{bmatrix} \frac{1}{L_1} \\ 0 \\ 0 \\ 0 \end{bmatrix} [\hat{v}_s] \tag{14}$$

and $\hat{v}_0 = [0 \ 0 \ 0 \ 1] \begin{bmatrix} \hat{i}_{L_1} \\ \hat{i}_{L_2} \\ \hat{v}_{c_1} \\ \hat{v}_{c_2} \end{bmatrix}$

Using the above state-space equations, and neglecting any parasitic resistances, the transfer function is calculated as:

$$\frac{v_0(s)}{d(s)} \approx \frac{\left(1 - s \frac{L_1}{R} \frac{D^2}{D'^2}\right) \left(1 - s \frac{C_1(L_1+L_2)R}{L_1} \frac{D^2}{D'^2} + s^2 \frac{L_2 C_1}{D}\right)}{D'^2 \left(1 + \frac{s}{\omega_{01} Q_1} + \frac{s^2}{(\omega_{01})^2}\right) \left(1 + \frac{s}{\omega_{02} Q_2} + \frac{s^2}{(\omega_{02})^2}\right)} \tag{15}$$

where

$$\omega_{01} = \frac{1}{\sqrt{L_1 \left(C_2 \frac{D^2}{D'^2} + C_1\right) + L_2(C_1 + C_2)}} \tag{16}$$

$$\omega_{02} = \sqrt{\frac{1}{L_2 \frac{C_1}{D^2} \parallel \frac{C_2}{D'^2}} + \frac{1}{L_1 C_1 \parallel C_2}} \tag{17}$$

$$Q_1 = \frac{R}{\omega_{01} \left(L_1 \frac{D^2}{D'^2} + L_2\right)} \tag{18}$$

$$Q_2 = \frac{R}{\omega_{02} (L_1 + L_2) \frac{\omega_{01}^2}{\omega_{02}^2}} \tag{19}$$

The design of the converter is dependent on the frequency of operation at which the switch operates. The higher switching frequency (f_s) allows the selection of reactive components like capacitors and inductors with smaller sizes. In this work, the operating switching frequency of the PWM signal selected is 100 kHz. The selected values of the circuit elements in the present work are listed in Table 1.

Substituting the calculated values for the components in equation (13), the transfer function of the original system is calculated as:

$$G_p(s) = \frac{-980s^3 + 1.9 \times 10^8 s^2 - 6 \times 10^{10} s + 9.6 \times 10^{15}}{s^4 + 20s^3 + 5.3 \times 10^7 s^2 + 1 \times 10^9 s + 3.2 \times 10^{13}} \tag{20}$$

Table 1 Specifications of SEPIC converter

Parameters	Numerical values
Input voltage	48 V
Output voltage	72 V
Duty ratio	0.6
L_1	0.5 mH
L_2	1 mH
C_1	10 μ F
C_2	100 μ F
Load	100 Ω
f_s	100 KHz

2.3 Reduced Order Model of SEPIC Converter

Being a fourth-order nonminimum phase system, the application of some of the advanced controlling techniques on SEPIC converter leads to complex equations and sluggish response. Hence, a need for the reduction of the order of the system arises. In this work, ADRC control method has been proposed for the converter. The controller design based on the fourth-order system induces greater phase lag that deteriorates the performance of the controller in the case of the transients. To overcome the above-said difficulties and to apply the ADRC method, this converter is reduced to a second-order system. The order of the system is reduced on the basis of a balanced reduction method on the calculations of Hankel singular values. Firstly, a balanced reduction of the system is done to isolate the states whose contribution to the input–output response is negligible[17]. After the reduction of the system, Hankel singular values are calculated which has N small entries. A scientist named Hermann Hankel designed a method to obtain Hankel singular values based on the controllability Gramian, and the observability Gramian[19]. Actually, Hankel singular values provide a measure of energy for each state in a system. Hankel singular values are calculated as the square roots, of the eigenvalues. They are the basis for balanced model reduction, in which high-energy states are retained while low-energy states are discarded. The second-order reduced transfer function is obtained as:

$$G_{pr}(s) = \frac{299.9s^2 + 146.5s + 1.56 \times 10^{10}}{s^2 + 0.009676s + 5.203 \times 10^7} \tag{21}$$

The reduced model retains the important features of the original model as its bode plot resembles the original system. The bode plot for these two systems is shown in Fig. 4.

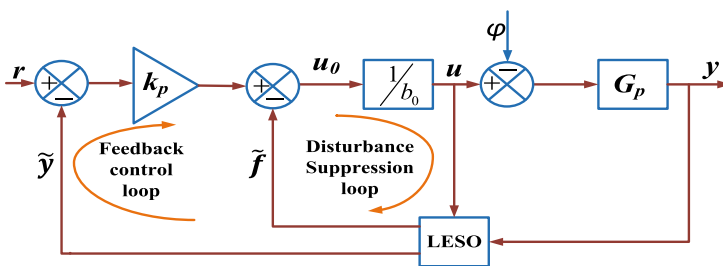


Fig. 4 Schematic block diagram of an ADRC controller

3 Design of ADRC Controller for SEPIC Converter

In the ADRC method, the disturbances whether known or unknown are initially clubbed together in a single variable and further they are estimated in a certain way with the help of an observer. The estimated disturbance is then suppressed with the action of the controller. The block diagram depicting the structure of the ADRC controller is shown by Fig. 4. b_o is defined as the gain of the system, u is considered as the control input and the system's output is denoted by y . Also, k_p is the controller's gain, \tilde{y} is the estimated output, \tilde{f} is the estimated value of disturbance and φ is the external disturbance. The SEPIC converter's reduced order transfer function model is characterized by Eq. (21). By taking the inverse Laplace transform of this transfer function, we get the dynamics of the SEPIC converter in the time domain in the form of a differential equation that relates the output with the input using the following relation:

$$\ddot{y} = (-a_0y - a_1\dot{y} + \delta b_o u - b_1\dot{u} + \delta b_o u + \delta(t)) + b_o u = f(y, \dot{y}, u, \dot{u}, w, \delta) + b_o u \quad (22)$$

where $f(y, \dot{y}, u, \dot{u}, w, \delta)$ has been considered as a generalized form of disturbance comprising of all the internal as well as the external disturbances. Since the system has been reduced to second order hence by the application of basic rules of ADRC method, a third-order observer is designed. The first two states in the observer are formed using the system dynamics while the third state in particular represents the cumulative disturbance. The estimated third state from the observer is canceled by the proper selection of the controller.

Let $x_0 = y$, $x_1 = \dot{y}$, $x_2 = h(\cdot)$, and $\dot{x}_2 = m$ where $h(\cdot)$ is differentiable and m is bounded. Using these assumptions, the state-space model for the system can be derived as:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Hm \\ y &= Cx(t) \end{aligned} \quad (23)$$

where

$$\dot{x}(t) = \begin{bmatrix} \dot{x}_0(t) \\ \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix}, A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ b_0 \\ 0 \end{bmatrix}, C = [1 \ 0 \ 0] \text{ and } H = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

The above state-space model can be used to design an ESO for the system given by:

$$\begin{aligned} \dot{z} &= Az + Bu + L(y - \hat{y}), \\ \hat{y} &= Cz \end{aligned} \quad (24)$$

where $z(t) = [z_0 \ z_1 \ z_2]^T$ is the estimated states and $L = [\beta_0 \ \beta_1 \ \beta_2]^T$ is the gain of the observer. A PD controller of the following form is considered:

$$u_0(t) = k_1(r - z_1) - k_2z_2 \quad (25)$$

The third state of the observer is rejected by the final control law given by:

$$u(t) = \frac{u_0(t) - z_3}{b_o} \quad (26)$$

where $k_p = [k_1 \ k_2 \ 1]$ is the controller's gain.

The previous works in [7, 12] have proposed some techniques for the selection of bandwidths of the observer (ω_0) and controller (ω_c). The characteristics equation of the controller is compared with the second-order equation tuned in the form of the controller's bandwidth (ω_c) that configures the controller's gain as:

$$\lambda(s) = s^2 + k_1s + k_2 = (s + \omega_c)^2 \quad (27)$$

From this, the controller's bandwidth is calculated as:

$$k_1 = 2\omega_c, k_2 = \omega_c^2$$

Similarly, the characteristics equation of the observer is compared with the third-order equation tuned in the form of the observer's bandwidth (ω_0).

$$\psi(s) = s^3 + \beta_0s^2 + \beta_1s + \beta_2 = (s + \omega_0)^3 \quad (28)$$

From this equation observer's bandwidth is calculated as:

$$\beta_0 = 3\omega_0, \beta_1 = 3\omega_0^2 \text{ and } \beta_2 = \omega_0^3.$$

The values of ω_c and ω_0 are selected as 900 rad/sec and 12600 rad/sec respectively based on the calculation of maximum sensitivity as 1.4.

4 Simulation Results

The circuit for the converter along with the controller was designed in the MATLAB/Simulink environment. The simulation results have been plotted for all three possible cases, i.e., for variation in set point voltage, variation in input voltage, and variation in load resistances. The performance parameters consisting of settling time, overshoot/undershoot, peak inductor current, and integral square error (IAE) has been calculated in each of these cases and listed in Table 2.

4.1 Servo Performance

In this case, the reference tracking capability of the controller is studied. Figures 5 and 6 depict the output voltages, the inductor currents, and the load currents obtained through simulation in boost and buck modes respectively. For the boost mode of operation, the input voltage is fixed to 30 V, while the load resistance is set to 100 Ω . At the time $t = 0.2$ s, the reference is step changed from 48 to 60 V, and again

Table 2 Performance parameters of the proposed method

Types of disturbances	T_s	Overshoot (V)	I_{L1} (A)	I_{L2} (A)	IAE
Servo performance (boost mode)	0.01	0	2.6	1.9	0.0402
Input voltage variation (boost-buck mode)	0.02	4	6	4	0.0287
Load variation (boost mode)	0.01	0.2	1.2	0.7	0.0003

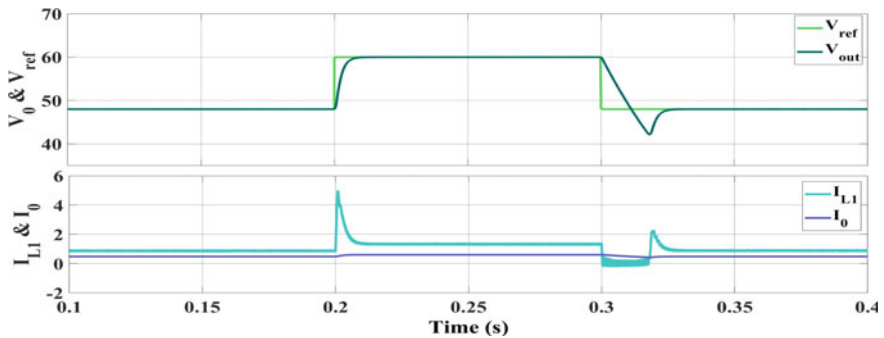


Fig. 5 Responses for change in reference in boost mode

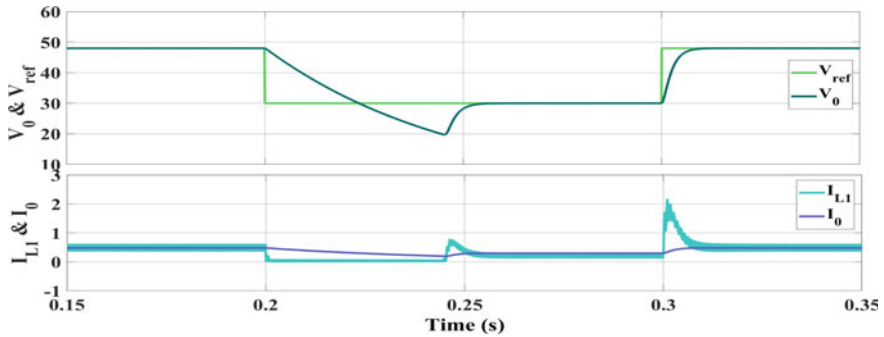


Fig. 6 Responses for change in reference in buck mode

back to 48 V at time $t = 0.3$ s. In the buck mode, the input voltage is set to 60 V, and at time $t = 0.2$ s, the reference voltage is decreased from 48 to 30 V and again to 48 V at time $t = 0.3$ s. The performance parameters have been listed in Table 2. From this, it is observed that the settling time is around 0.01 s when the reference is increased, while the curve settles at around 0.03 s when the reference is decreased.

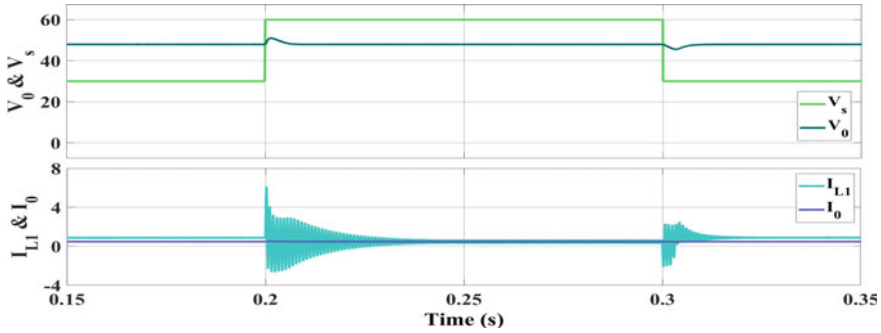


Fig. 7 Responses for change in input voltage

4.2 Regulatory Performance for Variation in Input Voltage

In this case, the effect of varying input was studied by keeping the load resistance and reference voltage fixed to $100\ \Omega$ and $48\ \text{V}$ respectively. Firstly, the input voltage was changed from 30 to $60\ \text{V}$ (boost-buck mode) at time $t = 0.2\ \text{s}$, and again from 60 to $30\ \text{V}$ (buck-boost mode) at time $t = 0.3\ \text{s}$, keeping the reference constant at $48\ \text{V}$. Figure 7 shows the simulation results under the assumed conditions. Also, the performance parameters are listed in Table 2. It is observed that the output voltage remains settled at $60\ \text{V}$ with very little overshoot/undershoot at the input voltage transition phase thereby proving a faster disturbance rejection capability.

4.3 Regulatory Performance for Variation in Load Resistance

Here the reference voltage was set to $48\ \text{V}$ for observing the effect of varying load. In the buck mode of operation, the input voltage was adjusted to $60\ \text{V}$ and the load resistance was varied from 100 to $50\ \Omega$ at time $t = 0.1\ \text{s}$ and then back to $100\ \Omega$ at time $t = 0.15\ \text{s}$. For analyzing the boost mode of operation, the input voltage was fixed to $30\ \text{V}$, and at time $t = 0.1\ \text{s}$, load resistance was varied from $100\ \Omega$ to $50\ \Omega$ and back to $100\ \Omega$ at time $t = 0.15\ \text{s}$. The corresponding responses are shown in Figs. 8 and 9, and the performance parameters are listed in Table 2. It can be analyzed that a small overshoot/undershoot of around $0.2\ \text{V}$ is present. Also, the settling time of around $0.005\ \text{s}$ was seen. The load voltage remains settled at the reference voltage, while the load current and primary inductor current changed instantly due to the change in the load as the power demand increased/decreased with increasing/decreasing load resistance.

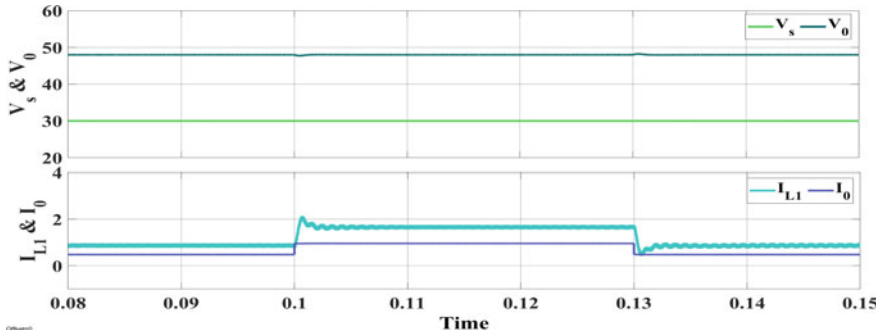


Fig. 8 Responses for change in load resistance in boost mode

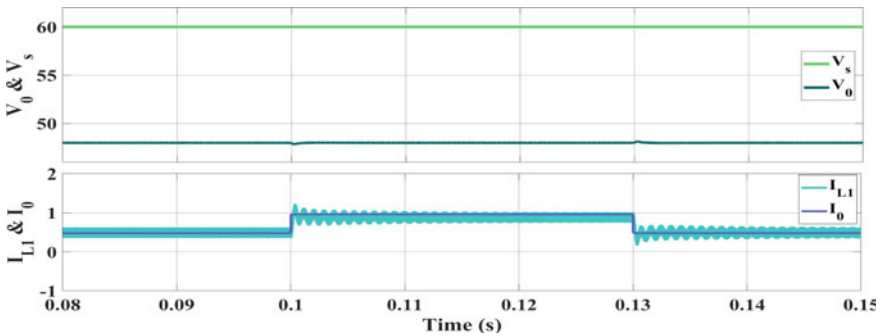


Fig. 9 Responses for change in load resistance in buck mode

5 Conclusion

In the present work, active disturbance rejection control (ADRC) is proposed for the SEPIC converter with high switching frequency. Higher switching frequency has allowed the selection of components with lower sizes. An ADRC-based control structure has been proposed for this converter which is derived based on the reduced order model of the SEPIC converter obtained through a balanced reduction technique. The suggested ADRC method results in outstanding performance and satisfactory robustness toward varying input voltages and loading conditions. This SEPIC converter along with suggested ADRC control may be used to achieve constant DC voltage from renewable energy sources like solar PV panels and windmill in varying environmental conditions.

References

1. Ahmad S, Ali A (2019) Active disturbance rejection control of dc-dc boost converter: a review with modifications for improved performance. *IET Power Electron* 12(8):2095–2107
2. Chennoufi K, Ferfra M, Mokhlis M (2021) Design and implementation of efficient mppt controllers based on sdm and ddm using backstepping control and sepic converter. In: 2021 9th international renewable and sustainable energy conference (IRSEC). IEEE, pp 1–8
3. Chiang S, Shieh HJ, Chen MC (2008) Modeling and control of pv charger system with sepic converter. *IEEE Trans Ind Electron* 56(11):4344–4353
4. Chu Z, Wu C, Sepehri N (2019) Active disturbance rejection control applied to high-order systems with parametric uncertainties. *Int J Cont Automat Syst* 17(6):1483–1493
5. Dahale S, Das A, Pindoriya NM, Rajendran S (2017) An overview of dc-dc converter topologies and controls in dc microgrid. In: 2017 7th international conference on power systems (ICPS). IEEE, pp 410–415
6. El Khateb A, Abd Rahim N, Selvaraj J, Uddin MN (2014) Fuzzy-logic-controller-based sepic converter for maximum power point tracking. *IEEE Trans Ind Appl* 50(4):2349–2358
7. Fu C, Tan W (2021) Analysis and tuning of reduced-order active disturbance rejection control. *J Frankl Inst* 358(1):339–362
8. Gao Z et al (2003) Scaling and bandwidth-parameterization based controller tuning. In: ACC, pp 4989–4996
9. Gireesh G, Seema P (2015) High frequency sepic converter with pwm integral sliding mode control. In: 2015 International conference on technological advancements in power and energy (TAP Energy). IEEE, pp 393–397
10. Gopi RR, Sreejith S (2018) Converter topologies in photovoltaic applications-a review. *Renew Sust Energy Rev* 94:1–14
11. Guan Y, Wang Y, Wang W, Xu D (2017) A high-frequency clcl converter based on leakage inductance and variable width winding planar magnetics. *IEEE Trans Ind Electron* 65(1):280–290
12. He T, Wu Z, Li D, Wang J (2019) A tuning method of active disturbance rejection control for a class of high-order processes. *IEEE Trans Ind Electron* 67(4):3191–3201
13. Herbst G (2013) A simulative study on active disturbance rejection control (adrc) as a control tool for practitioners. *Electronics* 2(3):246–279
14. Hwu K, Peng T (2011) A novel buck-boost converter combining ky and buck converters. *IEEE Trans Power Electron* 27(5):2236–2241
15. Kobaku T, Patwardhan SC, Agarwal V (2017) Experimental evaluation of internal model control scheme on a dc-dc boost converter exhibiting nonminimum phase behavior. *IEEE Trans Power Electron* 32(11):8880–8891
16. Komurcugil H, Biricik S, Guler N (2019) Indirect sliding mode control for dc-dc sepic converters. *IEEE Trans Ind Inf* 16(6):4099–4108
17. Laub A, Heath M, Paige C, Ward R (1987) Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms. *IEEE Trans Autom Cont* 32(2):115–122
18. Luo FL, Ye H (2016) *Advanced dc/dc converters*. CRC Press
19. Moore B (1981) Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans Autom Cont* 26(1):17–32
20. Sel A, Güneş U, Kasnaoğlu C (2020) Design of output feedback sliding mode controller for sepic converter for robustness. *Int J Electron* 107(2):239–249
21. Smedley KM, Cuk S (1995) Dynamics of one-cycle controlled cuk converters. *IEEE Trans Power Electron* 10(6):634–639
22. Yao C, Ruan X, Wang X, Chi KT (2011) Isolated buck-boost dc/dc converters suitable for wide input-voltage range. *IEEE Trans Power Electron* 26(9):2599–2613
23. Zhao Q, Lee FC (2003) High-efficiency, high step-up dc-dc converters. *IEEE Trans Power Electron* 18(1):65–73
24. Zhou R, Fu C, Tan W (2020) Implementation of linear controllers via active disturbance rejection control structure. *IEEE Trans Ind Electron* 68(7):6217–6226

Designing a Mapless Navigation Mobile Robot Using Deep Q Learning



Chinmay Kolhatkar and Kranti Wagle

Abstract Mobile robots are becoming part of our daily life. Autonomous navigation is an important feature of any mobile robot. Techniques such as GPS tracking cannot be implemented for indoor navigation. Also earlier techniques such as SLAM and PRM need the predefined map and are affected by the dynamically changing position of landmarks, Visual SLAM fails if there are fewer visual features in the environment such as (all white walls) can cause problem of loop closure. SLAM-based navigation algorithm needs a predefined map; hence, it demands for accurate sensors for precise map building (Ibragimov IZ, Afanasyev IM in Comparison of ROS-based Visual SLAM methods in homogeneous indoor environment, IEEE, 2017). To overcome these challenges, the approach of reinforcement learning can be used (Appl Sci 10:8386, 2020). In reinforcement learning, the robot acts as an agent and performs navigation based on the received rewards. Here, our goal is to implement a DQN-based navigation algorithm on real 2WD robot for performing a mapless navigation. The sensor data such as data from LIDAR, IMU and motor encoders is fused together to perform the localization and navigation using the DQN algorithm. The results of navigation are compared against the benchmark SLAM algorithms like GMapping, Hector SLAM and Cartographer.

Keywords SLAM · DQN · LIDAR · Reinforcement learning · Gazebo · ROS

1 Introduction

SLAM has disadvantage that it needs highly accurate sensor and it cannot work in mapless environment [4]. A predefined map is generated based on either the obstacles present in the surrounding area or based on the features abstracted from the environment. SLAM techniques need predefined map to be generated using techniques such as Hector SLAM, GMapping, DPPTAM, ORB-SLAM. All these techniques require the predefined generation of map before performing navigation. SLAM encounters

C. Kolhatkar (✉) · K. Wagle

Fr. Conceicao Rodrigues College of Engineering Bandra, Mumbai, India

e-mail: chinmayhere84@gmail.com

multiple problems such as loop closure, demanding highly accurate sensors, feature full environment and many more [5]. In order to tackle these issues and in order to perform mapless navigation with highest possible accuracy, machine learning algorithms can be implemented. So, the idea here is to navigate the robot in an environment without the help of predefined or pregenerated map of the environment. So, as we are dealing with the situation where we do not have a fixed set of data or we cannot collect the data posterior, we use a technique called reinforcement learning [2]. Reinforcement learning is a biologically inspired approach which can be utilized to navigate the robot in an unknown environment same as humans do. The basic analogy behind the project is that, as soldiers are trained in a simulated but rough terrain environment of army base camps but once they pass all the huddles during their training, based on their experiences, they can be able to survive in any similar situation anywhere in the world. Same analogy is applied here for robots. Reinforcement learning has four parameters agent (robot), environment, reward and action. Based on the action performed by an agent, the environment will make an observation; based on that observation, the environment will offer the reward. The reward can either be a negative reward or it can be a positive reward. Based on the inputs obtained during each episode, the agent will create an optimal strategy. The robot is acting as an agent or soldier who is getting trained by performing multiple actions in an environment and learning based on the experiences collected during the training period. The reward strategies are developed so that the correct action will be repeated again and again, and punishment is given for wrong actions so that wrong actions can be avoided. The reward system is called as a policy of the environment or policy for the reward. The policy is designed such that it will optimally converge and end up taking accurate actions [13]. The robot is trained for multiple episodes in which it learns from its experiences collected as state action and reward pair. The policy is designed such that the agent, i.e., robot will explore the environment and at a same time exploit the environment with learning curve over the time period. We are implementing value-based off-policy learning strategy for mobile robot navigation. The neural network is designed with robots pose/states as inputs and Q values as output of the network. The network has multiple hidden layers and backpropagation layers which train the network. A CUDA-based NVidia GPU kit is used on a mobile robot platform to run the autonomous navigation algorithm.

2 Methodology

The proposed model is implemented in two parts. The first part consists of designing of hardware mobile robotic base platform and the second part is designing of software module to run the mobile robot and implement the deep reinforcement algorithm for autonomous navigation. Extensive literature survey about mobile robot platform has been done, in which we observed that most of researchers choose to build their own mobile robot platform instead of readymade platforms like TurtleBot, Kobuki, Husky and Rumba as they are quite costly and not affordable for students or individual

researchers. Hence, here we are designing our own mobile robot base as good as readymade one. So, first, let us see the basic block diagram of our mobile robotic platform (Figs. 1 and 2).

The hardware platform uses Cuda development kit, a microcontroller, LIDAR LDS, motors with encoders and an IMU sensor. The RP-LIDAR is basically used to create a laser-scanned point cloud of environment which will provide distance from an obstacle. The LDS has about 12 m scanning range which is enough for our experimentation purpose. For successful implementation of mapless navigation algorithm, robot has to know its own position in an environment. Hence, we have used DC motors with encoders which will help us determine the exact pose of the robot. Along with pose of the robot, we also want to know the direction, tilt and acceleration of robot; hence, we have used MPU6050 gyroscopic IMU sensor. This complete odometry data is fed into ROS navigation stack for further processing. The ROS core is running on NVidia Jetson Nano Development Board, and for performing other control actions, an Arduino Mega board is used as slave controller. In Arduino environment, LinoBase and ROS libraries are used to integrate the functionality of the robot. The ROS core communicates over topics. Hence, the communication between Arduino and Jetson Nano is done over USB/UART Protocol. Arduino is basically responsible to handle the encoder and IMU sensor, and we have used bidirectional

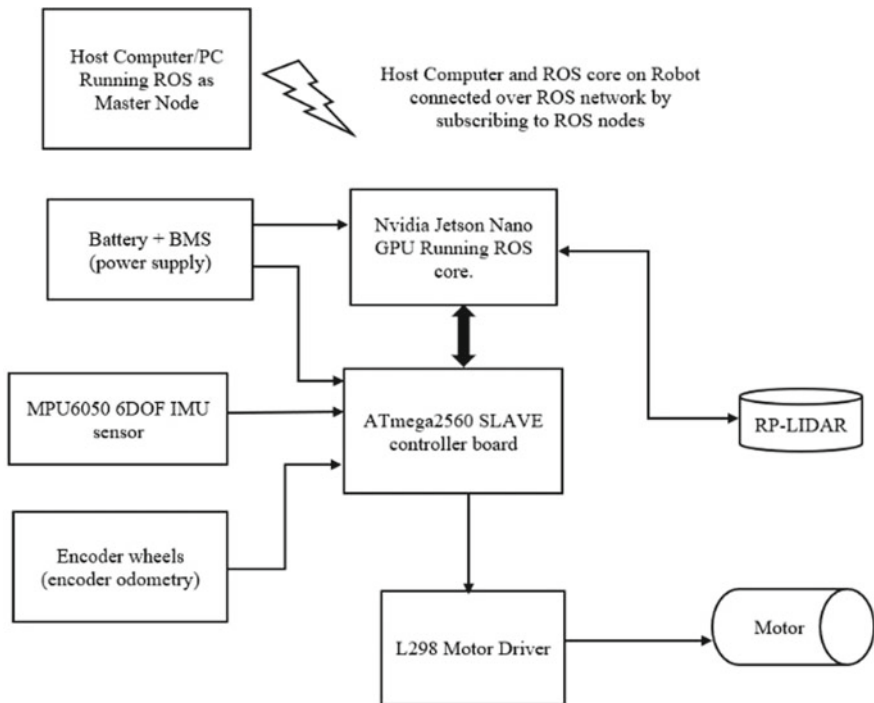
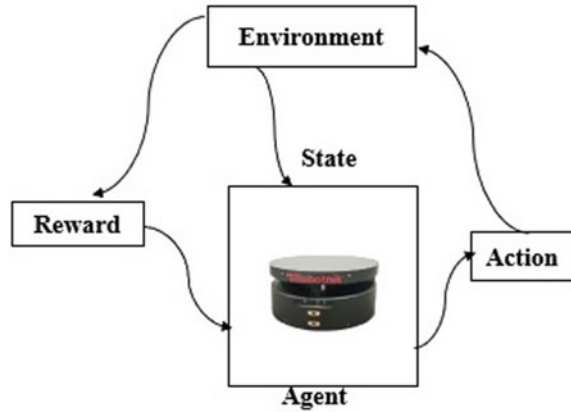


Fig. 1 Block diagram of robot base

Fig. 2 Reinforcement learning environment



Hall effect-type encoder motors which give information about the distance as well as direction of rotation. Encoders are attached to the interrupt pin of the ATmega2560. IMU sensor MPU6050 is connected to the I2C pins of Arduino. Arduino Mega also runs a PID sketch to apply accurate moves and keep wheel steady in same direction. Also, PWM signal is used to keep the control on robot's speed. Apart from these two control boards, we have used a LIDAR sensor, to create a point cloud map of environment. LIDAR is a 360° 2D laser scanning device. We have used RP-LIDAR A1 series from ROBOpeak. This LIDAR gives serial data which is interfaced with the Jetson Nano board via USB-TTL converter device. We have used this LDS to detect the distance from obstacles. Along with this, the odometry data used by the ROS navigation stack is taken from the motor encoders. We have used 190 RPM motor with 950 CPR encoder sensors. Encoder sensors are connected to the interrupt pins of microcontroller; each tick of pulse from encoder is converted to distance value using this formula:

$$\text{Distance travelled} = (\text{Encoder ticks}/360) * \text{circumference}.$$

In order to drive motor, motor driver L298N H-Bridge 2 A motor driver is used. Motor driver receives the control signal and PWM signal from Arduino Mega. Along with motor driver for powering the whole mobile robot, a portable 10,000 mAh power bank and 11.1 V Li-ion rechargeable battery are used, and to control the supply to whole robot and all sensors, battery management system (BMS) is used.

The further implementation includes the software written in Arduino IDE for communication with ROS navigation stack to achieve the autonomous navigation of mobile robot. Here, we have implemented deep Q network (DQN) algorithm for autonomous navigation of mobile robot. Q learning is an off-policy value-based reinforcement learning algorithm which is used to find the best action that can be taken given the current state. It is off-policy algorithm because it chose the random action from the action set and not from the designed fix policy. This randomization

is also helpful in our use case of robotics applications. Q learning has two important equations: one is Bellman’s equation and other is Markov’s decision process. The Bellman equation is basically used as an optimality equation where the optimal q value is found for each state called q_{max} or Q optimum. Based on this, the agent learns and generates the policy represented by π . This policy is such that it will converge optimally spitting out the optimal Q value which will maximize the reward in a given state provided action is taken. In order to make agent explore as well as exploit the environment, Markov’s decision process (MDP) is used and MDP is a 5 tuple $(S, A, p_a(s' | s, a), R(r, s))$ where s is finite set of states, A is finite set of actions, $p_a(s' | s, a)$ is the transition probability to define the chances that next states will be picked as a consequence of action a . $R(r, s)$ is the expected reward received due to action ‘ a ’ while in state s' . r is a signed value used for reward or punishment. From Bellman’s equation, γ is a discount factor $\gamma [0, 1]$. And the learning rate is given by $\alpha [0, 1]$ in our case as we are dealing with the discrete values. For testing of Q learning algorithm, we are using Gazebo environment. In Gazebo, we are simulating the algorithm on turtlebot3 robot. The turtlebot3 is a differential drive mobile robot with sensors like laser distance sensor (LDS), inertial and measurement unit (IMU) and motor encoders. The state of the robot in the Gazebo environment is determined by these sensors. Turtlbot3 performs five discrete actions shown in Fig. 3. We also determine the state of the robot as an input to our neural network (Figs. 4 and 5).

Turtlebot3 has total 26 state values, 24 values of laser distance (LDS) plus the distance of robot from the goal and the angle to the goal, where we are taking 24 LDS reading from LIDAR sensor per rotation and one reading about the distance of robot from the goal and one reading of angle from the goal using IMU sensor, which will total our state value = 26. We are feeding these state values to our neural network, and we expect the action at the output of the neural network. The architecture of neural network is shown below. Here, we are trying to implement the deep Q learning algorithm. In order to implement the algorithm, we are making use of Python programming language and Python 2.7 is used for development of the

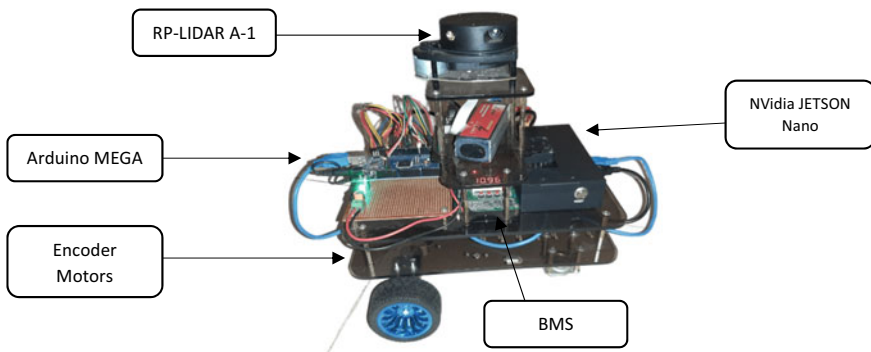


Fig. 3 2WD robotic platform designed for experimentation

Fig. 4 Robots action set
(linear velocity action set)

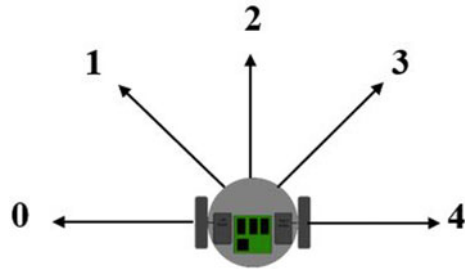
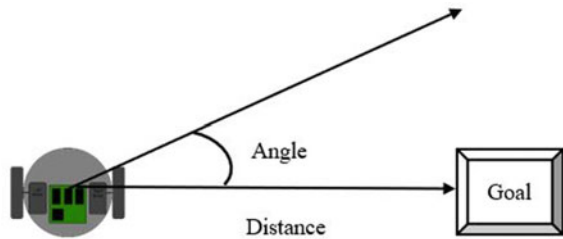


Fig. 5 Robots state in the environment from LDS and odometry



whole project. We are also using the Anaconda 5.2, TensorFlow and Keras packages for implementation of machine learning and neural network (Fig. 6).

Implementation of algorithm: Here is the standard Q learning algorithm, which we are trying to implement in Gazebo-simulated environment and our own designed mobile robot base.

Algorithm:

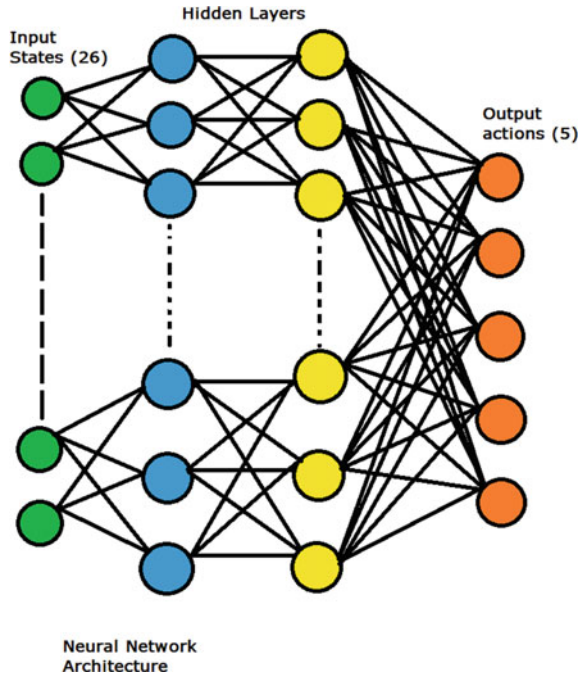
```

Start with  $Q_0(s, a)$  for all  $s, a$ .
Get initial state  $s$ 
For  $k = 1, 2, \dots$  till convergence
    Sample action  $a$ , get next state  $s'$ 
    If  $s'$  is terminal:
        target =  $R(s, a, s')$ .
        Sample new initial state  $s'$ 
    else:
        target =  $R(s, a, s') + \gamma \max Q_k(s', a')$ .
     $Q_{k+1}(s, a) \leftarrow (1 - \alpha)Q_k(s, a) + \alpha$  [target].
     $s \leftarrow s'$ 
    
```

The steps followed in implementation of this algorithm are as follows:

- (a) Initialize the replay memory capacity, i.e., represented by N .
- (b) Initialize network with random weights.

Fig. 6 Neural network architecture for DQN algorithm



- (c) For each episode:
 - (a) Initialize the starting state.
 - (b) For each time step:
 - (i) Select an action.
 - 1. Via exploration or exploitation.
 - (ii) Execute the selected action.
 - (iii) Observe the reward and next state.
 - (iv) Store the experience in replay memory.
 - (v) Sample the random batch from replay memory.
 - (vi) Preprocess states from each batch.
 - (vii) Pass the batch of preprocessed states to policy network.
 - (viii) Calculate loss between output Q value and Actual Q value.
 - (ix) Second pass to the network for the next state.
 - (x) Gradient decent updates weights in the policy network to minimize loss.

The idea here is to use the neural network instead of value iteration, in order to process the data at faster rate. The goal here is to minimize the loss and find out the optimal Q-function. The optimal Q-function, of course, should satisfy the Bellman equation. The optimal Q-function is represented by the Bellman equation given as:

$$q_*(s, a) = \max q_\pi(s, a),$$

where π represents the policy. In the algorithm, the iteration is performed for multiple episodes. One episode is when robot moves into environment optimizing the Q value, for given state and action pair. The episode ends when it either lands on a goal, crashes to the wall or timeout happens. During initial stage, the robot does not have idea from where to start the navigation. So from its current position, it randomly chooses the action from the action set, and once it starts moving into the environment, it iteratively updates the Q values for each state/action pair. The Q values are stored inside the q table. This Q table is nothing but our replay memory. The robot will each time look into the batch of preprocess states, and based on the highest Q value in the table, it performs the action resulting into better reward. For training purpose, the limited number of episodes are chosen. Let us say number of episodes are 700. During these 700 episodes, the model is trained for multiple Q values and state/action pairs. Now as robot has no predefined information about the environment or it's starting position, it is hard to determine the first action to be taken at the start of the episode. This is usually done by some kind of strategy. Here, we are making use of epsilon greedy strategy represented by ϵ . Robot uses this strategy to keep balance between the exploration and exploitation. The exploration is when robot explores the environment irrespective of reward it is gaining in the next state, whereas the exploitation means the robot performs only those action which will make it gain more reward. The epsilon greedy strategy helps robot to keep balance between the exploitation and exploration. Now let us discuss some implementation details of the algorithm. The reward policy for turtlebot3 is as follows.

<p>if $-\pi/2 < \theta < \pi/2$ then $R_\theta \geq 0$</p> <p>else $R_\theta < 0$</p> <p>if $D_c < D_g$ then $R_d > 2$</p> <p>else $1 < R_d \leq 2$</p> <p>finally $R = R_\theta * R_d$</p>	<p>where the θ is the angle from turtlebot3 and goal</p> <p>the D_c is current distance of robot from goal</p> <p>D_g is absolute distance from the goal</p> <p>R_θ is a reward obtained from θ</p> <p>R_d is as reward obtained from distance</p> <p>R is the total reward.</p>
--	--

The angle θ is calculated by

$$\theta = \pi/2 + \text{action} * \pi/8 + \emptyset \text{ where } \emptyset \text{ is Yaw of turtlebot3.}$$

The reward from the θ is calculated by,

$$R_\theta = 5 * (1 - \theta) \theta \text{ is angle from goal.}$$

The reward from the distance is calculated by,

$$R_d = 2^{(D_c/D_g)}$$

Whenever the robot reaches its goal, it gets the reward + 100, whereas when robot crashes into an obstacle, it gets the negative reward of - 100.

Currently in order to test the above algorithm, we are making use of Gazebo simulation environment in ROS. The algorithm is tested inside two different environments. First is environment with static obstacles and another one is with static as well as moving obstacles. The environment is created using the predefined materials and models available in Gazebo program. The robot is trained inside simulated environment for multiple episodes, and the weights/epsilon values in our case are stored in form of models. The model is stored as .json file which is then be ported to actual robot for navigating into similar environment. Now let us have a look at some hyperparameters for our neural network. The hyperparameters are those parameters which are always fixed or which cannot be evaluated during the process. The hyperparameters for our neural network are as follows, mentioned in Table 1 (Figs. 7 and 8).

Table 1 Hyperparameters of neural network

Hyperparameter	Value	Description
Episode_step	6000	One episode time
Target_update	2000	Update rate of the target network
Discount factor	0.99	Represents how much future events lose their value according to how far away
Learning rate	0.00025	Learning speed. If the value is too large, learning does not work
Epsilon	1	Initially kept 1 then decay
Epsilon decay	0.99	Reduction rate of epsilon
Epsilon minimum	0.05	Minimum epsilon value
Bathc size	64	Size of group of training samples
Train start	65	Start the training if replay memory size is greater than 64
Memory	1,000,000	The size of replay memory

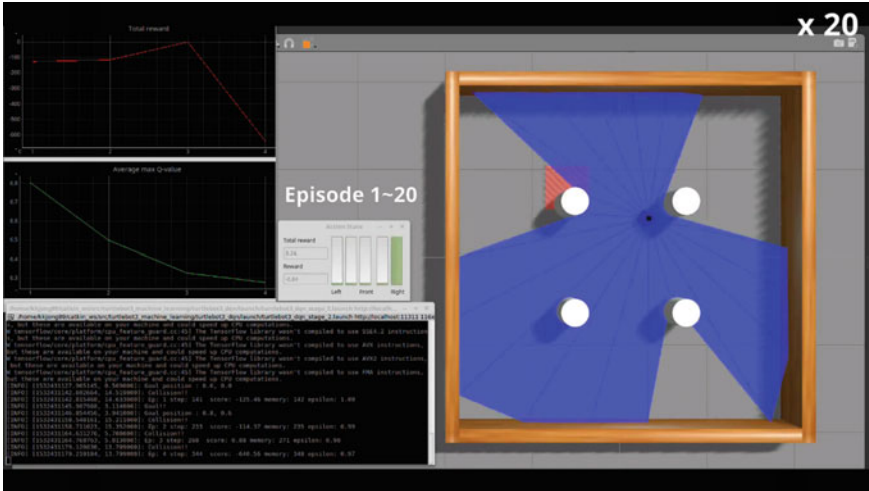


Fig. 7 DQN algorithm implementation and testing in simulation (Gazebo) and plotting the graph of average Max-Q value versus the total obtained reward for environment with static walls and moving objects

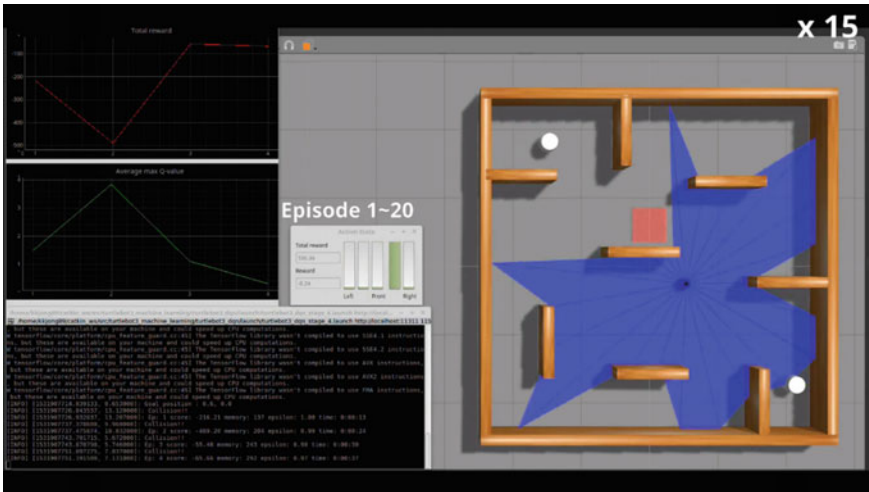


Fig. 8 DQN algorithm implementation and testing in simulation (Gazebo) and plotting the graph of average Max-Q value versus the total obtained reward for environment with moving objects/obstacles

3 Conclusion

The SLAM-based navigation approach is formerly a chicken-egg problem, where initiating the mapping/localization becomes difficulty; due to this interlock condition, pregeneration of map becomes mandatory. The problem of navigation of a mobile robot in an unknown and random environment is solved more precisely using Q learning algorithm rather than conventional SLAM-based approach. Dependencies on sensory network can be reduced up to some extent by using the reinforcement learning approach. As number of training episodes increases and more data is collected from the environment, navigation becomes more smooth and reliable. Instead of using expensive 3D camera-based approach, even a 2D LIDAR-based navigation algorithm performs well with reinforcement learning approach. Sensor fusion techniques such as integrating the encoder odometry and IMU sensor data with LIDAR data show better results in navigation. Designing customized robotic platform gives more freedom for implementing different strategies for navigation and also provides freedom to adjust the computational capabilities. The epsilon greedy strategy algorithm is not always the best fit for our navigation problem, because after a while, it makes robot to exploit more than exploration. DQN not only helps in navigation but also in obstacle avoidance in an unknown and randomly changing environment; even moving obstacles can be avoided and ultimately policy of navigation converges optimally. During literature survey, we observed that SLAM algorithms like DPP-SLAM, RTAB-MAP, GMapping, etc. need a predefined map and accurate sensor input every time while localizing as well as mapping. But in the case of reinforcement learning, a fully trained agent can navigate even with lesser accuracy of sensors. We observed that during testing of robot in live environment, one of the encoder motors was moving a bit out of control due to some shaft-related issue but still the robot managed to perform well compared to the SLAM algorithm.

References

1. Ibragimov IZ, Afanasyev IM (2017) Comparison of ROS-based visual SLAM methods in homogeneous indoor environment. In: 14th WPNC. IEEE
2. A deep reinforcement learning approach for active SLAM. *Appl Sci* 10:8386 (2020). <https://doi.org/10.3390/app10238386>
3. Autonomous navigation of a mobile robot in dynamic indoor environments using SLAM and reinforcement learning. In: 2nd international conference on advances in mechanical engineering (ICAME 2018) IOP publishing conference series: materials science and engineering
4. Li Y, Shi C (2018) Localization and navigation for indoor mobile robot based on ROS. IEEE
5. Gatesichapakorn S, Takamatsu J (2019) ROS based autonomous mobile robot navigation using 2D LiDAR and RGB-D camera. IEEE
6. Syaqr WA, Kamarudin K (2018) Mobile robot based simultaneous localization and mapping in unimap's unknown environment. IEEE
7. Ocando MG, Certad N (2017) Autonomous 2D SLAM and 3D mapping of an environment using a single 2D LIDAR and ROS. IEEE
8. Cheng Y, Wang GY (2018) Mobile robot navigation based on Lidar. IEEE

9. Madhira K, Patel J (2017) A quantitative study of mapping and localization algorithm on ROS based differential robot. In: NUiCONE. IEEE
10. da Silva BMF, Xavier RS (2017) Experimental evaluation of ROS compatible SLAM algorithm for RGB-D sensors. IEEE
11. Filipenko M, Afanasyev I (2018) Comparison of various SLAM systems for mobile robot in an indoor environment. IEEE
12. Yagfarov R, Ivanou M (2018) Map comparison of LIDAR based 2D SLAM algorithms using precise ground truth. In: ICARCV. IEEE
13. Chewu CCE, Manoj Kumar V (2018) Autonomous navigation of a mobile robot in dynamic indoor environments using SLAM and reinforcement learning. In: 2nd International conference on advances in mechanical engineering (ICAME 2018)
14. Megalingam RK, Teja CR (2018) ROS based autonomous indoor navigations imulation using slam algorithm. Int J Pure Appl Math 118(7):199–205
15. Hwang S-Y, Song J-B (2013) Clustering and probabilistic matching of arbitrarily shaped ceiling features for monocular vision-based SLAM. Adv Robot 27(10):739–747
16. Khan MU (2019) Balkan J Electr Comput Eng 7(3)
17. Xin J, Zhao H, Liu D (2017) Application of deep reinforcement learning in mobile robot path planning. IEEE
18. Zamora I, Lopez NG, Vilches VM (2017) Extending the OpenAI Gym for robotics: a toolkit for reinforcement learning using ROS and Gazebo
19. Tai L, Liu M (2016) A robot exploration strategy based on Q-learning network. IEEE

Implications of Location of Strain Gauges and Excitation Voltage Over the Metrological Performance of Trapezoidal-Shaped Force Transducer



Richa Saxena and Harish Kumar

Abstract The present study discusses the effect of location of strain gauges and excitation voltage of the digital indicator over the metrological characterization of the force transducer. A trapezoidal-shaped force transducer of nominal capacity 50 kN has been investigated in which two locations for mounting the strain gauges have been identified based on FEA. The evaluation of metrological characteristics has been based on ISO 376: 2011. The relative deviations due to repeatability, reproducibility, interpolation, and expanded uncertainty of measurement has been reported in this work. A comparative analysis has been performed between the observations recorded at 5 V and 10 V excitation voltages. Results show that the effect of increased excitation voltage of the digital indicator is insignificant over the uncertainty of measurement of the force transducer. The location of strain gauges affects the sensitivity of the transducer, significantly.

Keywords Excitation voltage · Force measurement · Metrological characteristics · Uncertainty of measurement

1 Introduction

A strain-gauged trapezoidal-shaped force transducer has been investigated in this study for force measurement in applications like onsite calibration of large testing machines to test the value of externally applied loadings, identification of materials' strength, verification of uniaxial testing machines, calibration of hardness blocks, etc. [1, 2]. A Wheatstone bridge configuration is used for measurement of strain produced, upon applying an external unknown force. The gauge factor represents the sensitivity of the strain gauge.

R. Saxena (✉)

Department of Mechanical Engineering, Government Polytechnic Kotana, Baghpat, UP, India
e-mail: richa.pn@gmail.com

H. Kumar

Department of Mechanical Engineering, National Institute of Technology Delhi, Delhi, India

Mounting of strain gauges at the maximum stress/strain positions is desirable for achieving the maximum sensitivity of SGFT [3]. In this investigation, an attempt has been made to identify the performance of SGFT by mounting the strain gauges at two different locations. Excitation voltage is another parameter which could affect the transducer's sensitivity [4]. A comparison of metrological characterization of EN 8-based trapezoidal-shaped SGFT at 5 and 10 V excitation voltage is being discussed here.

2 Fabrication

The trapezoid geometry is machined from a EN 8 steel specimen using vertical milling machine. The dimensions are considered as follows: 180 mm outer length, 160 mm inner length, 35 mm width, and 30 mm end boss diameter [5]. The machined component is annealed at 800 °C for relieving the internal stresses. Annealing causes softening of the component. So, hardening is performed for increasing the hardness by oil quenching [6]. Surface finishing operation is performed for mounting strain gauges for force measurement [7, 8].

2.1 Mounting of Strain Gauges

The maximum stress–strain locations are selected for mounting strain gauges. Two positions are being identified in this case for examining the location-effect upon the metrological performance of the transducer [3, 9]. Strain gauges 1 and 2 are mounted orthogonally on the outer surface of the trapezoid geometry. Strain gauges 3 and 4 are mounted on the inner surface, orthogonally. These four strain gauges form a full Wheatstone bridge at Position A. Strain gauges 5, 6, 7, and 8 form another full Wheatstone bridge at Position B, as shown in Figs. 1 and 2.

3 Evaluation of Uncertainty of Measurement

The measurement result has to be stated as a magnitude value of force along with its uncertainty, including the SI unit of force. The measured force value must be linked to a reference through a recognized continual traceability chain. An essential tool in ensuring the traceability of measurement is the calibration of the force measuring instrument. Calibration establishes the performance characteristics of a transducer before its actual use [10]. Calibration in accordance to ISO 376: 2011 [11] is followed for evaluating the metrological performance of the developed force proving instrument and transducer. It is a crucial step in determining the performance characteristics of the force measurement device. The concept of uncertainty was introduced in the

Fig. 1 A schematic diagram showing arrangement of four strain gauges at Position A

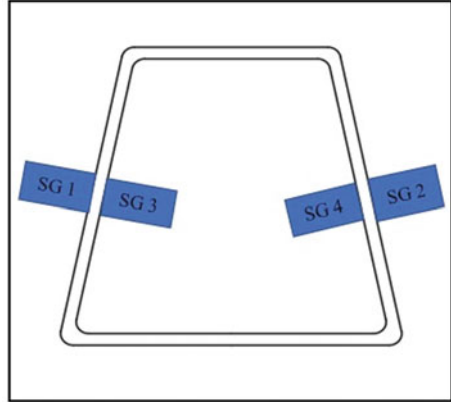
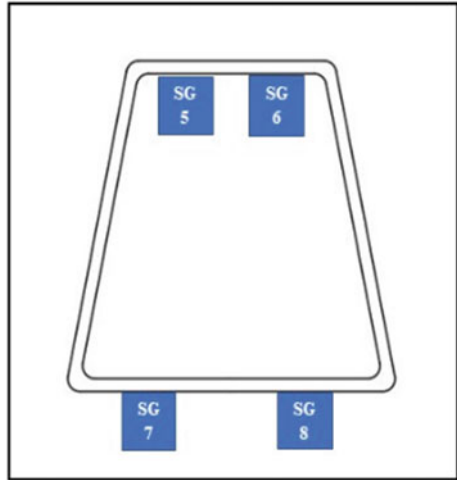


Fig. 2 A schematic diagram showing arrangement of four strain gauges at Position B



late 1980s. It is a statistical expression that displays the dispersion of values from the true value of the measurand [12, 13]. Following are the steps involved in evaluating UoM [14]:

- (a) Specification of measurand
- (b) Mathematical model of measurand as per ISO 376: 2011
- (c) Identification of sources of uncertainty
- (d) Evaluation of input quantities
- (e) Determine standard UoM of each component
- (f) Calculate combined standard UoM
- (g) Calculate expanded UoM
- (h) Result analysis.

SGFT has been calibrated in compression and tension for the nominal load capacity of 15 kN. The transducer has been subjected to 12% overload to the nominal capacity for 90 s four times. The overload test has been found satisfactory. 50 kN deadweight force machine has been used to calibrate the transducer following the calibration steps. The observations are recorded using a high-resolution digital indicator (10–5 mV/V) [5]. The resolution of an instrument plays a crucial role in displaying the output nearest to the true value. Higher the resolution, closest would be the output to the true value. The temperature compensation mechanism has not been included as the calibration is to be done in a controlled environment, as specified in ISO 376:2011 metrology standard. Analysis of the observations has been conducted for evaluating the UoM.

Excitation voltage of 5 V is given as an input to power the bridge initially. Three sets of observations have been recorded for a nominal tensile force of 15 kN

- Voltage output at Position A (5 V, excitation voltage)
- Voltage output at Position B (5 V, excitation voltage)
- Voltage output at Position A (10 V, excitation voltage).

3.1 Metrological Characterization of EN 8 SGFT, Position A

Initially, 5 V is given to the bridge connected at Position A, and the observations are recorded for further evaluation. The relative uncertainty contribution of individual parameters is summarized in Table 1 [15, 16].

Table 1 UoM of EN 8 trapezoidal-shaped SGFT (tension) with contributing factors (Position A, 5 V)

Force (kN)	Mean value of observations (mV/V)	Relative contribution to the combined standard UoM (%)						Expanded UoM (%) $k = 2$
		zer (2a)	res (2a)	rpr (2a)	rep (2a)	int (2a)	cmc (2a)	
1	0.03950	0.019	0.025	0.228	0.152	0.169	0.015	0.178
3	0.11822	0.019	0.008	0.102	0.118	0.044	0.015	0.092
5	0.19686	0.019	0.005	0.168	0.036	0.036	0.015	0.083
7	0.27552	0.019	0.004	0.152	0.025	0.030	0.015	0.075
9	0.35412	0.019	0.003	0.116	0.008	0.006	0.015	0.058
11	0.43269	0.019	0.002	0.090	0.039	0.024	0.015	0.057
12	0.47204	0.019	0.002	0.085	0.023	0.022	0.015	0.052
13	0.51148	0.019	0.002	0.082	0.033	0.005	0.015	0.051
14	0.55087	0.019	0.002	0.044	0.005	0.001	0.015	0.037
15	0.59034	0.019	0.002	0.068	0.039	0.018	0.015	0.050

Table 2 UoM of EN 8 trapezoidal-shaped SGFT (tension) with contributing factors (Position B, 5 V)

Force (kN)	Mean value of observations (mV/V)	Relative contribution to the combined standard UoM (%)						Expanded UoM (%) $k = 2$
		zer (2a)	res (2a)	rpr (2a)	rep (2a)	int (2a)	cmc (2a)	
1	0.00630	0.032	0.159	0.317	0.159	0.369	0.015	0.308
3	0.01887	0.032	0.053	0.212	0.106	0.168	0.015	0.164
5	0.03137	0.032	0.032	0.159	0.096	0.015	0.015	0.094
7	0.04387	0.032	0.023	0.137	0.091	0.037	0.015	0.089
9	0.05646	0.032	0.018	0.124	0.089	0.067	0.015	0.094
11	0.06890	0.032	0.015	0.131	0.087	0.079	0.015	0.099
12	0.07518	0.032	0.013	0.133	0.093	0.051	0.015	0.093
13	0.08154	0.032	0.012	0.135	0.098	0.057	0.015	0.096
14	0.08777	0.032	0.011	0.125	0.091	0.006	0.015	0.083
15	0.09404	0.032	0.011	0.138	0.096	0.000	0.015	0.087

UoM is found upto 0.10% within 20–100% of the transducer’s working range. The relative uncertainty of the contributing factors is also within the permissible limits.

3.2 Metrological Characterization of EN 8 SGFT, Position B

The metrological performance evaluation of EN 8 SGFT at Position B would reveal the effect of the changed location over the metrological performance of the transducer. A summarized analysis of UoM of EN 8 SGFT at Position B has been given in Table 2.

Based on the mean values of observations at various force steps, a comparative graphical analysis is being presented in Fig. 3. It is observed that the output values are too less at Position B for the same applied force, affecting the transducer’s sensitivity.

A comparative graphical analysis of UoM of EN 8 SGFT at Position A and Position B is shown in Fig. 4. This figure depicts that the higher UoM is observed at Position B than at Position A. However, the decreasing pattern of the UoM values is identical as of Position A.

3.3 Effect of Excitation Voltage Over Metrological Characteristics of EN 8 SGFT, Position A

Excitation voltage is given as an input to the Wheatstone bridge that powers the bridge. The observations of the SGFT’s output are recorded at 5 and 10 V excitation

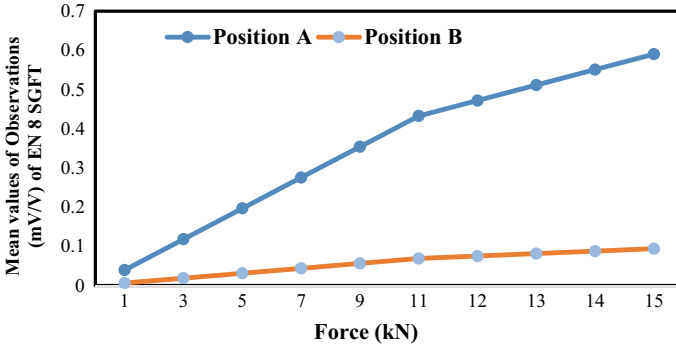


Fig. 3 A comparative graphical analysis between the mean values of observations (mV/V) obtained at Position A and Position B

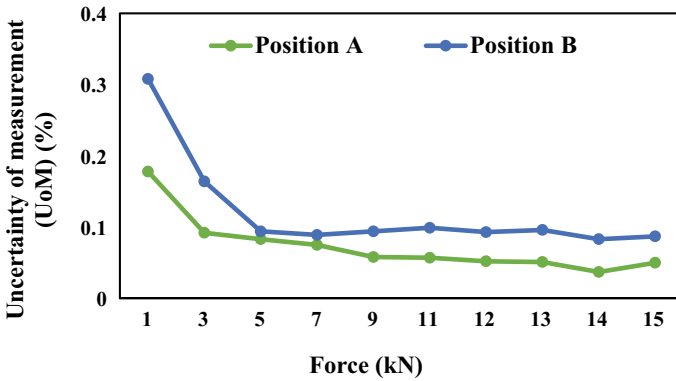


Fig. 4 Plot showing UoM ($k = 2$) of EN 8 SGFT at two positions of strain gauge arrangements (Position A and Position B)

voltage at Position A. The effect of excitation voltage over the metrology of EN 8 SGFT has been investigated based on metrological characteristics evaluation of the SGFT's output at 10 V excitation voltage, Table 3 [4].

It has been observed that an increase in excitation voltage does not have much impact on UoM. But it has a minor effect on the mean observations. The deviation between the UoM obtained from the two excitation voltages is shown in Fig. 5.

Table 3 UoM of EN 8 trapezoidal-shaped SGFT (tension) with contributing factors (Position A, 10 V)

Force (kN)	Mean value of observations (mV/V)	Relative contribution to the combined standard UoM (%)						Expanded UoM (%) $k = 2$
		zer (2a)	res (2a)	rpr (2a)	rep (2a)	int (2a)	cmc (2a)	
1	0.03755	0.019	0.027	0.053	0.133	0.178	0.015	0.152
3	0.11624	0.019	0.009	0.017	0.103	0.017	0.015	0.069
5	0.19492	0.019	0.005	0.118	0.021	0.037	0.015	0.065
7	0.27357	0.019	0.004	0.117	0.015	0.030	0.015	0.062
9	0.35218	0.019	0.003	0.088	0.023	0.006	0.015	0.051
11	0.43074	0.019	0.002	0.067	0.065	0.024	0.015	0.059
12	0.47009	0.019	0.002	0.064	0.000	0.022	0.015	0.045
13	0.50953	0.019	0.002	0.063	0.012	0.005	0.015	0.042
14	0.54892	0.019	0.002	0.046	0.026	0.001	0.015	0.040
15	0.58839	0.019	0.002	0.068	0.058	0.018	0.015	0.056

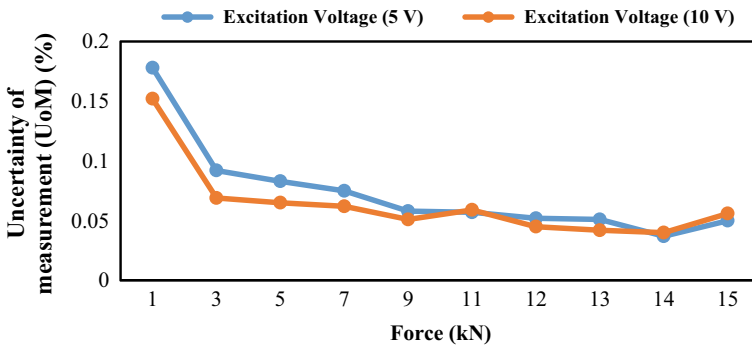


Fig. 5 Plot showing UoM ($k = 2$) of EN 8 SGFT at Position A; 5 and 10 V excitation voltages

4 Conclusion

The metrological characteristics of EN 8 SGFT have been found within the permissible limits to be applied in precision force measurement applications. The selection of the location of strain gauging affects the transducer’s sensitivity that has been concluded from an experiment performed using EN 8 SGFT. The strain gauging is performed at Position A and Position B. It is observed that the mean values of observations are less at Position B, because of the presence of low stress–strain values at this location. The effect of excitation voltage has been examined by supplying an excitation voltage of 10 V to EN 8 SGFT at Position A. A comparative graph of

UoM between 5 V and 10 V excitation voltage shows that the difference between the values is insignificant.

Nomenclature

FEA	Finite element analysis
DWFM	Deadweight force machine
SGFT	Strain gauge force transducer
zer	Relative uncertainty due to zero offset
res	Relative uncertainty due to resolution
rep	Relative uncertainty due to repeatability
rpr	Relative uncertainty due to reproducibility
int	Relative uncertainty due to interpolation
cmc	Machine uncertainty
UoM	Uncertainty of measurement
<i>k</i>	Convergence factor
<i>a</i>	Parameter refers to the difference between mean and maximum values of a factor

References

1. Fraden J (2016) Force and strain—handbook of modern sensors. Springer, Cham. https://doi.org/10.1007/978-3-319-19303-8_10
2. Stefanescu DM (2011) Handbook of force transducers: principles and components. Springer Science & Business Media
3. Chen B, Wu X, Peng X (2007) Finite element analysis of ring strain sensor. Sens Actuators A 139:66–69
4. Kumar H, Kumar A (2012) Influence of excitation voltage over metrological performance of force transducers. In: XX IMEKO world congress metrology for green growth, Sept 9–14, Busan, Republic of Korea
5. Saxena R, Ghoshal SK, Kumar H (2021) Design, development and metrological investigation of a trapezoidal-shaped force transducer. IJPAP 59:835–844
6. Bhateja A, Verma A, Kashyap A (2012) Study the effect on the hardness of three sample grades of tool steel i.e., EN-31, EN-8, and D-3 after heat treatment process such as annealing, normalizing, and hardening and tempering. Int J Eng Sci 1(2):253–259
7. Hoffmann K (1966) Practice hints for installation of strain gauges. Hottinger Baldwin Messtechnik GmbH, HBM, pp 1–56
8. Stefanescu DM (2011) Strain gauges and Wheatstone bridges—basic instrumentation and new applications for electrical measurement of non-electrical quantities. In: 8th international multi-conference on systems. Signals & Devices
9. Kumar H, Sharma C (2012) Role of finite element analysis in improving hysteresis error of force transducers. Trans Inst Meas Control 34(8):1019–1024
10. Czichos H, Saito T, Smith L (2011) Springer handbook of metrology and testing, 2nd edn. Springer, London-New York

11. Metallic materials—calibration of force proving instruments used for verification of uniaxial testing machines. ISO 376-2004/2011
12. Placko D (2006) Metrology in industry—the key for quality, 1st edn. Wiley-ISTE Publisher
13. (2007) Uncertainty in measurement. In: Measurement uncertainty. Springer Series in Reliability Engineering. Springer, Boston, MA
14. Mihaiela I, Rodica R (2012) Management of uncertainty evaluation process in calibrating a force measuring device. *J Electr Electron Eng* 5(1):95–100
15. Saxena R, Ghoshal SK, Kumar H (2019) A critical assessment of simple shaped force transducers: design and metrological consideration. *Indian J Pure Appl Phys* 57:42–51
16. Kumar H, Sharma C, Kumar A, Arora PK (2015) Design, development and metrological characterization of a low-capacity precision industrial force transducer. *ISA Trans* 58:659–666

Improvisation in Human–Robot Interaction Using Optimized Multimodal Operational Techniques



Pratiksha Prakash Jawale and Shantipal S. Ohol

Abstract Industry 4.0 fosters the use of cyber-physical systems which aims at combining the efforts from humans, robots with cutting-edge technologies to enhance productivity in industrial setting. Human–robot interaction (HRI) supports this aim by combining the skills of human and robot. This paper presents the HRI in industrial scenario. Utilizing a multimodal operational strategy can improve flexibility and dependability and provide several options for user interaction to better suit their demands. Review of different multimodal techniques is also presented in this paper. Each input method has benefits and drawbacks of its own. This paper also highlights the way to overcome these limitations.

Keywords Human–robot interaction · Collaborative robot · Multimodal operational techniques

1 Introduction

Industry 4.0 fosters the use of cyber-physical systems (CPS). Cyber-physical systems (CPS) are a brand-new class of systems that can interact with people using a wide range of cutting-edge communication modes. Its ability to interact with and improve physical environment capabilities through computing make it acceptable [1]. The ability to communicate and control is a crucial component in developing technology. Future production scenarios will employ modular and effective manufacturing technologies that will allow goods to manage their own manufacturing processes [2].

Human–robot interaction (HRI) refers to the interaction between humans and robots. The goal of HRI is to create robots that are able to interact with humans in a

P. P. Jawale (✉) · S. S. Ohol

Mechanical Engineering Department, COEP Technological University, Pune, Maharashtra, India
e-mail: ppj21.mech@coeptech.ac.in

S. S. Ohol

e-mail: sso.mech@coeptech.ac.in

natural and intuitive manner, allowing for effective communication and collaboration between humans and robots. This allows an operator to work shoulder-by-shoulder with robot instead of the caged isolation. This involves not only the development of the physical and computational components of the robot, but also the design of user interfaces, algorithms for perception and decision-making, and models of human behavior and cognition. It is a field of study that involves the design, implementation, and evaluation of interfaces and algorithms that allow for effective communication and interaction between humans and robots. Human–robot interaction (HRI) and cyber-physical systems (CPS) are closely connected since both involve the fusion of computer and physical systems. Robots are seen as a sort of CPS, while HRI refers to interactions between humans and robots. Multimodal operational techniques for robots refer to the integration of multiple control modalities or methods for controlling the behavior of robots. The modalities can include visual, auditory, haptic, and other sensory inputs, as well as direct human interaction through gestures, touch, or speech. The goal of multimodal control is to provide robots with a more flexible and intuitive means of interaction, making them easier to use and more versatile in a wider range of environments and tasks. The paper is further organized as in first section we first discuss human–robot interaction and collaborative robot. In second section different multimodal techniques to be discussed. Need for optimization of multimodal operational techniques is discussed in further section.

2 Human–Robot Interaction

Many a times human–robot interaction is also called human–robot collaboration but there is thin distinguished line that can be seen between these two terms. These are communication and collaboration, respectively [3]. The distinction between coexistence, cooperation, and collaboration is made by Schmidtler et al. [3], and it can be seen in Fig. 1

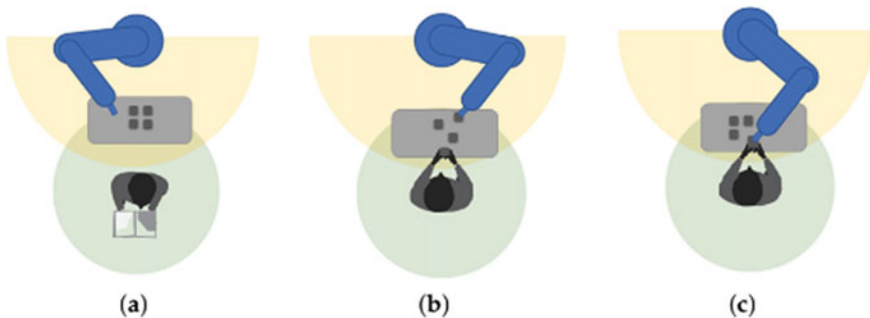


Fig. 1 Collaboration levels **a** coexistence, **b** cooperation, **c** collaboration [4]

In coexistence, humans and robots coexist by working closely together without interfering with one another's jobs. Cooperation refers to both working together to achieve a common goal at the same time, and collaboration is a cooperation in which robot gives real-time response to worker movement [3]. Hence, it is also called human–robot collaboration (HRC). This type of human–robot collaboration includes the design of user interfaces, algorithms for perception and decision-making, and models of human behavior and cognition. HRI is broad field of research.

Human–robot interaction is for combining the skills of both human and robot to achieve better industrial task execution. The fact that HRC integrates the abilities of both humans and robots is what has primarily made it successful in industrial settings [4]. This teamwork can overcome each other's limits and bring about the necessary changes. Robots, for instance, are unable to make decisions in undesirable situations, but humans can think of alternatives right away and instruct robots to take the proper action [4]. Human–robot interaction is primary concern in majority of robotic applications. Panagiota et al. [5] give highlights on human–robot interaction in an assembly cell. Lunghi et al. [6] described the accessibility of remote robotic intervention in dangerous environments via human–robot interfaces.

2.1 Need of HRI

Mass customization, or ever-higher levels of flexible production with mass efficiency, is what modern manufacturers strive for [7]. Industry 4.0 is an industrial shift toward smart manufacturing. The need for manufacturers to offer flexible goods and services is growing [8]. In order to overcome this difficulty, precise jobs that call for dexterity, flexibility, and cognitive decision-making are typically performed by human operators [9].

Interfaces are necessary for a tight relationship between a robot and a person, as well as for information exchange and programming. The interaction modules control how the human–robot collaborative work is carried out in terms of human and robot communication. The body motion capture and gesture recognition modules are two examples of such modules. It uses low-cost depth sensors to allow for natural interaction and robot control [10].

2.2 Collaborative Robot—An Application of HRI

Collaborative robots or cobots are revolutionizing human–robot interaction in the manufacturing sector. Cobots, also referred to as collaborative robots, are developed to collaborate with humans in a shared workspace. Collaborative robots are designed to handle jobs that are too risky, boring, or physically taxing for human workers to complete alone, hence enhancing human capabilities and increasing total production.

Typically including safety measures like sensors, force-limiting technology, and software algorithms that let them to halt or slow down in response to unexpected contact with human employees, collaborative robots are made to be safe to work with [11]. As a result, they can work close to people without endangering their safety.

Numerous industries, including as manufacturing, assembly, packing, material handling, and many more, can benefit from the deployment of collaborative robots. Collaborative robots are revolutionizing how humans and robots interact in the manufacturing sector. Collaborative robots are becoming increasingly popular among businesses of all shapes and sizes, and it makes sense for them to do so. They can be set up to work in a more adaptable and flexible way, or they can be programmed to carry out specific tasks.

2.3 Research Areas in HRI

Human–robot interaction (HRI) is a multidisciplinary field that includes elements of computer science, psychology, engineering, and design [4]. To better understand how people and robots interact in various circumstances, broad research is going on. Following are some of the major research areas in HRI [4, 10]:

1. Task dynamic analysis
2. Human-centered design
3. Cognitive psychology
4. Task planning and programming
5. Techniques and sensors for HRI
6. Intuitive interface design.

Robots used in HRI in industry are usually controlled by rigid, automatically generated programs that perform predetermined tasks. Robot programming and the requirement for specialist skills for robot [re]programming hinder the widespread deployment of HRI in industrial tasks. Furthermore, the dynamic HRC environment frequently results in the replanning and reprogramming of activities in response to modifications in an industrial activity [12]. Engineers are less able to adapt when faced with uncertainty since they have been forced to focus more on traditional offline/online programming. To address these problems, many researchers have used the multimodal operational technique. Therefore, we shall explore multimodal operational strategies utilized to improve the HRI in the following section.

3 Multimodal Operational Techniques

Intuitive interaction between humans and robots depends on how robot quickly adapting to human behavior. To achieve this, a multimodal control-based paradigm was presented by many academics. For a better and more effective HRI in industrial

settings, effective communication between human and robot partners is essential. As a result, the robot should be able to control a variety of social behaviors, voices, gestures, and faces. This is referred to as a cognitive human–robot interaction [13]. Multimodal operation techniques for robots refer to the integration of multiple control modalities or methods for controlling the behavior of robots.

3.1 Research in Multimodal Interaction Techniques

Liu et al. [14] in his work used the vocal command, hand motion, and body motion as input modality. A deep learning algorithms are used for multimodal fusion in a robust multimodal HRC manufacturing system architecture. In [15], multimodal interface is developed which make use of two camera sensor to enable a robot to assist a human operator in a screwing operation. It is a sensor-based method called visual servoing. The two sensors that are employed for intuitive interface execution are Microsoft Kinect, which monitors the workspace from a fixed stance, and a camera that is rigidly attached to the robot end effectors. Liu et al. [16], proposed a system that uses multimodal intuitive control and event-driven FBs to enable operator interaction with a multirobot work cell in a secure and open environment. Sensorless haptics, gestures, and voice are the modalities used in his work. This system describes distributed control and decision-making that is made possible by the running of algorithms built into FBs. In order to make industrial robot programming easier, Arsenopoulos and Benardos [17] created an RGB-D sensor-based human–robot interface (HRI). In his study, two novel user interaction approaches are utilized to gather the operator's motions and gestures and use them to position the joints of a 6 DOF articulated robot arm. The first three robot joints are located by measuring the operator's torso, shoulder, and elbow; the following three joints are located by measuring the operator's motions. Thus, the robotic arm is controlled by a mix of body position and hand gestures. A technique for streamlining industrial robot programming by using optical sensors that recognize human actions is put forth by Panagiota et al. [18].

According to Kandalaf et al. [19], a robotic arm is designed and built using 3D printing and is controlled by gesture movements and voice instructions. Accelerometer and speech recognition software based on a Bluetooth connection protocol for the microcontroller are used to produce the gesture data. Another multimodal control for human–robot collaborative assembly was created by Liu et al. [20]. In this, he has used brainwave, gesture, and voice as input modalities and used deep learning algorithms, LSTM hand gesture recognition and CNN for voice recognition. A convolutional network is created for the precise and dependable conversion of brainwave command words into orders for robots. A multimodal human–robot interaction based on the merging of voice and gesture is presented by Deng et al. [21]. The interface is developed with a robot control command system that can translate user words and gestures into commands the robot can carry out. This system uses the Microsoft Speech SDK to record operator speech. Then, directives are generated using a corpus-based technique of maximum entropy classification for natural language understanding. In this

system, the operator’s gesture is captured using leap motion. In order to estimate the collected data and decrease the sensor’s inherent noise, interval Kalman filter is used. Depending on the environment’s, task’s, and interaction type’s features, Kardos et al. [22] give a thorough explanation of the elements and specifications of a human–machine interface meeting the needs of human–robot interaction. The interaction context may change from operation to operation. Neto et al. [23] suggest using a robot to help a human coworker transport tools and parts and hold things during an assembly activity. Inertial measurement units [IMUs], which are wearable sensors, are used to record human upper body motions. An artificial neural network [ANN] is fed blocks of static and dynamic data to classify static, dynamic, and composed gestures. A context-sensitive, multimodal human–machine interface that was developed and put to the test in a trial implementation was also offered as a potential solution. Some authors used face recognition as an interaction mode with robots. El Makrini et al., perform a face recognition to authorize a worker to allow to work with robot. It is performed through a Kinect camera. The images are processed through IAI Kinect 2 ROS package [24]. The various multimodal operational techniques employed by the researchers are depicted in Fig. 2. We will discuss these modalities one by one.

Hand Gesture Recognition:

Hand gesture recognition in multimodal interaction can provide a more natural and intuitive way for users to interact with computers and has the potential to revolutionize the way that people interact with technology. The method of identifying and understanding human hand motions using computer vision algorithms is known as hand gesture recognition. Typically, to recognize a hand gesture, photos or videos of the hand must be taken, the data must be processed to extract the necessary information, and the gesture must then be classified using machine learning techniques. The position and orientation of the hand, its shape, and the movement of the fingers are some of the details that are commonly derived from the collected data.

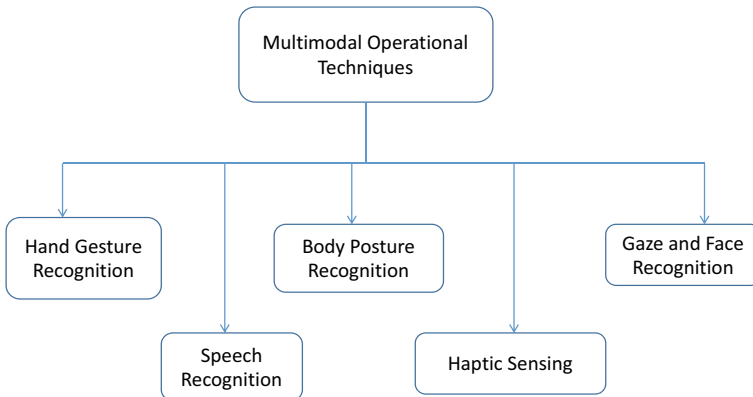


Fig. 2 Multimodal operational approaches

There are several methods for recognizing hand gestures, including template-based matching, skin color analysis, feature-based methods, and deep learning-based methods. Each of these approaches has its own strengths and limitations, and the method of choice is determined by the application's particular requirements. In recent years, deep learning-based methods have become increasingly popular for hand gesture recognition due to their high accuracy and robustness, recurrent neural networks (RNNs) and convolutional neural networks (CNNs) [16] are two commonly used deep learning architectures for hand gesture recognition. The sensors used in hand gesture recognition system are cameras and depth sensors [17, 25], inertial measurement units [23], leap motion sensor [20], etc. The Kinect V2 sensor used by many authors to recognize gestures and speech by many authors [26], consists of three parts: a depth sensor, an RGB camera, and a set of four microphones. These three parts allow it the ability to be controlled by gestures and speech. A dependable, infrared thermal imaging-based hand gesture detection system that is independent of light imaging is suggested by Breland et al. [27]. 14,400 thermal hand movements are compiled into a dataset and divided into two color tones. In order to effectively classify high-resolution hand motions, the author also suggests utilizing a deep CNN.

Body Posture Recognition:

Robots can use body posture recognition to better understand human intentions and emotions and to respond in a more natural and intuitive way. To perform body posture recognition, researchers typically use machine learning algorithms to process data from sensors, cameras, or other sources [18]. The algorithms can be trained on large datasets of body posture data to learn the features that are most relevant for recognition, such as the position of the arms, legs, and head, or the overall body posture.

Speech Recognition:

The connection between people and robots can be made better overall with the use of speech recognition technologies, making it easier and more intuitive for humans to control and interact with the robot. In order to effectively implement speech recognition technology in robots, it is crucial to take into account elements like the accuracy and reliability of the speech recognition system, the robot's ability to understand and respond to different accents and speaking styles, and the robot's ability to distinguish between multiple speakers in a noisy environment. The performance of speech recognition systems depends on factors such as the quality of the speech signal, the accuracy of the acoustic and language models, and the size and quality of the training data used to develop the models.

Noise disturbances are the main problem in implementing speech as an input modality. Carlos Toshinori Ishi created a robust voice recognition system in a noisy environment [28]. Author has achieved word accuracy of more than 80% for both adults and 70 dBA of background restaurant noise during a child's speaking. For adult speech, further testing in a noisy environment produced word accuracy of 73%. In [26], author developed speech recognition system which gives 91.43% recognition accuracy with 60 dB noise interferences. Using a microphone, the MS

Speech Recognition Libraries API was used to implement the speech recognition development.

Haptic Sensing:

Haptic sensing is the practice of providing the user with feedback through touch. In order to interact with the physical environment through touch, it uses sensors and actuators that can detect and produce touch sensations. In human–robot interaction, haptic sensing can be used to improve the realism and naturalness of the interaction by providing touch-based feedback to the user.

Gaze and Face Recognition:

Gaze recognition refers to the ability to identify and interpret the direction and focus of a person’s gaze. Face recognition in HRI is mainly for allowing operator to control the robot. Both of these have limited use as modality in industrial setting because unintentional movements can also be recognized and wrong interpretation can happen.

4 Review and Discussion

Unimodal interfaces, in which a single mode is used to operate the robot, while multimodal interfaces, offer two or more input modes. According to numerous researches, multimodal interfaces are preferred over unimodal ones because of the advantages outlined below. This provides interaction alternative to fulfill the need of diverse users according to their usage patterns and preferences [16]. Alternation and integrated use are possible. It can eliminate the limitation of modalities by effectively utilizing other modality instead. Compared to a speech-only interface [25, 26], they can enable shorter and simpler spoken utterances, which lead to fewer disfluencies and more accurate speech recognition.

The third advantage ultimately results in improved error prevention and simplicity in error resolution, works efficiently at all times irrespective of environmental conditions, and provides flexibility and reliability in operation gaze recognition that has limited use in industry as some users’ eye movements are frequently unintentional. As a result, the system responds in an undesired manner. It is expensive technology due to costly hardware requirements. It is difficult to control eye position accurately all the times. Human action recognition like body recognition and face recognition has limited use in industrial application as it does not provide simple way of communication though they are intuitive ways. In body recognition, operator has to take a definite posture which is not ergonomic way of communication. Face recognition is limited to the identification of operator for controlling of robot. Haptic sensing needs expensive haptic devices for its input. Most intuitive multimodal inputs used reliably in industrial application are gesture and speech.

The numerous multimodal methods to raise the HRI in an industrial setting are described above. Some authors have combined two or three different modes to obtain the desired multimodal interface, and others have employed numerous cameras to

make it multimodal [15]. Again, many authors have employed these various techniques to carry out various tasks in one operation cycle [14]. For instance, moving the arm with the voice and a gesture can operate grippers in a pick-and-place operation. Looking at Industry 4.0, there is a need to use multimodal operational approach in a productive way in light of this extensive field of research, the benefits of a multimodal approach, and the increased use of human–robot interaction in the industrial sector.

One of the ways to put this to industrial use is by optimizing these multimodal inputs. The most effective approach to control a robot can be determined and can be done through voice, gesture, or any other modality as stated above. In some scenarios, combining different modes may increase engagement, yet in others, a single mode may be effective. Figure 3 shows the possible combinations of industrial applications and multimodal operational techniques. This gives an operator flexible use of modes according to his/her capabilities. Also, this can accommodate differently able person in industrial era. The most important we can choose the most efficient mode of operation. Optimizing the multimodal operation can reduce the time of selection of modality and provide efficient mode of operation for understanding the application and ease of use of modality.

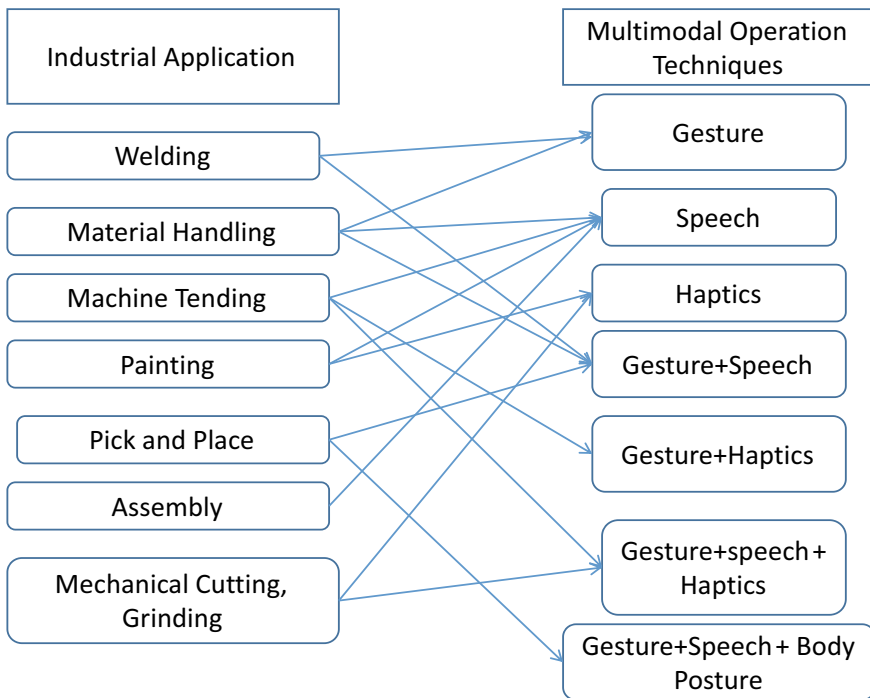


Fig. 3 Possible combination of industrial applications and multimodal operational techniques

5 Conclusion

Human–robot interaction in industrial settings has received a lot of attention recently due to its importance as a modern manufacturing tool in industrial robotization. The many of industrial robotic applications place a high priority on human–robot interaction. The versatile usage of input modes, such as alternation and integrated use, is made possible by multimodal interfaces. The use of multimodal techniques will improve error prevention and easier error resolution. A wider spectrum of users can use it. They can adjust to constantly shifting environmental conditions. They can handle a greater variety of manufacturing processes as well. All the mentioned modalities may not be useful in industrial applications. Each input method has benefits and drawbacks of its own. Combination of two or three or using single modality may perform better depending on an application. Optimization of the multimodal operations can be one of the ways toward enhancing human–robot interaction in industrial scenario.

References

1. Baheti R, Helen G (2011) Cyber-physical systems. *Impact Control Technol* 12(1):161–166
2. Lasi H, Fettke P, Kemper HG et al (2014) Industry 4.0. *Bus Inf Syst Eng* 6:239–242
3. Schmidler J, Knott V, Hölzel C, Bengler K (2015) Human centered assistance applications for the working environment of the future. *OER* 12:83–95
4. Panagiota T, Sotiris M, George C (2016) Human–robot interaction review and challenges on task planning and programming. *Int J Comput Integr Manuf* 29(8):916–931. <https://doi.org/10.1080/0951192X.2015.1130251>
5. Panagiota T, Alexandros-Stereos M, Sotiris M, George C (2017) On human robot interaction in an assembly cell. *Int J Comput Integr Manuf* 30:580–589
6. Lunghi G, Marin R, Di Castro M, Masi A, Sanz PJ (2019) Multimodal human-robot interface for accessible remote robotic interventions in hazardous environments. *IEEE Access* 7:127290–127319. <https://doi.org/10.1109/ACCESS.2019.2939493>
7. Wang Y et al (2017) Industry 4.0: a way from mass customization to mass personalization production. *Adv Manuf* 5:311–320. <https://doi.org/10.1007/s40436-017-0204-7>
8. Solvang B, Sziebig G (2012) On industrial robots and cognitive info-communication. *IEEE 3rd international conference on cognitive infocommunications (CogInfoCom)*. IEEE; pp 459–464
9. Zhong RY, Xu X, Klotz E, Newman ST (2017) Intelligent manufacturing in the context of industry 4.0: a review. *Engineering* 3(5):616–630. ISSN 2095-8099. <https://doi.org/10.1016/J.ENG.2017.05.015>
10. Berg J, Lu S (2020) Review of interfaces for industrial human-robot interaction. *Curr Robot Rep* 1:27–34. <https://doi.org/10.1007/s43154-020-00005-6>
11. Javaid M, Haleem A, Singh RP, Rab S, Suman R (2022) Significant applications of Cobots in the field of manufacturing. *Cogn Robot* 2:222–233. ISSN: 2667-2413. <https://doi.org/10.1016/j.cogr.2022.10.001>
12. Wang L, Gao R, Vánca J, Krüger J, Wang XV, Makris S, Chryssolouris G (2019) Symbiotic human-robot collaborative assembly. *CIRP Annals* 68(2):701–726
13. Sbanca MP, Mogan GL (2016) Cooperative assembly using two industrial robots. In: Borangiu T (ed) *Advances in robot design and intelligent control*. Springer International Publishing, Cham, pp 47–57

14. Liu H, Fang T, Zhou T, Wang L (2018) Towards robust human-robot collaborative manufacturing: multimodal fusion. *IEEE Access* 6:74762–74771. <https://doi.org/10.1109/ACCESS.2018.2884793>
15. Cherubini A, Passama R, Philippe F, André C (2015) A unified multimodal control framework for human–robot interaction. *Robot Auton Syst* 70:106–115. ISSN: 0921-8890
16. Liu S, Wang L, Wang XV (2021) Function block-based multimodal control for symbiotic human–robot collaborative assembly. *J Manuf Sci Eng* 143(9):091001
17. Arsenopoulos K, Benardos P (2021) A human-robot interface for industrial robot programming using RGB-D sensor. *Procedia Manuf* 55:9–15
18. Panagiota T, Athanasios A, Sotiris M, Xenofon C, George C (2016) High level robot programming using body and hand gestures. *Procedia CIRP* 55:1–5. ISSN: 2212-8271
19. Kandalafi N, Kalidindi P, Narra S, Saha HN (2018) Robotic arm using voice and Gesture recognition. In: *IEEE 9th annual information technology, electronics and mobile communication conference (IEMCON)*, pp 1060–1064. <https://doi.org/10.1109/IEMCON.2018.8615055>
20. Liu S, Wang L, Wang Xi V (2022) Multimodal data driven human–robot collaborative assembly. *J Manuf Sci Eng* 144(5):051012. <https://doi.org/10.1115/1.4053806>
21. Deng Y, Fang L, Xin H (2018) Research on multimodal human-robot interaction based on speech and gesture. *Comput Electr Eng* 72:443–454. ISSN: 0045-7906
22. Kardos C, Kemény Z, Kovács A, Balázs E, Váncza J (2018) Context-dependent multimodal communication in human-robot interaction. *Procedia CIRP* 72:15–20. ISSN: 2212-8271
23. Neto P, Simão M, Mendes N et al (2019) Gesture-based human-robot interaction for human assistance in manufacturing. *Int J Adv Manuf Technol* 101:119–135
24. El Makrini, Merckaert K, Lefeber D, Vanderborcht B (2017) Design of a collaborative architecture for human-robot assembly tasks. In: *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Vancouver, BC, Canada, pp 1624–1629
25. Auquilla AR, Salamea HT, Alvarado-Cando O, Molina JK, Cedillo PAS (2019) Implementation of a telerobotic system based on the kinect sensor for the inclusion of people with physical disabilities in the industrial sector. In: *IEEE 4th Colombian conference on automatic control (CCAC)*, Medellin, Colombia, pp 1–6. <https://doi.org/10.1109/CCAC.2019.8921359>
26. Kaczmarek W, Panasiuk J, Borys S, Banach P (2020) Industrial robot control by means of gestures and voice commands in off-line and on-line mode. *Sensors* 20(21):6358. <https://doi.org/10.3390/s20216358>
27. Breland DS, Dayal A, Jha A, Yalavarthy PK, Pandey OJ, Cenkeramaddi LR (2021) Robust hand gestures recognition using a deep CNN and thermal images. *IEEE Sensors J* 21(23):26602–26614. <https://doi.org/10.1109/JSEN.2021.3119977>
28. Ishi CT et al (2008) A robust speech recognition system for communication robots in noisy environments. *IEEE Trans Robot* 24(3):759–763. <https://doi.org/10.1109/TRO.2008.919305>

A Co-occurring Rule Mining Approach to Discover Viewers Watching Behavior of OTT Platform



Apurva Vashist, Suchismita Mishra, and Anil Kumar Sagar

Abstract People can watch movies at their convenience from their hectic lives in the present era of web platforms. There are many platforms that offer movie material. Yet, the user finds it quite difficult to search for and select a movie because there are such large movie databases. When I hear the phrase “OTT,” I think back to a time when CD players were still widely used and DVD players were in style. The availability of cable connections, the emergence of the pen drive, and YouTube’s prominence as a significant viewing platform all come about gradually as technology improves. With these variations, OTT video streaming is one of the most recent market trends. Users are motivated to sign up for OTT video streaming services due to the availability of material, ease of access to a range of shows, user-friendliness, and consistency of the shows. The current study attempts to look into the connections between OTT platform purchase intention, consumer engagement, brand image, and digital media marketing (DMM) in the Indian context. The discovered data was applied to the Apriori algorithm to identify the co-relation components as well as any undiscovered, intriguing relationships between them. In this work, the hidden patterns of co-occurring relations are extracted with modified-Apriori algorithm and visualization techniques are used to represent the knowledge. This might help in making better use of these digital platforms.

Keywords Over the top (OTT) · Data mining · Data visualization · Platformic strategies

A. Vashist

Department of Computer Science, Ambedkar Institute of Technology, Delhi 110092, India

S. Mishra (✉)

Department of Computer Science, Indira Gandhi Delhi Technical University for Women, Delhi 110006, India

e-mail: suchismita.mishra8@gmail.com

A. K. Sagar

Department of Computer Science and Engineering, Sharda University, Greater Noida, Uttar Pradesh 201310, India

e-mail: anil.sagar@sharda.ac.in

1 Introduction

When a customer bought a cable subscription in the past, their cable TV provider was in charge of the availability and supply of programming. Presently, the cable company can only give you access to the Internet; they have no influence over your consumption. This division has a big impact on advertising. The cable company only provides the Internet connection; it no longer monitors what you consume. Although many OTT services are subscription-based and feature advertisements, many also have tiered packages that allow users to pay for ad-free experiences. Because OTT is still a relatively new idea, there is a ton of possibility for growth. With more companies entering the OTT market, customers have more alternatives, and marketers have access to more ad inventory. Over the top, or OTT as it is more often known, is a technology that gives content owners control over how video material is delivered across a variety of devices with the aid of the Internet [8]. OTT is steadily gaining acceptance as a replacement for traditional cable television as a programming choice. The word over the top (OTT) refers to services like movies and television shows that are provided by content owners, broadcasters, distributors, etc., like production firms, and broadcasting channels over the top of the Internet.

Client-side and service-side setups for OTT marketing are the two main categories.

With a client-side arrangement, when streaming media is being viewed, the advertisement loads before the episode or movie. Due to the advertisement's seamless integration into the media's frames, server-side ad insertion eliminates ad-blocking. Because it is technically far more challenging to support server-side ad insertion, it is still a rather uncommon practice. The development of OTT has been attributed to a number of factors, including smartphones, incredibly fast IP networks, open source platforms, creative services, cutting-edge capabilities, and a shift in consumer preferences toward its "freemium"-based business models [6] (Fig. 1).

OTT streaming services function like a television network that shows your favorite shows. OTT differs from traditional broadcasting in that it reaches users directly over the Internet rather than through a cable operator. The top OTT services are Netflix, HBO, Disney + Hot Star, Amazon Prime Video, etc. (Fig. 2).

2 Literature Review

Another crucial problem that significantly affects the growth of OTT mobile consumers is the price model for mobile content. In the next days, it will gradually become necessary and a regular portion of free time. The future of entertainment is on-demand. India is the second-largest and smartphone market with the quickest growth [6].

According to Ko and colleagues [4], a set of numbers is said to be sparse if a large fraction of the values are assigned to the constant default value. By utilizing a

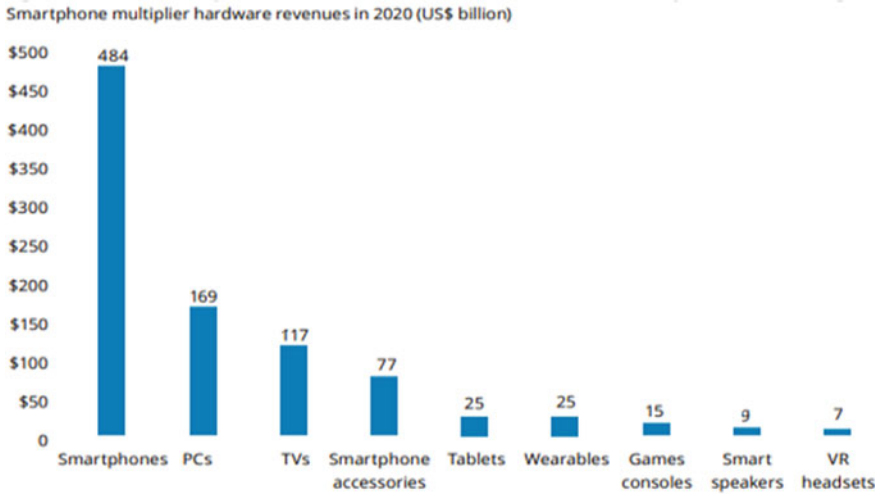


Fig. 1 High rise in smartphone sales in 2020. Source Yeole [8]



Fig. 2 OTT platform working model

hybrid recommendation model, this issue is resolved. Their method met the requirement for the rating data by incorporating the information. This algorithm uses both collaborative filtering and content-based filtering techniques.

Another study that compared new media to traditional media claimed that the rise of the Internet will unnecessarily lead to the extinction of newspapers and television [1]. The Internet attracts to certain user segments exclusively (those between the ages of 18 and 37), while other user segments still have varying degrees of commitment to traditional media, despite the fact that numerous studies have demonstrated the Internet’s enormous potential. Even if the Internet finally totally replaces traditional

media, as time goes on, it will happen more slowly than anticipated because the outright replacement is being driven by a small age group (those between the ages of 18 and 37).

OTT was essential during the lockdown to ease strain in the lives of the working populace, although not being as strong and vital as a bigger screen, Dolby digital sound, etc. (Karim 2020). The 4G and 5G revolutions, easy access to mobile phones at low rates, the telecom sector transformation, and other factors have all contributed to a huge rise in OTT mobile subscribers.

3 Methodology

In this current research work, the descriptive kind of research design has been used. To reach all the specified goals, both primary and secondary data were gathered. Secondary information was gathered from a variety of published sources, including books, journals, periodicals, websites, etc. Using a survey tool, the customer provided the primary data. A well-constructed questionnaire was created to address the study's varied aspects.

3.1 Model Design

This study was conducted over 1,458 number of OTT platform users in 16 number of different cities across India. The purpose is to determine the usage of several aspects of the OTT platforms and its users. The different cities covered representing across different levels are:

Metro cities—Mumbai, Delhi, and Kolkata.

Tier1 cities—Hyderabad, Pune, Bengaluru, Jaipur, Ahmedabad.

Tier2 cities—Bhubaneshwar, Jamshedpur, Ludhiana, Panipat, Patna, Nagpur, Madurai, Coimbatore.

A series of questions that were answered in-person at each of the centers served as the foundation for the inspection. The purpose of the questionnaire was to gather information about the respondents' OTT usage patterns in relation to the following crucial areas:

- Average watching session duration
- The consumption of long-form versus short-form content
- Movie-watching habits
- Analysis of the consumption of content by day and by time zone
- The number of platforms used and subscribed to by responders
- Motives for subscribing and not subscribing
- Popularity of OTT subscriptions over other similar category services.

3.2 *Experimental Setup*

This study has been carried out in broadly two phases:

Phase 1: Exploratory.

Phase 2: Measurement.

By using the multiple correspondence analysis (MCA) approach, the significant factors are identified. The found out data has been implemented with Apriori algorithm to determine the co-relation factors.

All of the information was manually gathered from several partners throughout the study's initial exploratory phase using pen and paper to describe their levels of satisfaction at various touch points. The points of contact are:

- Average watching duration of each session
- Length of watching content over OTT
- Movie consumption behavior
- Content viewing in day and time band analysis
- Number of OTT platforms being viewed and subscribed by users category
- Growth of popularity of OTT subscriptions
- Reasons for subscription and non-subscription of OTTs.

One of the most intriguing marketplaces and businesses in the world might be found in India's online video industry. After China alone, India has been predicted to reach above 500 million numbers of online platform users by the FY 2023, which will result as the second-largest OTT market worldwide.

This report was put together with a well-balanced viewpoint of online contents, flouting down and analyzing these aspects of viewers, viewing habits, manner of consumption, device kinds, distribution methods, and globalization of video on-demand, with forward-looking proclamations which are based on extensive business analysis, consumer behavior research, real and open data of such online platforms and its services. The outcomes underline once more how important it is for all businesses and coworkers to support the expansion and improvement of this ecosystem.

On a scale of 1 to 10, with 10 representing the best, the outcome has been recorded. To get the intended result, the given data set has been applied to Weka 7.0.

4 **Result Analysis**

For Indian audiences, the online video business has developed into a staple source of entertainment. According to this data, users spent an average of hour 10 min per day. Also, we can say 8.2 h spend per week, watching videos on OTT platforms. For an average session length of about 40 min the viewers watched 12.5 times on an average per week (Fig. 3).

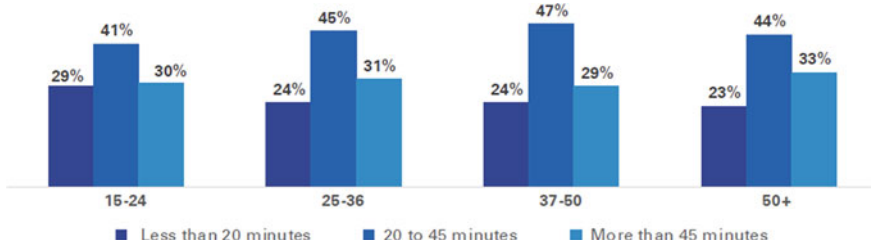


Fig. 3 Average session durations

With the exception of movies, most content types have short enough durations to be consumed in a single sitting. Longer forms of material like movies may now be seen in a variety of ways thanks to OTT services, which enable access from anywhere at any time.

According to the findings of our survey, viewers are starting to exercise more judgment when watching movies on OTT platforms. Around half (approximately 58%) of the viewers who favored movies viewed it multiple times.

This percentage was much greater (around about 85%) for respondents in metro areas compared to the other two types of cities, which may indicate that metro areas have less free time than other types of cities.

In South, a majority (almost 60%) preferred to view movies in a single session, whereas a substantial preference of flexibility in East zone, around 80% of the consumers favoring numerous online sessions. Typical session lengths recorded in our study provided information on patterns of short- and long-forms of online content consumption on OTT platforms.

Smaller session lengths have gained more popularity in younger viewers, maybe as a result of their shorter attention spans, according to the age-based analysis. OTT platforms must be aware of such developments as well as implement content approaches that maintain a good balance of both long and short sessions to appeal to different watchers niches.

Users can now easily access online video content from any location thanks to the availability of cheap mobile data, and the user experience is enhanced by video platforms’ adoption of various strategies like malleable resolutions of video, offline watching facilities to give users a continuous experience along with varying network speeds. The significance of such accessibility has been made clear in this survey, where approximately 2 out of 10 of the participants who are using paid OTT accounts did this, only to have access all contents everywhere. Due to longer commuting hours, this aspect is more significant in metro areas. About 30% of subscribers pay to watch movies wherever they are, making this one of the top three reasons for doing so. The adaptability of online video viewing on smartphones has led to a reasonably spread pattern of content consumption throughout the day.

It was found that during their daily commute, at least 12% of respondents from all professions watched web movies (7 a.m.–10 a.m.). Also, 29% of the viewer’s choosing this session; more than this percentage of viewers are quadrupled for the

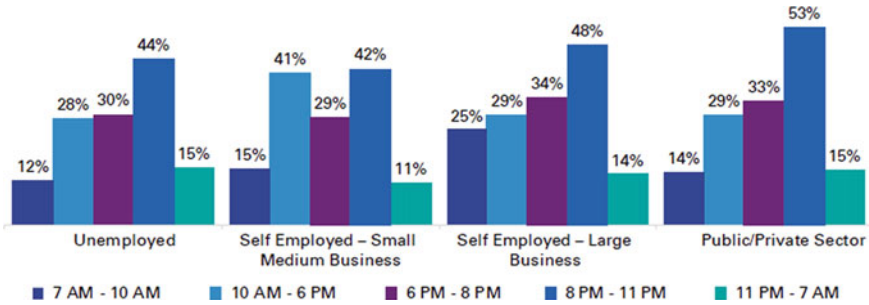


Fig. 4 Preference of time slots for OTT content consumption

night travel (mostly found during 6 pm–9 pm). In addition, it was discovered that, contrary to popular belief, a sizable portion of viewers (approximately 28%) watch during regular business hours (which is from 10 am–6 pm). OTT platforms should strive to maximize customer arrangement by establishing a content approach that should cater to viewer’s needs according to time, given that video consumption is evenly distributed throughout several time periods (Fig. 4).

Content has gained importance for telcos worldwide since 2016, when the trend toward 4G services gained traction. Telco has been collaborating along with numerous OTT platforms (across both SVOD and AVOD) having a variety of content-oriented contributions, including Live TV events, sports events, and many digital originals in an effort to retain subscribers. Telcos serve as crucial distribution channels for online video platforms as well, helping them achieve an acute mass of audiences having little expenditure on their own consumer acquisition expenditures.

The availability of “Live Television” on these telecom applications was discovered to be both the major and determining criterion for usage, with 37% approximate of Airtel TV viewers and around 56% of the Jio TV viewers among the participants indicating the same opinion. The significance of creative material was likewise made clear by the nearly 32% of respondents who cited it as one of the reasons they chose Jio TV and Airtel TV, respectively.

Users can access material behind a paywall using these telco applications without individually subscribing because of the variety of content options available on telco platforms. According to the findings of our survey, roughly 43% of users who access SVOD content but do not pay for it do so via telco apps. Even if the telco is collecting a portion of the cost of the material from customers through the market’s available voice, data, and entertainment package deals.

While practically every OTT platform in India has taken advantage of telco distribution’s reach to increase their audience, several platforms have developed sizable viewership bases on the telco platform through distribution partnerships. The Indian OTT ecosystem’s SVOD platforms Eros Now and Zee5 have some of the widest telco relationships.

Nearly about 34% of participants in metro areas who didn’t subscribe SVOD content because they found enough content available for free of cost or they don’t

Fig. 5 SVOD consumers who do not subscribe to paid OTT

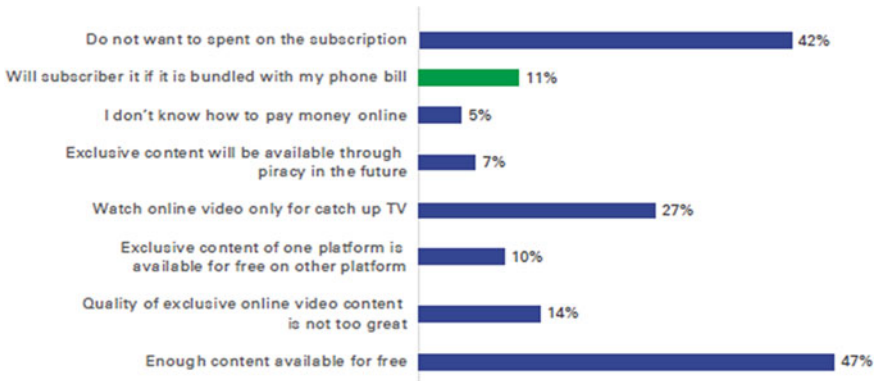
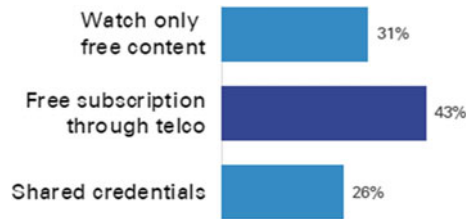


Fig. 6 Factors for AVOD non-subscribed users in metros cities

want to spend on it. Intriguingly, almost 1% customers claimed they would utilize SVOD services if their telco bill includes some subscriptions to Internet video. Recently, initiatives taken by telcos like Vodafone Idea, Airtel who are giving previews on offers available on SVOD platforms as a portion of their premium segment postpaid services in advance finally switching to the carrier billing-based subscription, supporting the aforementioned findings (Figs. 5 and 6).

It was discovered that, in this case, users’ decisions to subscribe to a paid online video platform were most strongly influenced by the content. The topmost three motives respondents chose to pay for an OTT platform included “content quality” and “content exclusivity on the platform.” The same emphasizes the significance of exclusive, top-notch tent-pole content in luring consumers to platforms (Fig. 7).

One of the most significant considerations for choosing between online video platforms, according to respondents, is content. Favorite content on another platform at that specific time was the main driver of platform switching, with “content freshness” coming in third. This emphasizes the significance of tent-pole or hook-based content that can cause consumers to platform switching, which makes it necessary for the online platform to continuously update its collection with new contents in order to maintain user attraction and retention. Yet, given the prevalence of freshness of contents, quality, and pricing does not appear to be the key component in policy-making concerning the migration between OTT platforms, despite the fact that the most important factors are related to financial rewards.

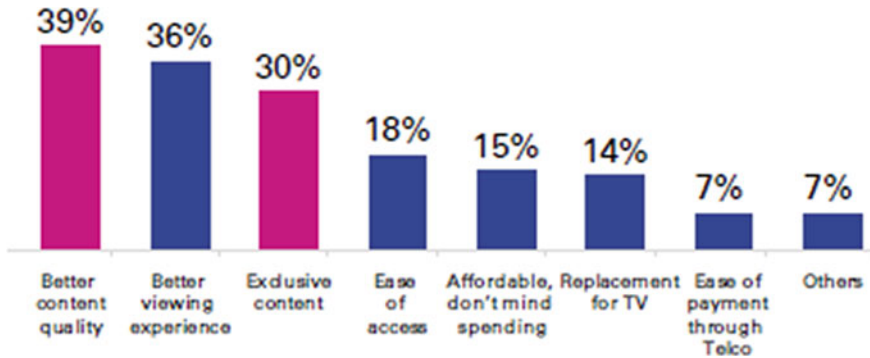


Fig. 7 Factors for subscribing to OTT platform (in terms of %)

5 Conclusion and Future Scope

The study was carried out to get a minute view on the watching behavioral pattern and preference of the viewers. With the easy availability of Internet all over the India, the users prefer to watch every kind of content in their smartphone to retain their privacy of entertainment. This has shown a boom in growth in various OTT platforms in the Indian entertainment sector. Our study findings are describing this behavior of increased usage of OTT platform. Also, it is found that other paid and free platforms of entertainment are in competence with the OTTs for the future.

The study can be wider based on content quality and censoring of the content as the contents is also being viewed by all age groups. Future research might compare the generation and their propensity for using OTT platforms. While this study is primarily concentrated on an urban sample, different findings regarding the rural population and its propensity for OTT may be explored and addressed.

References

- Gaskins B, Jerit J (2012) Internet news is it a replacement for traditional media outlets. *Int J Press/Politics* 17(2):190–213
- ISO/IEC 23000-19 (2018) Information technology—coding of audio-visual objects—part 19: common media application format (CMAF) for segmented media. <https://www.iso.org/standard/71975.html>
- Jemy VC, Dermawan W, Yos S (2018) A proposed selection process in over-the top project portfolio management. *J Ind Eng Manag JIEM* 11(3):371–389, Online ISSN: 2013-0953, Print ISSN: 2013-8423. <https://doi.org/10.3926/jiem.2448>
- Ko H et al (2022) A survey of recommendation systems: recommendation models, techniques, and application fields. *Electronics* 11.1:141
- Lee C, Han D, Han K, Yi M (2022) Improving graph-based movie recommender system using cinematic experience. *Appl Sci* 12(3):1493. <https://doi.org/10.3390/app12031493>
- Rohit JJ (2020) Factors influencing the shift from traditional TV to OTT platforms in India. *Int J Adv Sci Technol* 29(7s):4044–4051

7. Xiangbo L, Jordi C, Yuriy R (2019) Research gate-towards mass deployment of CMAF Robert Peck
8. Yeole S, Bhaisare (2022) A study on user perspective on OTT platform in India. *J Positive School Psychol* 6(3):7351–7364
9. Zahra A-M, Hamed V-N, Javad S (2021) Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Syst Appl* 167

Using Interferometer-Based Photonic Integrated Circuits to Perform Rapid Unitary Transformations



Sathvik Redrouthu, Jagadeepam Maddipatla, David Cao, Ryan Kim, and Steven Lu

Abstract Artificial intelligence (AI) has been growing exponentially over the past two decades. AI-based computational demands have thus been skyrocketing, while at the same time, Moore's Law is beginning to end. Thus, we introduce a transistor-independent optical chip able to perform matrix-vector multiplication. Matrix-vector multiplication is fundamental for computation within AI, graphics, machine learning, image processing, computer vision, graph theory, and quantum computing. We specifically discuss a layout capable of performing a unitary matrix transform by routing light through a lattice of Mach-Zehnder interferometers. We discuss how this transformation can be built up to any arbitrary matrix-vector multiplication. For a given 4×4 unitary matrix, the chip performs at 7 teraflops with a 245 teraflops/watt performance per watt rating. Data analysis from simulations and testing is shown, including measured loss at different wavelengths.

Keywords Data analytics · High speed computing · Machine learning · Matrix-vector multiplication · Neural networks · Optical computing

S. Redrouthu (✉) · J. Maddipatla (✉) · D. Cao · R. Kim (✉) · S. Lu
Procyon Corp, Clearwater, United States
e-mail: 2024sredrout@tjhsst.edu

J. Maddipatla
e-mail: jag.maddipatla@gmail.com

R. Kim
e-mail: 2024rkim@tjhsst.edu

D. Cao
e-mail: 2024dcao@tjhsst.edu

S. Lu
e-mail: 2024slu1@tjhsst.edu

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
R. N. Shaw et al. (eds.), *Innovations in Electrical and Electronic Engineering*, Lecture Notes in Electrical Engineering 1109, https://doi.org/10.1007/978-981-99-8289-9_33

1 Introduction

There has been an exponential growth in the field of generative artificial intelligence (AI), driven by advances in machine learning algorithms, hardware capabilities, and the availability of vast amounts of data. Generative AI models, such as generative adversarial networks (GANs) and variational autoencoders (VAEs), and even newer models constructed using the transformer architecture, have shown remarkable potential in various applications, including image synthesis, text generation, and music composition [1].

One of the key challenges in training and running generative AI models lies in the computational demands imposed by their intricate architectures and the large-scale matrix operations involved. Traditional digital electronic processors, including graphics processing units (GPUs) and tensor processing units (TPUs), have been widely employed to accelerate AI computations [2, 3]. These processors rely on the transistor-based digital electronics that have dominated the computing industry for decades. However, as the limits of Moore's Law and Dennard Scaling loom on the horizon, alternative computing paradigms are gaining attention to overcome these limitations [4].

An emerging avenue for addressing the computational challenges of generative AI is the exploration of unconventional computing architectures, particularly in the realm of optical computing. Unlike traditional electronic processors, optical computing harnesses light-based technologies to perform computations. Optical computing exhibits unique advantages, including high parallelism, low power consumption, and potential for ultra-fast processing speeds. These characteristics make optical computing an attractive candidate for accelerating the training and inference of generative AI models, which often require vast amounts of computational resources.

Recent advancements in the field of silicon photonics have paved the way for the development of novel hardware accelerators tailored specifically for generative AI tasks [5]. Silicon photonics leverages the integration of photonic components on silicon chips, enabling the manipulation and transmission of light signals with high efficiency. In this research paper, we delve into the design and implementation of such an optical chip to perform AI operations. By leveraging the parallelism and speed of photonics to perform important unitary matrix transformations, we strive to enhance the training and inference performance of AI models beyond the constraints imposed by Moore's Law.

We begin by providing an overview of the accelerator's architecture, highlighting the key components and their functionality. The integration of silicon photonics with digital electronic circuits allows us to perform matrix-vector multiplication (MVM) efficiently, a fundamental operation in generative AI computations. We discuss the design considerations and optimization techniques employed to achieve high-speed and low-power MVM on the accelerator.

Finally, we discuss the implications of our findings and the potential impact of silicon photonic chips on the future of AI. The integration of photonics into the AI hardware landscape opens up new avenues for faster and more energy-efficient

computations, circumventing the limitations imposed by Moore's Law and Dennard Scaling. We highlight the benefits of optical computing for generative AI applications, such as real-time image and video generation, interactive content creation, and rapid prototyping in creative industries.

2 Chip Design

2.1 Interferometer

Matrix-vector multiplication is often considered the backbone of artificial intelligence (AI) [6]. One particular matrix, the simple yet fundamental unitary operator, rotates any given vector while preserving magnitude. In the realm of photonics, unitary matrices represent light passing through an interferometer. In the interferometer, the difference between the amplitudes of the input and output lasers is what allows for computations. In this paper, we focus on 4×4 matrices hence an array of 4 lasers. In order to model the 4×4 matrix, we implement 2×2 unitary matrix transformations, which require Mach-Zehnder interferometers with 50 : 50 directional couplers as inputs and outputs, each with a transmission coefficient $\tau = \frac{1}{\sqrt{2}}$ and a coupling coefficient $\kappa = \frac{1}{\sqrt{2}}$. Thus, the circuit model based on the amplitude of the waveguides going into and out of an interferometer is given by

$$\begin{bmatrix} E_{out1} \\ E_{out2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{-j\phi_1} & \\ & e^{-j\phi_2} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} E_{in1} \\ E_{in2} \end{bmatrix} \quad (1)$$

where $\phi_{1,2} = \beta L_{1,2} - \frac{\alpha_{1,2}}{2} L_{1,2}$ and $\beta = \frac{2\pi}{\lambda} n_{eff}$. This setup has previously been demonstrated to achieve a two-dimensional vector rotation [7]:

$$\begin{aligned} \begin{bmatrix} E_{out1} \\ E_{out2} \end{bmatrix} &= \begin{bmatrix} e^{j\theta} \cos \phi & -e^{-j\theta} \sin \phi \\ e^{-j\theta} \sin \phi & e^{j\theta} \cos \phi \end{bmatrix} \begin{bmatrix} E_{in1} \\ E_{in2} \end{bmatrix} \\ &= U^{(2)}(\theta, \phi) \begin{bmatrix} E_{in1} \\ E_{in2} \end{bmatrix} \end{aligned} \quad (2)$$

where $\theta = \frac{\phi_1 + \phi_2}{2}$ and $\phi = \frac{\phi_2 - \phi_1}{2}$.

In terms of the original phase shifts $\phi_{1,2}$, evaluating E_{out1} and E_{out2} with matrix Eq. (1), the transfer functions for each port are given by

$$E_{out1} = \frac{1}{2} (E_{in1}(e^{j\phi_1} - e^{j\phi_2}) + E_{in2}(-e^{j\phi_1} - e^{j\phi_2})) \quad (3a)$$

$$E_{out2} = \frac{1}{2} (E_{in1}(e^{j\phi_1} + e^{j\phi_2}) + E_{in2}(-e^{j\phi_1} + e^{j\phi_2})) \quad (3b)$$

In terms of intensity:

$$I_{out1} = \frac{1}{4} \left| \sqrt{I_{in1}}(e^{j\phi_1} - e^{j\phi_2}) + \sqrt{I_{in2}}(-e^{j\phi_1} - e^{j\phi_2}) \right|^2 \quad (4a)$$

$$I_{out2} = \frac{1}{4} \left| \sqrt{I_{in1}}(e^{j\phi_1} + e^{j\phi_2}) + \sqrt{I_{in2}}(-e^{j\phi_1} + e^{j\phi_2}) \right|^2 \quad (4b)$$

since intensity is the square of the absolute value of the amplitude. Turning off certain input ports, we find that the transfer functions for each port, computed by the quotient of I_{out} and I_{in} , are

$$T_{11} = \frac{I_{out1}}{I_{in1}} I_{in2} = 0 = \frac{1}{4} |e^{j\phi_1} - e^{j\phi_2}|^2 = \frac{1}{2} [1 - \cos(\beta \Delta L)] \quad (5a)$$

$$T_{12} = \frac{I_{out2}}{I_{in1}} I_{in2} = 0 = \frac{1}{4} |e^{j\phi_1} + e^{j\phi_2}|^2 = \frac{1}{2} [1 + \cos(\beta \Delta L)] \quad (5b)$$

$$T_{21} = \frac{I_{out1}}{I_{in2}} I_{in1} = 0 = \frac{1}{4} |-e^{j\phi_1} - e^{j\phi_2}|^2 = \frac{1}{2} [1 + \cos(\beta \Delta L)] \quad (5c)$$

$$T_{22} = \frac{I_{out2}}{I_{in2}} I_{in1} = 0 = \frac{1}{4} |-e^{j\phi_1} + e^{j\phi_2}|^2 = \frac{1}{2} [1 - \cos(\beta \Delta L)] \quad (5d)$$

where $\Delta L = L_1 - L_2$. We neglect loss, so $\alpha_{1,2} = 0$.

2.2 Overall Circuit

Any $n \times n$ unitary matrix can be expressed as a product of $\frac{n(n-1)}{2}$ 2×2 unitary matrices as shown [8]:

$$U^{(n)} = \prod_{i=1}^n \prod_{j=1}^{n-i} T_{n-j, n-j} \quad (6)$$

Here, $T_{i,j}$ is an $n \times n$ matrix where the matrix at (i, j) is a 2×2 unitary matrix:

$$T_{i,j} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & 0 & 0 \\ \vdots & 0 & U^{(2)} & 0 & \vdots \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \quad (7)$$

Since an MZI implements a 2×2 rotation, it follows that any $n \times n$ unitary matrix transformation can be obtained by chaining a sequence of $n \times n$ matrices in the form of $T_{i,j}$. We focus on 4×4 matrices due to space constraints; we realize them by chaining many of the interferometers proposed in Sec 2.1 as shown in Fig. 1 [9]. From Fig. 1, it is apparent that

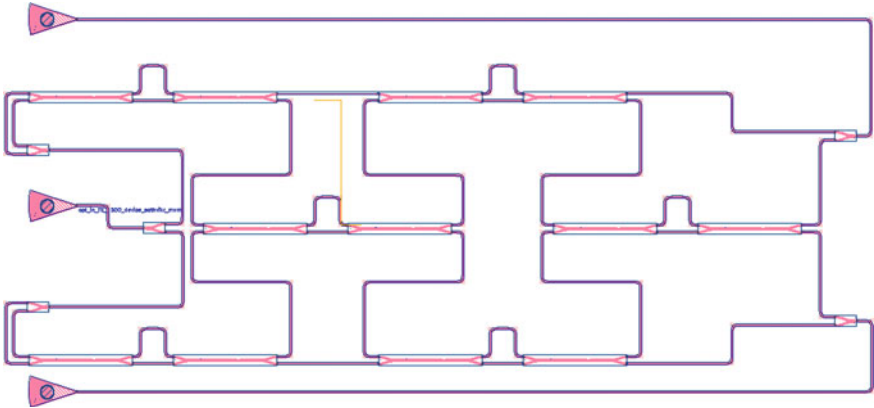


Fig. 1 Circuit to realize a 4×4 unitary matrix. Grating couplers are designed for the fundamental TE mode. The layout fits within the 605×410 space allotted, but parameter variations are not performed due to its relatively large size. Due to restrictions, the number of lasers shown on the figure is limited to 3, but the general structure remains very similar regardless

$$\begin{bmatrix} E_{out(mvm)1} \\ E_{out(mvm)2} \\ E_{out(mvm)3} \\ E_{out(mvm)4} \end{bmatrix} = U^{(4)} \begin{bmatrix} E_{in(mvm)1} \\ E_{in(mvm)2} \\ E_{in(mvm)3} \\ E_{in(mvm)4} \end{bmatrix} \tag{8}$$

where, in this case,

$$U^{(4)} = \prod_{i=1}^4 \prod_{j=1}^i T_{4-i,j} = T_{3,1} T_{3,2} T_{3,3} T_{2,1} T_{2,2} T_{1,1}. \tag{9}$$

Due to a restriction in the number of grating couplers available, a network of Y-splitters are included after the input grating coupler to allow the light to split into each port of the unitary matrix section. Thus,

$$E_{in(mvm)1} = \frac{1}{\sqrt{2}} E_{Y1} = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} E_{in} \right) = \frac{1}{2} E_{in} \tag{10a}$$

$$E_{in(mvm)2} = \frac{1}{\sqrt{2}} E_{Y1} = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} E_{in} \right) = \frac{1}{2} E_{in} \tag{10b}$$

$$E_{in(mvm)3} = \frac{1}{\sqrt{2}} E_{Y2} = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} E_{in} \right) = \frac{1}{2} E_{in} \tag{10c}$$

$$E_{in(mvm)4} = \frac{1}{\sqrt{2}} E_{Y2} = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} E_{in} \right) = \frac{1}{2} E_{in} \tag{10d}$$

Note *mvm* represents the command *mvmul* which multiplies a matrix by a vector on the chip [10]. Refer to Eq. (8). From the two grating couplers, the four output light beams are combined into two light beams:

$$E_{out1} = \frac{E_{out(mvm)1} + E_{out(mvm)2}}{\sqrt{2}} \quad (11a)$$

$$E_{out2} = \frac{E_{out(mvm)3} + E_{out(mvm)4}}{\sqrt{2}} \quad (11b)$$

Hence, the overall circuit is given by The unitary matrix can be built up to any arbitrary matrix through a reverse application of the singular value decomposition (SVD), which states that any matrix M can be expressed as the product of 2 unitary matrices and a diagonal matrix. Such a decomposition enables a larger variety of matrices to be represented with this chip.

3 Chip Implementation

Chip fabrication took place at the commercial Applied NanoTools (ANT) foundry. To implement our chip design and efficiently perform the unitary transformations when requested, we make use of a silicon wafer and various optical components.

3.1 Waveguide

We use strip waveguides throughout our design, each with a height of 220 nm and a width of 500 nm. These configurations allow one mode to primarily be present within the waveguide. The waveguide bend radius is 5, the smallest bend radius practical for transverse electric (TE) polarization. The waveguide mode profile is shown in Fig. 2. We show a wavelength sweep in Fig. 3. Through the obtained data, the compact model was calculated:

$$n_{eff(undoped)}(\lambda) = 2.56862 - 0.729403(\lambda - 1.55) + 0.0995556(\lambda - 1.55)^2 \quad (12)$$

We show the fit in Fig. 7 (Fig. 4).

3.2 Interferometer

The transmission spectrum for each interferometer port is shown in Fig. 5. The purpose of the transmission spectrum is to display the proportion of light intensity that

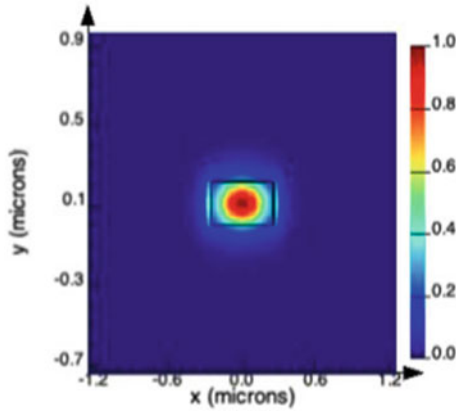


Fig. 2 Waveguide mode profile for quasi-TE polarized light

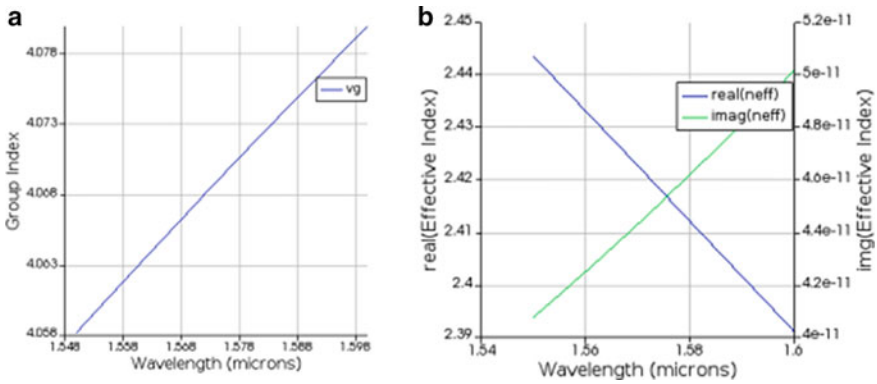


Fig. 3 **a** Group index versus wavelength. **b** Effective index versus wavelength

the device transmits versus the varying wavelengths. The maximum light intensity occurs between the wavelengths 1550 and 1560 nm.

3.3 Overall Circuit

The loss versus wavelength graph for the overall circuit is shown in Fig. 6. The circuit receives the least amount of loss between the 1520 and 1540 nm interval, where both tails of the graph are decreasing.

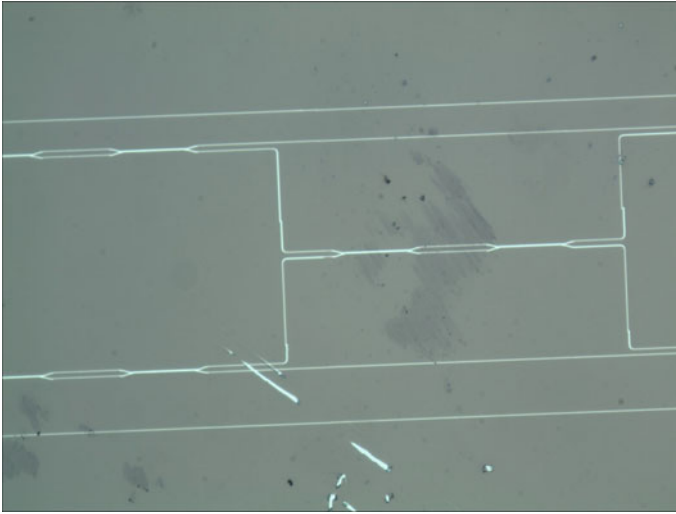


Fig. 4 SEM image of the fabricated chip. Retrieved after fabrication at the University of Maryland. Note the slight misalignment between waveguide and directional coupler components (attributed to slight fabrication error)

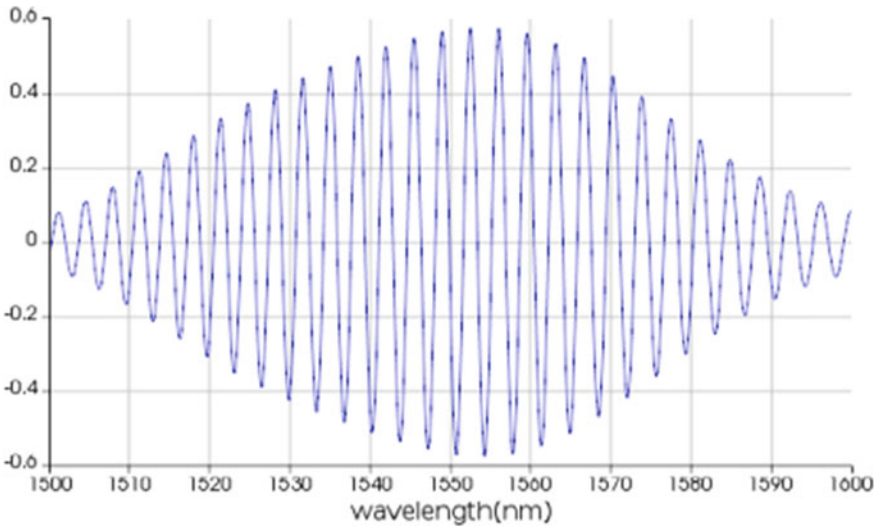


Fig. 5 Transmission spectrum for each port of the interferometer

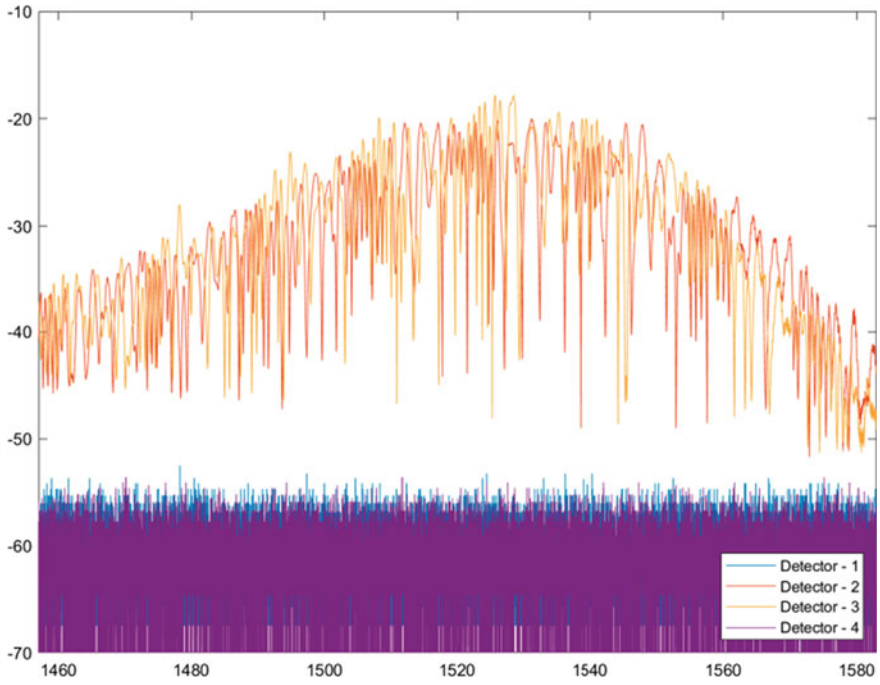


Fig. 6 Loss at various wavelengths for the overall matrix-vector multiplication circuit

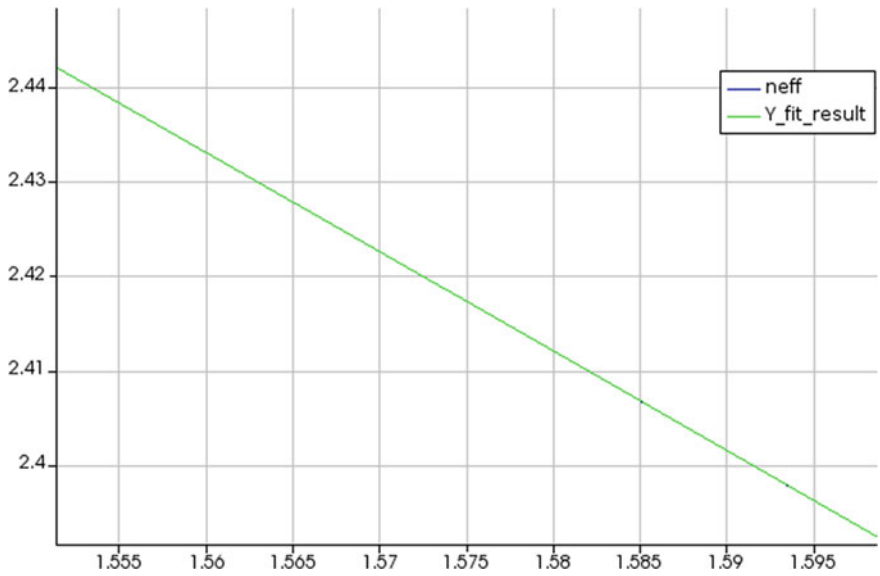


Fig. 7 Waveguide compact model fit to frequency sweep data

3.4 Wafer Preparation

The chip was fabricated using a standard complementary metal-oxide-semiconductor (CMOS) manufacturing process at ANT. A Silicon-on-Insulator (SOI) wafer with a silicon thickness of 220 nanometers and a silicon dioxide thickness of 3μ was utilized. The wafer was cleansed using distilled, deionized water and was dried in a hotplate bake. Since the fabrication process warranted a negative feature development, hydrogen silsesquioxane (HSQ) negative resist was used. The wafer was coated with this resist using a Headway resist spinner set at 4000 RPM. After the resist was applied, the wafer was hotplate baked again for 4 min at 80°C .

3.5 Fabrication & Development

The resist was polymerized using a JEOL JBX-6300FS electron beam lithography fabrication machine operated at 100 keV energy and a 8 nA beam current, with a $500 \mu\text{m}$ exposure field size. The resist was developed using a solution of tetramethylammonium hydroxide, and unexposed silicon was removed using a plasma etching process. Cladding oxide was deposited using a plasma-enhanced chemical vapor deposition (PECVD) system. The wafer was diced into $25 \times 25 \text{ mm}$ square dies. The design outlined in this paper was present on one such die. The process was in accordance with the fabrication process of the University of British Columbia [11].

3.6 Testing

The chip was tested using a setup created within the University of British Columbia. An optical single-die probe station of the MapleLeaf make was utilized. The testing process was initialized by first loading the chip onto a loading stage using an automated loading system. Optical alignment was performed between the chip and the fiber array head setup using a series of optical alignment markers that were fabricated onto the wafer dies. The optical alignment markers themselves were comprised of a series of grating couplers and connecting waveguides. The micropositioner system within the probing system then logged the coordinates of key features on the die, which activated the automated test cycle. A series of laser pulses ranging from 1480 to 1560 nm were transmitted through the chip. The recorded on-chip signal losses are reflected in Fig. 5.

A scanning electron microscope (SEM) was also utilized to visually evaluate the chip. It is important to note that the SEM capture was conducted at the University of Maryland on a new fabrication of the same design. When visually evaluating the images procured from the SEM capture, it is clear that fabrication errors were present (Fig. fig:interferspstransmissionspect). Specifically, misalign-

Table 1 Calculated chip statistics

Specification	Value (4×4)	Value (256×256)
Instructions	1 MVM/cycle	1 MVM/cycle
Clock	220 GHz	3.7 GHz
PPW	245 TFLOPS/watt	261 TFLOPS/watt
Performance	7 TFLOPS	478 TFLOPS

ments were present between individual waveguide structures and adjacent directional coupler components imported from the SiEPIC process design kit (PDK). This fabrication error can be attributed to general misalignment present in electron beam lithography machines, alongside the blocked component hierarchy present in the chip design. Such misalignment can be eliminated in the future with a cleaner integration of PDK components inside of CAD software.

The clock speed, performance per watt (PPW), and performance were calculated for both the fabricated chip and a scaled version of the chip capable of processing matrices of dimensions 256×256 . The calculated statistics are displayed in Table 1.

4 Conclusion

We investigate a Mach–Zehnder interferometer (MZI)-based layout capable of performing a 4×4 unitary matrix transformation. We first go over the MZI specifications, theory, and simulation results. Next, we go over the theory and simulation results for the overall circuit. We discuss practical aspects of our circuit, such as our waveguide configuration (strip) and fundamental mode polarization (transverse electric). Our fabricated chip shows that efficient unitary transformations are practical on silicon photonic hardware. Through integration with other setups such singular value decomposition based methods, more complex products can be realized. This will enable a wider array of algorithms to benefit from photonic acceleration.

The chip tapeout and the results of the subsequent tests support the conclusion that silicon photonics based hardware accelerators could be a viable alternative to traditional digital electronics. The calculated statistics indicate a clock speed that is approximately three orders of magnitude higher than that of current state-of-the-art GPU units [12]. The chip’s energy-efficient nature is also indicated by the PPW, which is relatively high when compared to digital electronic hardware accelerators in the status quo. Of course, a chip capable of processing 4×4 matrices is not terribly useful in the world of data processing and AI development. As innovation accelerates, the volume of data is only expected to grow. Still, when contextualized from a standpoint of 256×256 matrices, the chip retains superiority in energy efficiency and processing performance.

5 Discussion

The implications of a hardware accelerator grounded in silicon photonics are profound, particularly when contextualized with the consistent growth in data and AI throughout the world. Many problems have risen in our world due to a lack of compute potential. Genomic sequencing, for example, would undoubtedly benefit from expanded processing power and speed, potentially yielding insight into new medicines. In general, medicine has seen great strides with the development of AI models catered to diagnosis and drug development [13]. When considering autonomous vehicles, it becomes clear that light-speed computing is an advantage due to the split-second nature of driving [14]. The future of AI as a whole is dependent on capable hardware. Traditional digital electronics require exorbitant amounts of electricity at unsustainable rates, often too high to support the innovation of new AI architectures [15]. Bringing about a paradigm change could open society to another century of unrestricted innovation.

References

1. Merritt R (2022) What is a transformer model? <https://blogs.nvidia.com/blog/2022/03/25/what-is-a-transformer-model/>
2. Nvidia GeForce RTX 4090 specs (2022). <https://www.techpowerup.com/gpu-specs/geforce-rtx-4090.c3889>
3. Jouppi NP, Young C, Patil N, Patterson D, Agrawal G, Bajwa R, Bates S, Bhatia S, Boden N, Borchers A, Boyle R, Cantin P-I, Chao C, Clark C, Coriell J, Daley M, Dau M, Dean J, Gelb B, Ghaemmaghami TV, Gottipati R, Gulland W, Hagmann R, Ho CR, Hogberg D, Hu J, Hundt R, Hurt D, Ibarz J, Jaffey A, Jaworski A, Kaplan A, Khaitan H, Koch A, Kumar N, Lacy S, Laudon J, Law J, Le D, Leary C, Liu Z, Lucke K, Lundin A, MacKean G, Maggiore A, Mahony M, Miller K, Nagarajan R, Narayanaswami R, Ni R, Nix K, Norrie T, Omernick M, Penukonda N, Phelps A, Ross J, Ross M, Salek A, Samadiani E, Severn C, Sizikov G, Snellman M, Souter J, Steinberg D, Swing A, Tan M, Thorson G, Tian B, Toma H, Tuttle E, Vasudevan V, Walter R, Wang W, Wilcox E, Yoon, DH (2017) In-datacenter performance analysis of a tensor processing unit. arXiv (2017). <https://doi.org/10.48550/ARXIV.1704.04760>
4. Moore's Law or how overall processing power for computers will double every two years. <http://www.moorelaw.org/>
5. Lee J (2023) U.S. silicon photonics chip startup Ayar Labs gets AI funding boost. Thomson Reuters. <https://www.reuters.com/technology/us-silicon-photonics-chip-startup-ayar-labs-gets-ai-funding-boost-2023-05-24/>
6. Tayo BO (2022) Matrix multiplication for data science (or machine learning). <https://www.kdnuggets.com/2022/11/matrix-multiplication-data-science-machine-learning.html>
7. Fok MP, Prucnal PR, Mach-Zehnder interferometer. <https://www.sciencedirect.com/topics/engineering/mach-zehnder-interferometer>
8. Dita P (2001) Factorization of Unitary Matrices. arXiv. <https://doi.org/10.48550/ARXIV.MATH-PH/0103005> . <https://arxiv.org/abs/math-ph/0103005>
9. Clements WR, Humphreys PC, Metcalf BJ, Kolthammer WS, Walmsley IA (2016) Optimal design for universal multiport interferometers. *Optica* 3(12):1460–1465. <https://doi.org/10.1364/OPTICA.3.001460>
10. Redrouthu S, Athavale R (2022) Tensor Algebra on an Optoelectronic Microchip. arXiv. <https://doi.org/10.48550/ARXIV.2208.06749> . <https://arxiv.org/abs/2208.06749>

11. Khavasi A, Chrostowski L, Lu Z, Bojko R, <https://arxiv.org/pdf/1602.06158.pdf>
12. <https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/nvidia-a100-datasheet.pdf>
13. Maddipatla J (2022) Classaphasia: an ensemble machine learning network to improve aphasia diagnosis and determine severity. In: 2022 IEEE International conference on bioinformatics and biomedicine (BIBM), pp 3270–3273. <https://doi.org/10.1109/BIBM55620.2022.9995002>
14. Burkacky O, Deichmann J, Rott L, Falkenhausen AV (2021) Automotive semiconductors for the autonomous age. McKinsey amp; Company. <https://www.mckinsey.com/industries/industrials-and-electronics/our-insights/automotive-semiconductors-for-the-autonomous-age>
15. Walton J (2021) Graphics card power consumption and efficiency tested. Tom’s Hardware. <https://www.tomshardware.com/features/graphics-card-power-consumption-tested>

NASA Nearest Earth Object Classification Using Quantum Machine Learning: A Survey



Aman Singh Bhogal, Mausmi Sinha, and Pratiksha Meshram

Abstract Quantum computing has emerged as a revolutionary field with the potential to revolutionize computation and solve complex problems that are beyond the capabilities of classical computers. This literature review paper presents a comprehensive examination of the tools, libraries, and algorithms developed in the realm of quantum computing. The review aims to provide an extensive analysis of the current state of the field and explore the advancements made in harnessing quantum phenomena for computational purposes. The literature review begins by delving into the fundamental principles of quantum computing, including the concept of qubits, superposition, entanglement, and quantum parallelism. It then navigates through the developments made by companies like D-Wave, IBM and Google in the field of quantum computing and the libraries developed such as Qbsolv, Qiskit, and Cirq, exploring their features, functionalities, and applications. Additionally, an in-depth exploration of prominent quantum algorithms, including Grover's algorithm, Shor's algorithm, Quantum Bayesian networks, and Quantum Support Vector Machines (SVM), is presented alongside their classical counterparts for comparative analysis. By critically evaluating the existing literature, this review aims to identify the strengths, limitations, and potential applications of these tools and algorithms. Additionally, it sheds light on the significance of utilizing quantum computing in celestial object identification, showcasing the comparative analysis between classical and quantum algorithms for improved risk assessment in space exploration.

Keywords Quantum computing · Literature review · Tools · Libraries · Algorithms · Qubits · Superposition · Entanglement · Qiskit · Cirq · Qbsolv ·

A. S. Bhogal · M. Sinha (✉) · P. Meshram
NMIMS, Mukesh Patel School of Technology Management and Engineering Shirpur, Shirpur,
India

e-mail: sinhatanu2001@gmail.com

A. S. Bhogal
e-mail: bhogalamansingh22@gmail.com

P. Meshram
e-mail: Pratiksha.Meshram@nmims.edu

Grover's algorithm · Shor's algorithm · Quantum Bayesian networks · Quantum SVM · NASA

1 Introduction

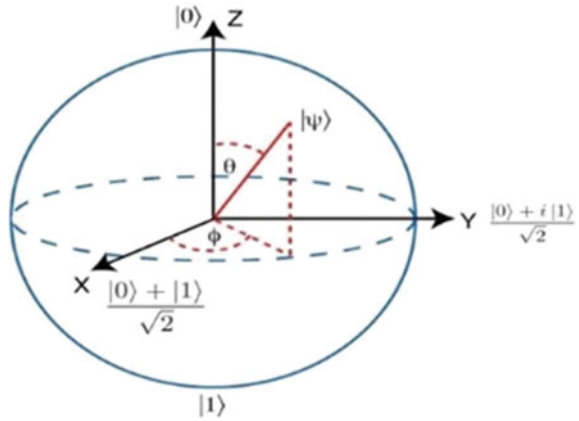
In the vast expanse of outer space, an infinite number of objects exists. Surprisingly, some of these objects are much closer to us than we might imagine. Although we may perceive a distance of 70,000 km as insignificant in terms of potential harm, at an astronomical scale, it is a minute distance that can have significant repercussions on various natural phenomena. These celestial objects, including asteroids, have the potential to pose risks and prove harmful. Therefore, it is prudent to be aware of our surroundings and identify those objects that could potentially endanger us. To facilitate this, we aim to apply Classical Naive Bayes Algorithm and Quantum Naive Bayes Algorithm on provided dataset to do a comparative study on performance. Before we move to the implement part, we have done a thorough literature review on the developments made in the field of quantum computing till now to help us in the process.

Quantum computing term was first coined by Richard Feynman in paper [1]. In this paper, Richard says that "It's true that with a suitable class of quantum machines you could imitate any quantum system, including the physical world." Quantum computing is a field of study and research that explores the use of quantum mechanics principles to perform computational tasks. It is a distinct paradigm of computing that takes advantage of the unique properties of quantum systems, such as superposition, entanglement, and interference, to process and manipulate information. In quantum computing, we use a "quantum bit," or qubit for short. It's like a basic building block that we use to do computations. It's similar to a regular bit that you find in everyday computers, which is the basic piece of information. A bit can have two possible values: 1 or 0. These values represent the "on" and "off" states of the bit.

In quantum computing, a qubit can take three possible states; they are 0, 1 and 0 or 1. In the book [2], Neilsen and Chaung describe quantum mechanical state of a qubit using Bloch sphere picture. The Bloch sphere is a sphere with a center and a surface of unit radius. It helps represent the state of a qubit using a vector that starts from the sphere's center and extends to a point on its surface. The direction of the vector on the sphere shows the state of the qubit. Visualize different qubit operations as rotations of this vector on the Bloch sphere. To better understand the behavior of a qubit, look at Fig. 1. This figure shows that a qubit has two complex numbers, α and β . The equation $|\alpha|^2 + |\beta|^2 = 1$ means that the vector representing the qubit is on the unit plane of the Bloch sphere. This constraint is needed to keep the qubit's state normalized. The formatter will need to create these components, incorporating the applicable criteria that follow.

This paper has been divided into three main sections. The first section focuses on conducting an in-depth literature review in the field of quantum computing through various articles and research papers. Second section focuses on conducting a study

Fig. 1 Chauhan et al. in [3]



on the recent developments in the field of tools and libraries for quantum computing, with a focus on the work done by companies such as IBM, Google, and others. The goal is to review the current state of the art in this area and identify any trends or promising approaches that could be useful for future research.

Third section focuses on various quantum computing algorithms that have been implemented so far, including but not limited to Shor’s algorithm for integer factorization and Grover’s algorithm for unstructured search. Additionally, this section includes a review of the developments made in implementing machine learning algorithms using quantum computing, such as Quantum Support Vector Machine and Quantum Bayesian Network. We will conduct this review by analyzing various research papers in this domain.

2 Related Work

According to Steffen’s paper [4], when we work with very small circuits made of only seven atoms, something special happens. These tiny atoms follow different rules than large objects because they follow the rules of quantum mechanics. By studying and manipulating these tiny atoms, we can explore new ways to design things. We can use systems that involve the levels of energy in the atoms or the way the atoms spin instead of regular circuits. In brief, the paper suggests that working with extremely small things like atoms opens up exciting opportunities to do things differently and potentially achieve new advancements in technology and design.

Again in reference to Steffen’s paper [4], quantum physics offers an intriguing foundation for achieving computational power to address specific categories of problems that are otherwise impossible to solve through traditional machine computation. In this document, we provide a brief overview of both current theoretical and experimental works in the emerging field of quantum computing. Although the implementation of a functioning quantum computer presents significant scientific and

technological challenges, the current pace of progress suggests that such challenges will be substantially addressed within the next decade. We present a brief outline of a quantum computing system and its integration with machine learning through various algorithms. Additionally, we will discuss tools and libraries developed by renowned companies, followed by a discussion of some algorithms.

3 Quantum Computing and Quantum Machine Learning

Quantum computing is a new approach to computation that takes advantage of the peculiar behavior of quantum physics. It utilizes three fundamental properties of quantum physics: superposition, interference, and entanglement. In addition to these properties, it allows for quantum parallelism, which makes computations exponentially faster. Classical computing is outperformed by quantum computing due to its inherent parallelism. The concept of a quantum Turing machine was subsequently developed, and D. Deutsch introduced the notion of quantum parallelism in [5].

Superposition is a fascinating quantum phenomenon where a quantum system can exist simultaneously in multiple states. In quantum computation, classical bits 0 and 1, representing basic states, are denoted as $|0\rangle$ and $|1\rangle$, respectively. Quantum states are represented using shorthand notation called “ket” vectors, enclosed within the symbol $| \rangle$. Ket vectors are column vectors, containing a single column with multiple rows. Apart from ket vectors, there are also “bra” vectors denoted by symbols consisting of a left angle bracket and a vertical line, such as $\langle |$. Bra vectors, on the other hand, are row vectors, containing one row but multiple columns (Fig. 2).

In order to represent single-qubit quantum states, we use a basic set of two vectors. This set of vectors is called a basis, and it can represent the entire vector space. This means that any vector within the state space can be expressed as a combination of these vectors using linear operations.

Qubits, unlike classical bits, can exist in a superposition of $|0\rangle$ and $|1\rangle$, which can be represented as a combination like $a|0\rangle + b|1\rangle$. In this case, a and b are complex numbers that satisfy the condition $|a|^2 + |b|^2 = 1$. In this case, a and b are complex numbers that satisfy the condition $|a|^2 + |b|^2 = 1$. When this superposition is measured using the basis $\{|0\rangle, |1\rangle\}$, the probability of measuring $|0\rangle$ is $|a|^2$ and the probability of measuring $|1\rangle$ is $|b|^2$. This is similar to the scenario of measuring photon polarization.

Fig. 2 Ket and bra vector

$$|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{"Ket Vector"}$$

$$\langle 1| = \begin{bmatrix} 0 & 1 \end{bmatrix} \quad \text{"Bra Vector"}$$

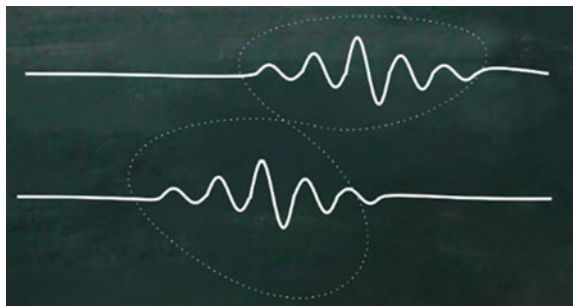
According to Rieffel in [6], if we have a classical function f with m input bits and k output bits, we can implement it on a quantum computer by assuming the existence of a quantum gate array U_f that can carry out the function f . The U_f gate array is a transformation of $m + k$ bits and takes the form $U_f: |x, y\rangle \rightarrow |x, y \oplus f(x)\rangle$, where \oplus represents the bitwise exclusive-OR operation. These U_f gate arrays are unitary, regardless of the function f . To compute the value of $f(x)$, we apply the U_f gate array to the input state $|x\rangle$ combined with k zeros, resulting in $|x, 0\rangle$. By using the property that $f(x) \oplus f(x)$ equals 0, we can conclude that $U_f U_f$ equals the identity matrix I .

When the quantum gate array U_f is applied to a qubit input that is in superposition state, the outcome is both straightforward and impactful. As U_f is a linear transformation, it operates on all the basis vectors within the superposition concurrently, generating a superposition of the resulting states. This extraordinary capability allows for the calculation of $f(x)$ for multiple x values in a single application of U_f . This phenomenon, referred to as quantum parallelism, enables computations on different inputs to occur simultaneously, leading to exponential improvements in computational efficiency compared to classical computing.

Quantum algorithms are powerful because they utilize the benefits of quantum parallelism and entanglement. Quantum entanglement is a fascinating phenomenon in quantum mechanics that holds significant importance. It refers to the ability of two or more quantum particles to become intricately connected, forming a unified system in which the quantum states of each individual particle cannot be independently defined without considering the quantum states of the other particles. This interdependence implies that any operation or manipulation applied to one particle is inherently associated with the other particles, even if they are physically separated [7]. The entanglement of qubits in quantum computing establishes a strong correlation between them. This correlation is such that measuring the state of one qubit instantly provides information about the state of the other, regardless of the distance between them. This property is essential in quantum computing as it enables efficient communication and computation by exploiting the entangled relationship between qubits (Fig. 3).

Quantum interference allows us to guide quantum systems toward their intended state. This involves creating an interference pattern where paths leading to incorrect

Fig. 3 Entanglement



outcomes are canceled out through destructive interference, while paths leading to the correct outcome are reinforced through constructive interference.

The immense power of quantum computing does not come from its processing speed, which is relatively slow, or its limited memory capacity of just a few quantum bits. Rather, its strength lies in the potential of the algorithms it enables. These algorithms have unique complexity traits that distinguish them from their classical counterparts. To understand the importance of this distinction, let's briefly examine complexity theory, which studies the computational resources required to execute an algorithm.

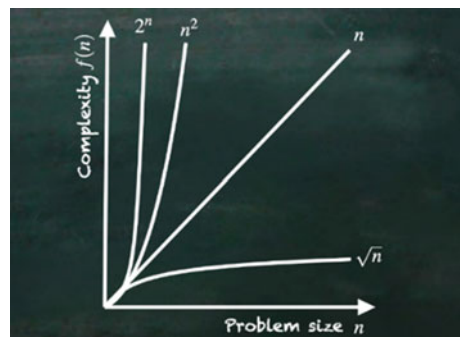
As an example, the computational complexity of addition is represented as $O(n)$, indicating that the effort required to add two numbers grows linearly with the size, which is determined by the number of digits in each number. In contrast, the computational complexity of multiplication is represented as $O(N^2)$, which means that the effort required increases by the square of the number size. These algorithms are considered solvable in polynomial time. Although our smartphones can multiply numbers with 800 digits within a few seconds, the factorization of such numbers takes approximately 2000 years when performed on a supercomputer (Fig. 4).

An efficient quantum algorithm, such as Shor's algorithm, uses superposition to simultaneously evaluate multiple potential factors of a number. Rather than directly computing the result, it utilizes interference to combine all possible solutions in a way that yields the correct answer. This algorithm effectively addresses factorization problems with a complexity denoted as $O((\log * n)^2(\log \log * n)(\log \log \log * n))$.

In other words, a quantum machine would need only polynomial resources in terms of time, memory, and space to simulate other quantum systems. In contrast, classical machines encounter an exponential resource requirement as the problems being simulated grow larger [8, 9]. Quantum computing is a powerful tool as it has the potential to solve specific mathematical calculations with reduced complexity.

The emergence of a new field known as Quantum Machine Learning (QML) has been facilitated by the development of quantum algorithms for machine learning. Machine learning problems can be broadly categorized into two key tasks: effectively handling extensive datasets and devising algorithms capable of processing such data

Fig. 4 Graph of common complexity functions



with utmost speed. In terms of the former task, quantum registers possess the ability to manage substantially larger volumes of data in comparison with classical registers.

A classical bit register of size n has the capacity to store only an n -size binary string. In contrast, an n -qubit register can store a binary string of size $2n$ by leveraging amplitude-based encoding. Although the extraction of all these strings presents a challenge due to state collapse upon measurement, resulting in obtaining only a single amplitude or string, these qubits possess inherent parallelism. Consequently, it becomes feasible to develop algorithms that can concurrently operate on all $2n$ strings, leading to an exponential acceleration when compared to classical algorithms.

Quantum Machine Learning (QML) represents the convergence of quantum computing and machine learning, resulting in a multidisciplinary field. This integration can be accomplished through four distinct approaches:

1. **Classical-Classical approach (CC):** This approach involves the utilization of classical algorithms that incorporate principles derived from quantum mechanics, quantum processing, or quantum information. Known as “quantum-inspired” algorithms, they are employed on classical data and run on classical computers.
2. **Classical-Quantum approach (CQ):** This approach involves the utilization of Quantum Machine Learning algorithms to handle classical data, aiming to enhance the efficiency of machine learning tasks. The goal is to identify quantum variations of classical machine learning algorithms or existing quantum algorithms that can effectively tackle machine learning challenges and deliver quantum advantages over classical counterparts.
3. **Quantum-Classical approach (QC):** This involves the application of classical machine learning methods and algorithms to process quantum data. This enables quantum computers to learn from and extract valuable insights from the distinct characteristics of quantum data.
4. **Quantum-Quantum approach (QQ):** The Quantum-Quantum approach is centered around using quantum algorithms and quantum data. In this approach, Quantum Machine Learning (QML) algorithms are utilized to manipulate quantum states, revealing hidden patterns and extracting valuable insights from the quantum data. While the CQ and CC approaches have received more attention and extensive exploration in the field of QML, the QQ approach is also of great importance [10].

Although QML is still in its early stages of development, it holds considerable promise for handling larger datasets and may eventually establish itself as a viable alternative to classical machine learning.

4 Tools and Libraries

See Table 1.

Table 1 Developments made by various companies in quantum computing

Parameters	D-Wave	IBM	Google
Business	Leading company to produce commercial quantum computing systems, software, and services	Offers quantum systems, software, and services for quantum computing	Demonstrated practical realization of quantum computing with NISQ systems
Technological achievement	Known for quantum annealing approach and quantum systems like advantage	Developed processors such as eagle, osprey, and planned condor with increasing qubit counts	Achieved breakthroughs with processors like sycamore, performing tasks in seconds that would take classical computers thousands of years
Focus	Installed over 1000 qubits and focuses on scalability at the quantum artificial lab	Focuses on modular systems and connections between independent processors for scalability	Pursues the development of logical qubits for error-free storage of quantum data
Tools	Qbsolv	Qiskit	Cirq
Motto	Provides cloud services, application development tools, and professional services	Pursuing the development of a quantum-centric supercomputer integrating classical and quantum systems	Emphasizes the potential of quantum computing for tackling complex calculations and surpassing classical computing capabilities
Future Work	Collaborating with customers like temporal defense system, Google/NASA/USRA, Virginia Tech, and the Hume center for defense and intelligence applications	Envisions multi-chip processors like Kookaburra to achieve high qubit counts	Focuses on scalability and commercial readiness of quantum computing
Computational model	Uses quantum annealing computational model	Uses universal gate computational model	Uses universal cate computational model

4.1 D-Wave

D-Wave is one of the leading quantum computing systems, software, and service providing company that is based in Canada. D-Wave builds and delivers quantum systems along with cloud services, application development tools, and professional services to support quantum computing.

Fig. 5 D-wave advantage.
Source [11]



Notably, D-Wave has successfully installed over 1000 qubits at the Quantum Artificial Lab located within the NASA Ames Research Centre. The company introduced its latest quantum computers called Advantage (2020). These new systems boast more than 5000 qubits and offer 15-way qubit connectivity. They are now accessible through D-Wave's Leap cloud computing platform. In comparison, the previous system had around 2000 qubits and featured six-way connectivity. By leveraging Leap's hybrid solver, which combines classical CPUs and GPUs with D-Wave's quantum system, users can now tackle significantly more intricate problems [11].

D-Wave CEO Alan Baratz responded to skeptics who claimed that D-Wave's technology would not scale by highlighting the significant advancements they have made. Baratz explained that despite having more than twice the number of qubits, more than twice the connectivity, and over five times the number of devices on the superconducting chip, D-Wave has successfully maintained the ability to program, read out, and operate the system within the same time frame and temperature (Fig. 5).

D-Wave is launching its third D-Wave Advantage system, which can now be accessed through both D-Wave's Leap quantum cloud and AWS' Amazon Braket service. Additionally, the company has updated its tools to empower enterprises in utilizing quantum annealing for tackling new optimization problems. The D-Wave Advantage system stands out with its impressive 5000+ qubits, positioning it ahead of competitors like IBM, who recently announced their intentions to develop a 4000-qubit system by 2025.

4.2 IBM

According to reference [12], IBM utilized its quantum processor Eagle in 2022, which had the capability to handle 127-qubit processing. They also introduced Qiskit Runtime, a runtime environment that combined classical and quantum systems, designed to execute containerized quantum circuits swiftly and at scale. The initial version of Qiskit Runtime demonstrated an impressive $120\times$ speedup on a quantum

workload designed for research purposes. Later in 2022, IBM introduced Osprey, a 433-qubit processor. Furthermore, IBM had plans to unveil the Condor processor with 1121 qubits in 2023. Looking ahead, IBM envisions the integration of three multi-chip processors, each consisting of 1386 qubits, known as Kookaburra, resulting in a combined total of 4158 qubits.

As per [13], IBM is pursuing the development of a quantum-centric supercomputer, which will revolutionize the computing landscape by integrating quantum processors, classical processors, quantum communication networks, and classical networks into a cohesive system.

IBM acknowledges that scaling the quantum computer system poses a significant challenge. However, their approach diverges from building large-scale quantum computers on a single massive chip. Instead, they are focused on establishing connections between independent processors, creating a modular system that can scale without limitations. To achieve this, IBM has devised three distinct approaches.

In 2023, IBM plans to introduce “Heron,” a 133-qubit processor equipped with control hardware that enables real-time processing. This represents the first step in their journey toward a scalable modular system (Fig. 6).

As part of their subsequent approach, IBM intends to introduce “Crossbill,” a 408-qubit processor composed of three chips interconnected using chip-to-chip couplers. This design enables the seamless implementation of heavy-hex lattices across multiple chips. The launch of “Crossbill” is planned for the future.

In 2024, IBM has outlined their third approach called “Flamingo,” which involves unveiling a 462-qubit processor. Notably, this processor will feature an integrated quantum communication link between the processors, facilitating quantum parallelization. The introduction of “Flamingo” marks another significant milestone in IBM’s quantum computing roadmap.

IBM aims to overcome scaling limitations of quantum processors by 2025 using modular hardware, control electronics, and cryogenic infrastructure. Their objective is to develop a scalable quantum computing system that pushes boundaries and fosters future advancements in the field (Fig. 7).

Fig. 6 133-qubit “Heron” processor, slated for 2023.

Source [13]

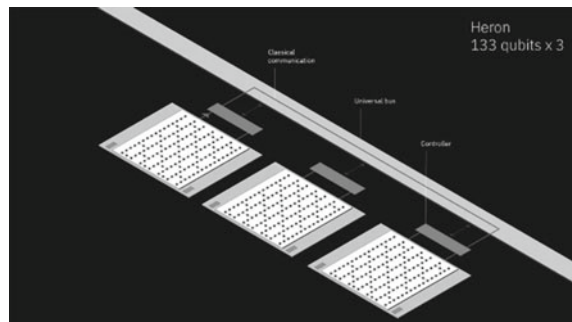
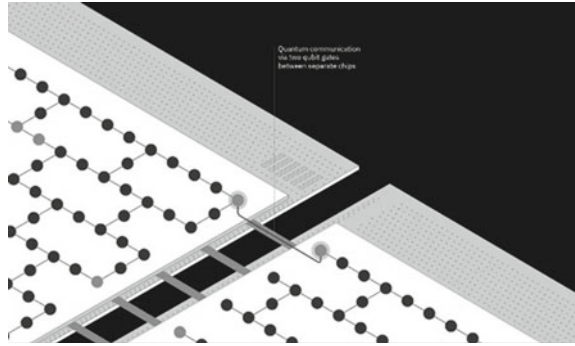


Fig. 7 Quantum communication via two-qubit gates between separate chips. *Source* [13]



4.3 Google

According to [14], in 2019, Google showcased the practical realization of quantum computing, shifting from a theoretical concept to a tangible reality. This marked a significant milestone as humanity entered the era of Noisy, Intermediate-Scale Quantum (NISQ) computing, departing from traditional classical computing methods. Google’s initial stage involved conducting an experiment that employed 12 qubits to simulate a basic chemical reaction on a quantum computer. This accomplishment represented a notable advancement for Google in the field of quantum computing (Fig. 8).

Following their initial experiment, Google proceeded to utilize 53 qubits to explore the ballistic propagation of quantum entanglement waves. Through their research, they discovered that by modifying the circuit’s composition, they could effectively

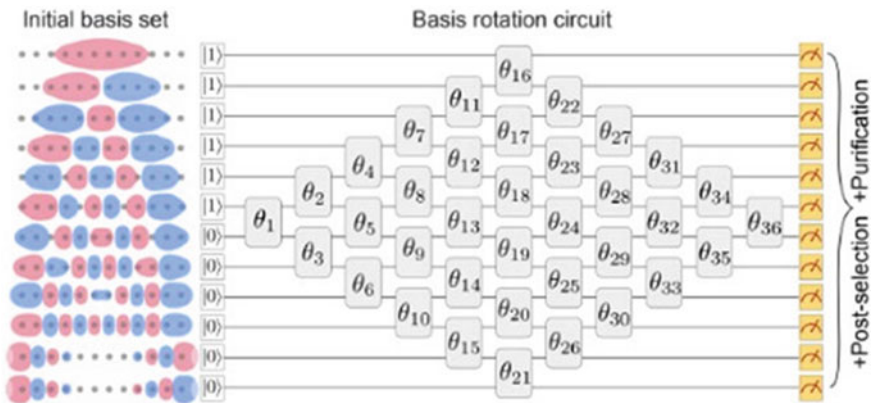


Fig. 8 *Source* Google’s Website

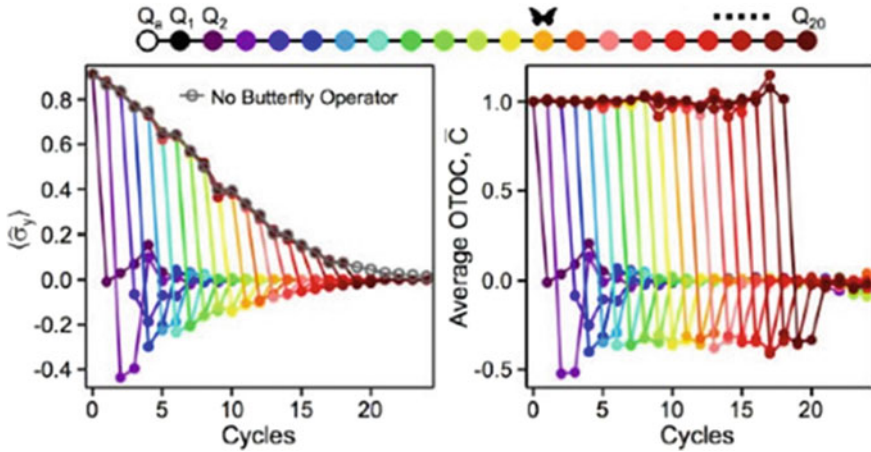


Fig. 9 Graphical results from Google’s Website

control the speed at which entanglement propagation occurs. This finding demonstrated the influence of circuit design on the dynamics of entanglement and further expanded Google’s understanding of quantum phenomena.

Google’s Sycamore quantum processor accomplished a task in 200 s that would have taken 10,000 years on a classical computer. To build a prototype of a logical qubit, Google believes 100 qubits are necessary. They plan to achieve scalability by reducing the error-rate by a factor of 10 with each qubit size increase. A logical qubit will enable error-free storage of quantum data for performing complex calculations. This milestone will mark quantum computing’s equivalent of the transistor moment, demonstrating its readiness for commercial scalability (Fig. 9).

5 Quantum Computing Libraries

5.1 D-Wave’s Qbsolv

D-Wave Qbsolv is a software tool created for leveraging quantum computing to solve optimization problems. Its underlying framework is based on the Quadratic Unconstrained Binary Optimization (QUBO) model, which mathematically represents optimization problems using binary variables.

A notable advantage of D-Wave Qbsolv is its superior efficiency in tackling intricate optimization problems compared to classical computers. This efficiency stems from its utilization of quantum annealing, a process that enables simultaneous exploration of an extensive search space containing potential solutions. Qbsolv employs decomposition solving techniques such as Divide and Conquer and Dynamic

Programming. These algorithms have a long-standing history in computer science, particularly for addressing problems with a substantial number of variables.

Many challenging problems that stand to benefit from quantum computing are too large to be directly mapped onto a Quantum Processing Unit (QPU). To overcome this limitation, Qbsolv employs a strategy where it divides the problem into smaller subproblems, solves these subproblems, and then reconstructs the original problem's solution using the solutions obtained from the subproblems.

D-Wave Qbsolv offers two interfaces:

- **Command Line Interface (CLI):** The CLI executes the tabu algorithm on the problem, which is partitioned into multiple subproblems, each containing several dozen variables.
- **Python Interface:** The Python interface includes a Qbsolv class wrapper for the qbsolv C code. Additionally, it allows for the substitution of the default tabu algorithm with a dimod sampler.

5.2 IBM's Qiskit

Qiskit is a comprehensive open-source software library for quantum computing that encompasses the entire stack, starting from interacting with IBM Q hardware, through simulation and emulation, and all the way to application-level algorithms. According to [15], the tool is organized into four libraries named after the classical elements: terra, aqua, aer, and ignis. Collectively, these four Qiskit libraries offer a complete suite of software solutions for quantum computing, seamlessly interconnected and communicating through shared data structure constructs.

Qiskit library [8] can be used for

- **Implementing Quantum Circuits:** Qiskit offers an extensive collection of quantum gates and a diverse range of pre-designed circuits, making it accessible to users of all proficiency levels for research and the development of applications.
- **As Transpiler:** The Qiskit transpiler facilitates the translation of Qiskit code into an optimized circuit that aligns with the native gate set of a specific back end. This enables users to program for a wide range of quantum processors. With Qiskit, users have the flexibility to transpile their code using the default optimization provided, configure a custom optimization, or even develop their own plugin to tailor the transpilation process according to their specific requirements.
- **Any Hardware:** Qiskit assists users in scheduling and executing quantum programs on diverse local simulators and cloud-based quantum processors. It provides support for various quantum hardware designs, including superconducting qubits and trapped ions, enabling users to explore and utilize different types of quantum technologies.

Qiskit empowers users to build quantum programs and execute them on simulators or actual quantum computers. Through its expansive network of providers, Qiskit offers the capability to compile your Qiskit code for an extensive array of diverse

back ends, surpassing the range provided by any other quantum framework available. It supports back ends like Amazon Braket, Azure Quantum, IBM Quantum, NVIDIA cuStateVec, QC Ware Forge, etc.

5.3 *Google's Cirq*

In 2022 [16], Google released first full version of the open-source quantum programming framework Cirq: Cirq 1.0. Cirq serves as an open-source platform specifically designed to facilitate the development of Noisy Intermediate-Scale Quantum (NISQ) algorithms. Coined by John Preskill [17], the NISQ era represents the current state of quantum computers, which necessitate error correction and feature a limited number of qubits. By offering access to Cirq, individuals are empowered to engage in their own research endeavors, exploring and investigating near-term applications tailored to the capabilities of quantum computers. This amalgamation of Cirq's capabilities and its availability fosters a collaborative environment for advancing the field of quantum computing.

The primary features offered by Cirq are

- **Circuits:** At the heart of Cirq lies the fundamental capability to build quantum circuits, encompassing the essential methods and data structures required for this purpose.
- **Devices:** Cirq offers interfaces that facilitate the execution of circuits on a wide range of quantum hardware provided by various services. It enables users to define the specific characteristics and constraints of quantum hardware devices, thereby supporting the seamless running of circuits on these devices.
- **Simulation:** Cirq includes pre-installed Python simulators that are ideal for testing small circuits. These simulators encompass two primary types of simulations supported by Cirq: pure state and mixed state. This enables users to efficiently evaluate and validate their circuits within a controlled and simulated environment.

6 Quantum Computing Algorithms

Alongside advancements in quantum hardware, quantum algorithms have made tremendous progress.

6.1 *Shor's Algorithm*

The field of quantum computing gained momentum in 1994 when Peter Shor presented a breakthrough quantum algorithm capable of polynomial-time factoring

of n -digit integers [18]. This propelled quantum computing to the forefront of scientific exploration, surpassing the previously known exponential Lenstra and Lenstra algorithm from 1993 for integer factorization.

Many factoring algorithms, including Shor's algorithm, employ a common approach of reducing the factoring problem to determining the period of a function [6]. Shor leverages quantum parallelism to compute the quantum Fourier transform of the function, concentrating the amplitudes around multiples of the reciprocal of the period. Measuring the state provides a high probability of obtaining the period, enabling the factorization of the integer M .

6.2 Grover's Algorithm

In 1996 [3], Indian-American computer scientist Lov Kumar Grover devised a quantum computing database search algorithm. Grover's algorithm is an unstructured search that efficiently finds a unique input function with high probability, generating an output value from a database of " N " entries. It achieves a time complexity of $O(\sqrt{N})$ and requires $O(\log N)$ storage space for " N " entries [19]. Unlike traditional linear algorithms, Grover's algorithm can search an unsorted database in $O(\sqrt{N})$ time complexity, representing a significant improvement.

Grover has extended the scope of his algorithm beyond search problems, attaining quadratic speedup for non-search tasks like calculating the mean and median of a function [19]. By employing similar techniques, Grover has demonstrated that specific search problems, which typically require $O(\log N)$ time complexity in classical computing, can be solved in constant time complexity ($O(1)$) on a quantum computer.

Additionally, Grover's search algorithm can be utilized as a subroutine in various other quantum computations. Biron et al. [19] have demonstrated the applicability of this technique with arbitrary initial amplitude distributions, while still maintaining a complexity of $O(\sqrt{N})$.

6.3 Quantum Support Vector Machine

Support Vector Machines (SVMs) have gained significant popularity in addressing classification tasks. Classification is a fundamental concept in supervised learning, aiming to predict the most probable pre-defined class for newly encountered data. Classification can be divided into binary or multi-class classification based on the number of pre-defined classes. SVM functions by identifying an optimal hyperplane that can successfully separate the classes. The hyperplane primarily emphasizes the data points situated at the boundary, known as support vectors.

There exist two implementations of QSVM [10], each employing a different approach. The first approach utilizes Grover's search, conducting a maximum search

on all generated potential solutions to find the hyperplane. This method achieves a quadratic speedup. The second approach, which offers exponential speedup, employs the HHL algorithm. It transforms the original quadratic programming problem into a linear system of equations problem, subsequently solved using HHL. Various enhancements to QSVM have been suggested in literature, with quantum versions providing exponential speedup over classical SVM.

In [20], an implementation of QSVM on noisy intermediate-scale quantum computers (NISQ) is suggested. It utilizes an optimized HHL quantum circuit with reduced depth, yielding more accurate results on current NISQ systems. Despite limitations like its applicability to linearly separable datasets, this implementation demonstrates the efficient execution of QML algorithms like SVM on NISQ computers. Additionally, a demonstration of QSVM on a four-bit quantum processor for optical character recognition training with characters “6” and “9” is showcased. Additionally, [21] discusses hybrid quantum–classical versions of SVM, suggesting that leveraging classical intractable quantum feature maps can address difficult machine learning tasks and enhance performance.

6.4 *Quantum Bayesian Network*

Bayesian Networks [9], also referred to as Bayesian belief networks, are powerful probabilistic graphical models employed for representing knowledge within uncertain domains. These networks adopt a directed acyclic graph structure comprising nodes and edges, where nodes symbolize random variables and edges denote probabilistic dependencies between them. Bayesian networks facilitate two main types of analysis: forward analysis, enabling probabilistic predictions of lower-level nodes using probability distributions of higher-level nodes, and inverse analysis, inferring values of higher-level nodes based on data from lower-level nodes.

In his work, primarily conducted by Tucci in 1995 [22], Quantum Bayesian Networks were introduced as a quantum analog of classical Bayesian networks. The goal was to enhance algorithm performance by leveraging quantum capabilities compared to traditional Bayesian probabilistic models. Quantum Bayesian Networks incorporate advanced transformation gates specifically designed to operate on qubits, enabling quantum-level processing and analysis. These transformation gates play a crucial role in manipulating and harnessing the power of quantum states within the network. They enable the representation and analysis of probabilistic relationships among quantum variables, making them valuable tools in quantum information science for modeling and reasoning about quantum systems with probabilistic behavior.

In a QBN, the nodes correspond to quantum variables, which can represent quantum systems or quantum measurements. The edges in the network signify the probabilistic dependencies between these quantum variables. These dependencies are quantified using quantum conditional probability distributions, providing a mathematical description of the probabilistic relationships within the network.

7 Research Methodology

This literature review paper on quantum computing aims to provide a comprehensive analysis of existing research and developments in the field. The initial stage of the research involved clearly defining the objectives of the literature review. Extensive literature search was conducted using various academic databases, such as IEEE Xplore, Springer Library, and ScienceDirect, among others. The retrieved literature was screened based on relevance to the research objectives. Emphasis was placed on recent publications to ensure the review reflected the current state of the field. The data extracted included details about the tools and libraries available for quantum computing. The extracted data was synthesized and organized to present a coherent overview of the current landscape of quantum computing. The findings were categorized based on the different tools, libraries, and algorithms discussed, enabling a comparative analysis of their features, advantages, and potential applications.

8 Conclusion

This literature review paper on quantum computing aims to provide a comprehensive analysis of existing research and developments in the field. The initial stage of the research involved clearly defining the objectives of the literature review. Extensive literature search was conducted using various academic databases, such as IEEE Xplore, Springer Library, and ScienceDirect, among others. By applying Classical Naive Bayes Algorithm and Quantum Naive Bayes Algorithm to the NASA—Nearest Earth Objects dataset, we aim to evaluate their performance in classifying objects of potential risk to Earth. Emphasis was placed on recent publications to ensure the review reflected the current state of the field. Key information, findings, and insights were extracted from the selected studies. The data extracted included details about the tools and libraries available for quantum computing. The extracted data was synthesized and organized to present a coherent overview of the current landscape of quantum computing.

References

1. Deutsch D (1985) Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc R Soc Lond A, Math Phys Sci* 400(1818):97–117
2. Nielsen M, Chuang I (2010) *Quantum computation and quantum information: 10th anniversary edition*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511976667>
3. Chauhan V, Negi S, Jain D, Singh P, Sagar AK, Sharma AK (2022) Quantum computers: a review on how quantum computing can boom AI. In: 2022 2nd international conference on advance computing and innovative technologies in engineering (ICACITE), Greater Noida, India, pp 559–563. <https://doi.org/10.1109/ICACITE53722.2022.9823619>

4. Steffen M, DiVincenzo DP, Chow JM, Theis TN, Ketchen MB (2011) Quantum computing: an IBM perspective. *IBM J Res Develop* 55(5):13:1–13:11. <https://doi.org/10.1147/JRD.2011.2165678>
5. Shor PW (1994) Algorithms for quantum computation: discrete logarithms and factoring. In: *Proceedings 35th annual symposium on foundations of computer science*, pp 124–134
6. Yanofsky N (2007) An introduction to quantum computing. https://doi.org/10.1007/978-94-007-0080-2_10
7. Horodecki R, Horodecki P, Horodecki M, Horodecki K (2009) Quantum entanglement. *Rev Mod Phys* 81(2):865–942. <https://doi.org/10.1103/revmodphys.81.865>
8. Qiskit Documentation. <https://qiskit.org/>
9. Borujeni SE, Nannapaneni S, Nguyen NH, Behrman EC, Steck JE (2021) Quantum circuit representation of Bayesian networks. *Expert Syst Appl* 176:114768. ISSN 0957-4174. <https://doi.org/10.1016/j.eswa.2021.114768>
10. Jadhav A, Rasool A, Gyanchandani M, Quantum machine learning: scope for real-world problems. *Proc Comput Sci*. <https://doi.org/10.1016/j.procs.2023.01.235>
11. Lardini F, D-wave launches its 5000+ qubit advantage system. <https://techcrunch.com/2020/09/29/d-wave-launches-its-5000-qubit-advantage-system/>
12. Cooney M, IBM wants a 4,000 qubit quantum computer by 2025. <https://www.networkworld.com/article/3659886/ibm-wants-a-4-000-qubit-quantum-computer-by-2025.html>
13. Gambetta J, Expanding the IBM quantum roadmap to anticipate the future of quantum-centric supercomputing. <https://research.ibm.com/blog/ibm-quantum-roadmap-2025>
14. Google’s Website. Our quantum computing journey. <https://quantumai.google/learn/map>
15. Wille R, Van Meter R, Naveh Y (2019) IBM’s Qiskit tool chain: working with and developing for real quantum computers. In: *2019 design, automation & test in Europe conference & exhibition (DATE)*, Florence, Italy, pp 1234–1240. <https://doi.org/10.23919/DATE.2019.8715261>
16. Bacon D, Broughton M, Cirq Turns 1.0. <https://opensource.googleblog.com/2022/07/Cirq-Turns-1.0.html>
17. Bassan T, Quantum programming with Google Cirq. <https://tanishabassan.medium.com/quantum-programming-with-google-cirq-3209805279bc>
18. Grover LK (1998) A framework for fast quantum mechanical algorithms. In: *Proceedings of the thirtieth annual ACM symposium on theory of computing*
19. Rieffel E, Polak W (2000) An introduction to quantum computing for non-physicists. *ACM Comput Surv* 32(3):300–335. <https://doi.org/10.1145/367701.367709>
20. Yang J, Awan AJ, Vall-Llosera G (2019) Support vector machines on noisy intermediate scale quantum computers
21. Lin J, Zhang D-B, Zhang S, Li T, Wang X, Bao W-S, Quantum-enhanced least-square support vector machine: simplified quantum algorithm and sparse solutions. <https://doi.org/10.1016/j.physleta.2020.126590>
22. Harikrishnakumar R, Borujeni S, Dand A, Nannapaneni S (2020) A quantum Bayesian approach for bike sharing demand prediction, 2401–2409. <https://doi.org/10.1109/BigData50022.2020.9378271>

Studying the Effect of Type of Surface Passivation Layer on Performance Parameters of AlGa_N MSM Detector



Harpreet Kaur and Manish Kumar Hooda

Abstract In past studies, it is reported that surface passivation helps to minimize surface leakage current in case of photodetectors. However, utilization of stack of passivation layers for Metal–Semiconductor–Metal (MSM) type detectors is studied in very few reports. In present study, we have investigated the effect of single passivation and stack of passivation layers on the performance parameters of AlGa_N MSM detector as Ga_N or AlGa_N layer exhibits higher surface state density, for three types of passivation layers and without passivation case also. Passivation layers like SiO₂, Al₂O₃, and Al₂O₃/SiO₂ (stack) have been utilized for investigation analysis. The comparison of performance parameters of passivated detectors and without passivation-based detector has been performed so that suitable selection of passivation can be done for enhanced performance. The considered performance parameters of detector are dark current density, electric field, recombination rate, and photocurrent-to-dark current density ratio. Comparison analysis shows that detector with Al₂O₃/SiO₂ passivation stack provides lowest dark current density of $5e-08$ A/cm² at top surface of AlGa_N layer as compared to other three types of detectors. Moreover, highest electric field, highest photocurrent-to-dark current density ratio, and lower recombination rate have been obtained with Al₂O₃/SiO₂ passivated stack-based detector. Present work helps to select the type of passivation for AlGa_N MSM detector so as to provide improved performance parameters and reliable UV detection for low noise applications.

Keywords MSM · Dark current density · Recombination rate · Passivation · AlGa_N · Performance

H. Kaur (✉)

Chitkara University School of Engineering and Technology, Chitkara University, Solan, Himachal Pradesh 174 103, India

e-mail: harpreet.kaur@chitkarauniversity.edu.in

M. K. Hooda

Semi-Conductor Laboratory, S.A.S Nagar, Mohali 160 071, India

e-mail: manishk@scl.gov.in

1 Introduction

The MSM detector structure is preferred among other types of detector structures (like, p - n , PIN and Schottky barrier) due to easy fabrication processes and planar design flexibility [1, 2]. The performance of photodetector is highly dependent on the material properties of photon-absorbing semiconductor layer and geometry parameters [3].

AlGa N or Ga N MSM detectors have become a promising solution for providing very low dark current/density due to wide band gap. These types of detectors are capable of providing high responsivity or photocurrent and thermally stable operation. Schottky contacts on AlGa N layer of MSM detector provide low dark current/density and low capacitance due to rectifying nature [2, 4]. MSM photodiodes have intrinsic ability of providing higher UV/visible ratio and linear optical power response [5].

The influence of surface passivation of AlGa N /Ga N detectors has been analyzed by authors and demonstrated an improvement in various device performance parameters and characteristics [6–8]. Ga N or Al $_x$ Ga $_{1-x}$ N material layer exhibits high surface density states; thus, surface passivation with insulating materials like SiO $_2$, Al $_2$ O $_3$, and SiN $_x$ helps to reduce undesirable effects of surface states and saturation current [9, 10].

For efficient UV light detection, photodetector with low dark current or dark current density is required so as to enhance UV light to dark current ratio and noise. Therefore, surface passivation and anti-reflecting coatings are generally required for photodetectors. In this paper, effect of different types of passivation layers and utilization of passivation stack on the performance parameters of detector has been studied.

This section is followed by device structure description and simulation methodology heading. Afterward, results and discussions have discussed. Finally, conclusion of present research work is presented.

2 Device Structure Description and Simulation Methodology

The geometry of AlGa N MSM structure has been designed using Silvaco Atlas tool. Both x and y dimensions of selected MSM structure are 50 μ m as shown in Fig. 1. The thickness of both Al $_{0.5}$ Ga $_{0.5}$ N and AlN layers is taken 500 nm. AlN layer is used as buffer layer between AlGa N and Sapphire substrate to minimize the crystal constants mismatch [11]. The Schottky contacts on AlGa N layer have been deposited by utilizing gold metal electrodes for getting higher Schottky Barrier Height (SBH). Moreover gold (Au) provides best transmittance in UV range as compared to platinum, nickel, etc. [12]. The thickness of four metal fingers or Schottky contacts is 20 nm. The width (W) and spacing (S) between electrodes are taken as 5 and 10 μ m.

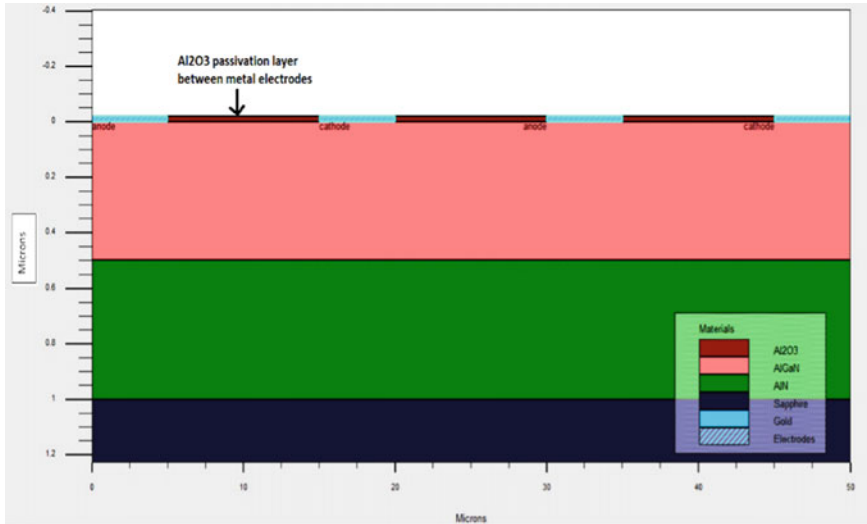


Fig. 1 Al_{0.5}Ga_{0.5}N/AlN/Sapphire MSM detector with Al₂O₃ surface passivation layer

The spacing between metal electrodes is taken twice so as to minimize reflection of incoming light and to absorb higher amount of incident light [13, 14]. However, transit time of photo-generated carriers can be slightly affected which could be overcome by applying high Reverse Bias voltage [4].

Table 1 shows doping levels of semiconductor layers and Sapphire substrate of proposed Al_{0.5}Ga_{0.5}N/AlN/Sapphire detectors.

The geometry parameters are taken same for all MSM detectors considered under investigation. These parameters include device dimensions, thickness of all layers, doping levels, number of metal electrodes (fingers), and their dimensions. In Fig. 1, dark blue color layer is representing Sapphire, Green color layer is AlN, and peach colored layer is of AlGaIn material that is known photo-absorbing layer. That’s why detector under investigation is written AlGaIn/AlN/Sapphire type MSM detector. On the top of AlGaIn layer, four sky blue colored metal electrodes are shown. In between these electrodes, Al₂O₃ passivation layer of red color is shown. The thickness of passivation layer is considered same as the thickness of metal electrodes.

In Fig. 2, the device dimensions and name of all structure layers are same as specified in Fig. 1. However, in between metal electrodes, Al₂O₃/SiO₂ surface passivation stack is shown. In this stack, SiO₂ passivation can be seen by sky blue color and

Table 1 Doping levels of AlGaIn/AlN/Sapphire layer for all selected detector MSM structures

Type of semiconductor layer	Doping level (per cm ³)
AlGaIn (<i>n</i> -type)	10 ¹⁹
AlN	10 ¹²
Sapphire	10 ¹²

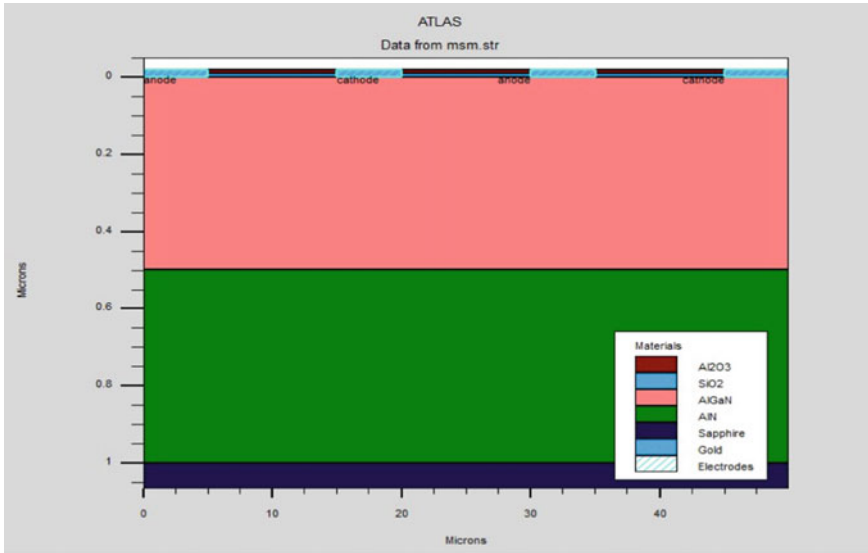


Fig. 2 $\text{Al}_{0.5}\text{Ga}_{0.5}\text{N}/\text{AlN}/\text{Sapphire}$ MSM detector with $\text{Al}_2\text{O}_3/\text{SiO}_2$ passivation stack

on it Al_2O_3 passivation of red color can be seen. The thickness of both the passivation layers in stack is 10 nm each. The metal electrodes are always interconnected to be termed as anode and cathode terminals. Then, reverse bias voltage is applied to both the electrodes so that photodetector converts incident light energy to electrical current. Thus, transport of current in metal–semiconductor contact is due to the movement of majority carriers [15].

Previous studies show that with the use of passivation layer, performance parameters of detector can be enhanced [16]. However, detector surface passivation using passivation layers stack can demonstrate more excellent improvement in performance as compared to single layer passivation [17].

Adequate list of material parameters associated to $\text{Al}_{0.5}\text{Ga}_{0.5}\text{N}$, AlN, and Sapphire have been utilized in Atlas code. These include bandgap at 300 K, electron affinity, permittivity, absorption coefficient (wavelength dependent), optical, and lattice constants. The material parameters for $\text{Al}_{0.5}\text{Ga}_{0.5}\text{N}$ and AlN layer are taken from several references [18–20]. Gold Schottky contact with work function 5.1 eV has been defined to get Schottky Barrier Height of 1.8 eV at metal–semiconductor interface. Past studies show that barrier height of Schottky contact gets increased due to surface modification that helps to reduce dark current [21, 22]. Physical models have defined using model and impact statements in the program code. Physical models related to carrier statistics, mobility, recombination, and impact ionization have been defined in the Atlas code. Fermi Dirac statistics have been used so that certain properties of highly doped material can be accounted. Effective density of states related to electrons and holes has been defined for semiconductor materials. Trap density of active material is assumed to be ideal. Default mobility model parameters for

Al_{0.5}Ga_{0.5}N, GaN, and AlN material have been utilized in program code are based on low field and are taken from [23]. The electron–hole pairs get generated in active layer when wavelength of incident light corresponds to band gap of semiconductor layer. The wavelength of incident light for Al_{0.5}Ga_{0.5}N/AlN/Sapphire detector is taken as 270 nm which leads to band gap of 4.4 eV which is mentioned in Atlas-Silvaco program code. The power intensity of UV light is taken 1 W/cm² for all detectors under simulation work.

3 Results and Discussions

Four Al_{0.5}Ga_{0.5}N/AlN/Sapphire MSM detectors with surface passivation layer have been designed and simulated for extracting performance parameters. Thickness of single passivation layer is taken same as that of thickness of metal electrodes. However, for passivation stack, thickness of each passivation is half of metal electrode thickness. First, MSM detector is without passivation layer. Second, detector is passivated with conventional SiO₂ material-based passivation. Third, MSM detector is Al₂O₃ material-based passivation layer and fourth is based on Al₂O₃/SiO₂ passivation stack. The cost of Al₂O₃ has reported to be low that makes it suitable for various cost-effective light detection applications [24].

Table 2 represents the dielectric values of some commonly utilized insulating materials that are preferred for surface passivation of AlGaN detectors.

Some of the materials like β-Ga₂O₃, SiN_x are reported non-suitable for the passivation of AlGaN type detectors. The reason is that β-Ga₂O₃ passivation requires high temperature oxidation which can cause damage to AlGaN layer. SiN_x passivation causes etching process trouble during fabrication processes [7].

The performance parameters of unpassivated, single passivation layer-based Al_{0.5}Ga_{0.5}N/AlN/Sapphire MSM detectors have been compared with Al₂O₃/SiO₂ passivation stack-based detector. The comparison of performance parameters of four MSM detectors is given in Table 3.

In Table 3, the average value of dark current density in MSM detector without passivation layer is 8e−08 A/cm². This value can be approximated with the dark current density mentioned in [26] which states that larger finger spacing provides lower dark current as compared to equal finger width and spacing-based MSM

Table 2 Dielectric constant of different materials [25]

Type of material	Dielectric constant
SiO ₂	3.9
Al ₂ O ₃	9
β-Ga ₂ O ₃	Between 9.93–10.2
Si ₃ N ₄	7
SiN _x	Between 8–10

Table 3 Performance parameters of AlGa_N MSM detectors under investigation

Type of passivation for MSM detector	Performance parameters				
	Average dark current density (in A/cm ²)	Peak electric field near electrodes under reverse bias of 20 V (in V/cm)	Average recombination rate under incident light (in per cm ³ s)	Average photocurrent density (in A/cm ²)	Photocurrent-to-dark current density ratio (multiplying factor of 1e10)
Without passivation layer	8e-08	2e06	3.25e24	3e05	375
With SiO ₂ passivation layer	1e-05	1.4e06	3.15e24	3e05	300
With Al ₂ O ₃ passivation layer	6e-08	1.8e06	2.5e24	3e05	500
With Al ₂ O ₃ /SiO ₂ passivation stack	5e-08	2.5e06	3e24	3e05	600

detector. Moreover, the dark current density plots of detector structures with passivation show less noise as compared to detector which is without passivation. The dark current density plots of unpassivated and Al₂O₃ passivation-based MSM detectors are shown in Fig. 3a, b. In simulation work, the horizontal cut-line has been drawn in Al_{0.5}Ga_{0.5}N layer at 100 nm from the top surface to extract these plots.

The dark current or leakage current gets reduced due to surface passivation [22] as the density of energy states gets reduced. Therefore, electrical characteristics of device get improved [27]. In Table 3, lowest dark current density of 5e-08 A/cm² and highest photocurrent-to-dark current density ratio have been obtained for Al₂O₃/

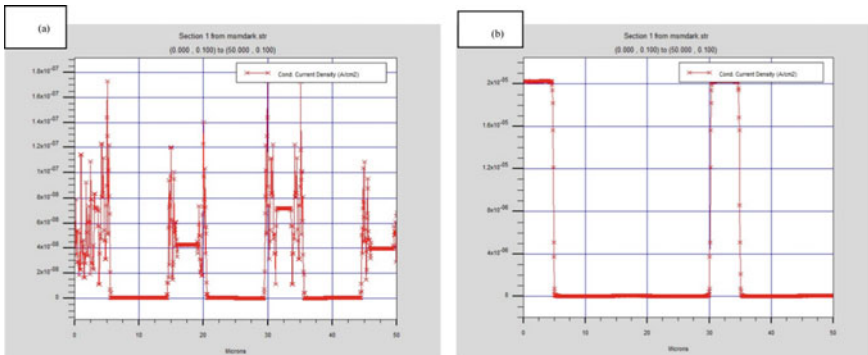


Fig. 3 Dark current density plots for MSM detectors **a** without passivation layer, **b** with SiO₂ passivation layer

SiO₂ passivation stack-based MSM detector. The reason for lower current density can be explained as the reduction of surface states of AlGa_N that lead to minimization of electronic state density at Al₂O₃/SiO₂/AlGa_N interface [28]. Therefore, device performance parameters get improved.

As specified in [23], photocurrent is a measure of photo-absorption rate in the active semiconductor layer and can be taken as current density. In present simulation work, similar photocurrent value has been obtained at 20 V for all detector structures in I-V characteristics plots. Figure 4 shows I-V characteristic plot of Al₂O₃/SiO₂ passivation stack-based detector in Reverse bias varying from 0 to 20 V. Red line shows dark current which is very low as compared to photocurrent. Green color line shows photocurrent which starts to increase from 15 V threshold voltage. The maximum value of photocurrent at 20 V is approximately 8.5 mA.

The electric field near the electrodes is highest in MSM detector which is based on Al₂O₃/SiO₂ passivation layer stack. It is mentioned in [14] that higher value of electric field helps to provide higher carrier velocity. Since electron and hole Shockley–Read–Hall recombination lifetimes for AlGa_N layer is fixed in each detector under investigation, therefore we obtained least dark current density in this type of MSM detector.

In spectral response, we obtained same cut-off wavelength for all detectors. The cut-off wavelength is 400 nm in case of Al₂O₃ passivated detector which is shown in Fig. 5.

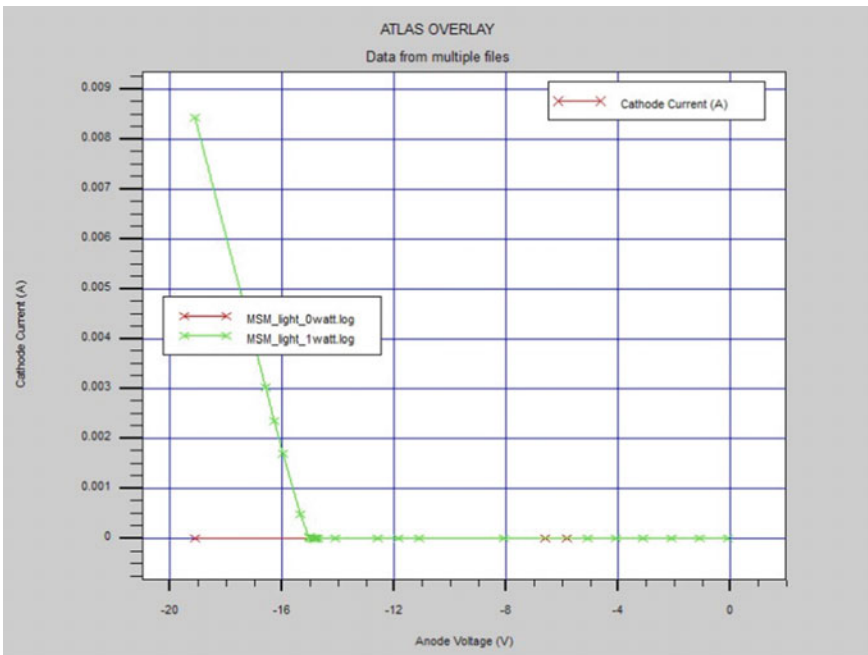


Fig. 4 I-V characteristic plot of MSM detector based on Al₂O₃/SiO₂ passivation layer stack

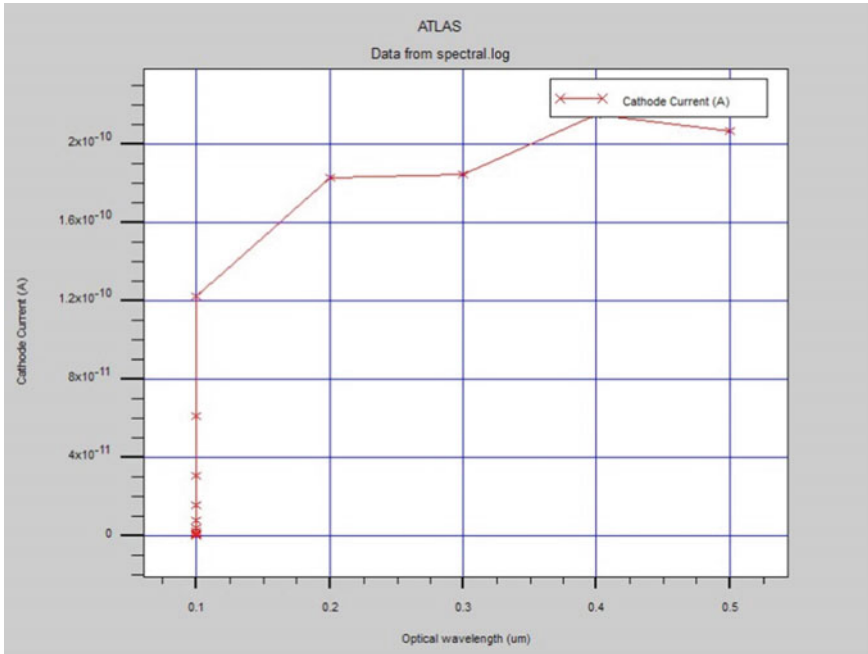


Fig. 5 Spectral response of Al_2O_3 passivated MSM detector

Results obtained in present work are helpful for analyzing the electrical behavior of AlGaIn MSM detectors with and without passivation layer. Therefore, we can state that surface passivation has very less effect on photocurrent and spectral response but significant impact on the dark current, electric field and recombination rate in photo-absorbing layer of MSM photodetector.

4 Conclusion

In present work, effect of passivation layer and passivation stack has been analyzed for AlGaIn MSM detector. The performance parameters of passivated detectors have been compared with unpassivated detector. Passivated detectors using SiO_2 , Al_2O_3 , $\text{Al}_2\text{O}_3/\text{SiO}_2$ stack layers have been investigated for performance parameters like dark current density, electric field, recombination rate, and photocurrent-to-dark current density ratio. It is concluded that detector passivated with stack exhibits lowest dark current density and highest photocurrent-to-dark current density ratio as compared to unpassivated and SiO_2 , Al_2O_3 -based detectors. Therefore, it is important to select appropriate type of passivation material for enhanced and reliable operation in efficient UV detection applications.

References

1. Monroy E, Calle F, Munoz E, Omnes F (1999) AlGa_N metal–semiconductor–metal photodiodes. *Appl Phys Lett* 74(22):3401–3403. <https://doi.org/10.1063/1.123358>
2. Muñoz E, Monroy E, Calle F, Omnes F, Gibart P (2000) AlGa_N photodiodes for monitoring solar UV radiation. *J Geophys Res Atmos* 105(D4):4865–4871. <https://doi.org/10.1029/1999JD900939>
3. Fernández S, Naranjo FB, Sánchez-García MÁ, Calleja E (2020) III-nitrides resonant cavity photodetector devices. *Materials* 13(19):4428. <https://doi.org/10.3390/ma13194428>
4. Xie C, Lu XT, Tong XW, Zhang ZX, Liang FX, Liang L et al (2019) Recent progress in solar-blind deep-ultraviolet photodetectors based on inorganic ultrawide bandgap semiconductors. *Adv Funct Mater* 29(9):1806006. <https://doi.org/10.1002/adfm.201806006>
5. Zou Y, Zhang Y, Hu Y, Gu H (2018) Ultraviolet detectors based on wide bandgap semiconductor nanowire: a review. *Sensors* 18(7):2072. <https://doi.org/10.3390/s18072072>
6. Gassoumi M, Gaquiere C, Maaref H (2011) Surface passivation effects on AlGa_N/Ga_N high electron mobility transistors with SiO₂. *Sensor Lett* 9(6):2175–2177. <https://doi.org/10.1166/sl.2011.1787>
7. Lee CJ, Park H (2016) Surface passivation method for Ga_N UV photodetectors using oxygen annealing treatment. *센서학회지* 25(4):252–256. <https://doi.org/10.5369/JSSST.2016.25.4.252>
8. Lu W, Kumar V, Schwindt R, Piner E, Adesida I (2002) A comparative study of surface passivation on AlGa_N/Ga_N HEMTs. *Solid-State Electron* 46(9):1441–1444. [https://doi.org/10.1016/S0038-1101\(02\)00089-8](https://doi.org/10.1016/S0038-1101(02)00089-8)
9. Luo B, Mehandru R, Kim J, Ren F, Gila BP, Onstine AH et al (2002) Comparison of surface passivation films for reduction of current collapse in AlGa_N/Ga_N high electron mobility transistors. *J Electrochem Soc* 149(11):G613. <https://doi.org/10.1149/1.1512675>
10. Enisherlova KL, Goryachev VG, Rusak TF, Kapilin SA (2016) Study of the effect of passivation layers on capacitance of AlGa_N/Ga_N heterostructures. *Modern Electron Mater* 2(4):131–137. <https://doi.org/10.1016/j.moem.2016.12.006>
11. Wang G, Xie F, Lu H, Chen D, Zhang R, Zheng Y et al (2013) Performance comparison of front-and back-illuminated AlGa_N-based metal–semiconductor–metal solar-blind ultraviolet photodetectors. *J Vac Sci Technol B* 31(1). <https://doi.org/10.1116/1.4769250>
12. Monroy E, Omnes F, Calle FJSS (2003) Wide-bandgap semiconductor ultraviolet photodetectors. *Semicond Sci Technol* 18(4):R33. <https://doi.org/10.1088/0268-1242/18/4/201>
13. Kaur H, Kaur HJ, Hooda MK (2022) Investigating the effect of number of metal electrodes on performance parameters of AlGa_N MSM photodetectors. *Int J Smart Sens Intell Syst* 15(1). <https://doi.org/10.2478/ijssis-2022-0015>
14. Averine SV, Chan YC, Lam YL (2001) Geometry optimization of interdigitated Schottky-barrier metal–semiconductor–metal photodiode structures. *Solid-State Electron* 45(3):441–446. [https://doi.org/10.1016/S0038-1101\(01\)00017-X](https://doi.org/10.1016/S0038-1101(01)00017-X)
15. Huang Y, Yang J, Zhao D, Zhang Y, Liu Z, Liang F, Chen P (2023) A study on the increase of leakage current in AlGa_N detectors with increasing Al composition. *Nanomaterials* 13(3):525. <https://doi.org/10.3390/nano13030525>
16. Liu HY, Liu GJ, Huang RC, Sun WC, Wei SY, Yu SM (2017) In situ growth of Al₂O₃ as a passivation and antireflection layer on TiO₂-based MSM photodetectors. *IEEE Sensors J* 17(16):5087–5092. <https://doi.org/10.1109/JSEN.2017.2724061>
17. Murugapandyan P, Nirmal D, Hasan MT, Varghese A, Ajayan J, Fletcher AA, Ramkumar N (2021) Influence of AlN passivation on thermal performance of AlGa_N/Ga_N high-electron mobility transistors on sapphire substrate: a simulation study. *Mater Sci Eng B* 273:115449. <https://doi.org/10.1016/j.mseb.2021.115449>
18. Wang X, Jing Y, Zhu X, Liu C, Yun J, Zhang Z (2013) Theoretical study on the optical and electrical properties of Al_xGa_{1-x}N crystals. <https://doi.org/10.4236/jmp.2013.43A063>
19. Murthy V, Srivani A, Raghavaiah G (2017) Physical studies in III-nitride semiconductor alloys. *Int J Thin Film Sci Technol* 6:15–27

20. Gelinas RJ (2005) A novel approach to modeling tunnel junction diodes using Silvaco ATLAS software. Doctoral dissertation, Monterey, California. Naval Postgraduate School)
21. Nico V (2005) Analysis and modelling of MSM and Schottky barrier GaN UV detectors. Doctoral dissertation, MS thesis, Dept. Appl. Phys., Eindhoven Univ. Technol., Eindhoven, The Netherlands
22. Garg M, Naik TR, Pathak R, Rao VR, Liao CH, Li KH et al (2018) Effect of surface passivation process for AlGaIn/GaN HEMT heterostructures using phenol functionalized-porphyrin based organic molecules. *J Appl Phys* 124(19). <https://doi.org/10.1063/1.5049873>
23. Manual AUS, Santa Clara CA (2016) ATLAS user's manual
24. Bernát J (2005) Fabrication and characterisation of AlGaIn/GaN high electron mobility transistors for power applications. Doctoral dissertation, Aachen, Techn. Hochsch., Diss.
25. Robertson J (2004) High dielectric constant oxides. *Eur Phys J Appl Phys* 28(3):265–291. <https://doi.org/10.1051/epjap:2004206>
26. Kaur H, Kaur HJ, Hooda MK, Dassi M (2023) Electrical performance analysis of Al_{0.5}Ga_{0.5}N/AlN/Sapphire-based MSM UV detector for high photocurrent. *J Optics* 52(1):355–364. <https://doi.org/10.1007/s12596-022-00904-1>
27. Gupta A, Chatterjee N, Kumar P, Pandey S (2017) Effect of surface passivation on the electrical characteristics of nanoscale AlGaIn/GaN HEMT. In: *IOP Conf Series Mater Sci Eng* 225(1):012095. <https://doi.org/10.1088/1757-899X/225/1/012095>
28. Yatabe Z, Asubar JT, Hashizume T (2016) Insulated gate and surface passivation structures for GaN-based power transistors. *J Phys D Appl Phys* 49(39):393001. <https://doi.org/10.1088/0022-3727/49/39/393001>

Tachyon: A Programmable Optoelectronic Hardware Accelerator for Ultrafast Tensor Arithmetic



Sathvik Redrouthu, Jagadeepam Maddipatla, Pranav Vadde, and Anant Khandelwal

Abstract We discuss the design and implementation of Tachyon, a high-speed, scalable optical hardware accelerator. In addition, we go over programming the accelerator with an instruction set architecture (ISA) capable of efficiently communicating with on-chip optical components to perform ultrafast $\mathcal{O}(1)$ matrix-vector multiplication (MVM). We first introduce the optoelectronic accelerator's computing scheme. We then go over our pipeline for programming the accelerator and introduce our optical ISA. Our results show that the accelerator's optical core is scalable and can perform an MVM in a single pass with a performance per watt (PPW) clocked 3 orders of magnitude higher than Google's Tensor Processing Unit. The core is rated at 7 TFLOPS and 245 TFLOPS/W for a 4×4 matrix tile, but scales to 478 TFLOPS and 261 TFLOPS/W for a 256×256 .

Keywords Data analytics · High-speed computing · Machine learning · Matrix-vector multiplication · Neural networks · Optical computing

1 Introduction

Over the past decade, there has been a dramatic increase in the parameter count of neural networks, driven by advances in machine learning algorithms, hardware, and data availability. This increase has enabled significant improvements in performance on a wide range of tasks, from image classification to natural language processing. In

S. Redrouthu (✉) · J. Maddipatla (✉) · P. Vadde · A. Khandelwal
Procyon Corp, Clearwater, FL, USA
e-mail: sathvikr@procyoncorp.io

J. Maddipatla
e-mail: jag.maddipatla@procyoncorp.io

P. Vadde
e-mail: pranav.vadde@procyoncorp.io

A. Khandelwal
e-mail: pranav.vadde@procyoncorp.io

2012, the state-of-the-art image classification model, AlexNet, had only 61 million parameters. By 2015, the VGG-19 model had 143 million parameters, and by 2016, the ResNet-152 model had 60 million parameters. In 2018, the DenseNet-264 model had 36.4 million parameters, while the EfficientNet-B7 model, released in 2019, had 66 million parameters [1]. The parameter counts of natural language models have also increased significantly. In 2015, the state-of-the-art language model had only 5 million parameters. By 2018, the OpenAI GPT-2 model had 1.5 billion parameters, and by 2020, the GPT-3 model had 175 billion parameters [2].

The increase in parameter count has enabled significant improvements in performance on a range of benchmarks. For example, in the ImageNet Large Scale Visual Recognition Challenge, the top-1 error rate of the winning models decreased from 26.2% in 2012 to 3.5% in 2021. Similarly, in the GLUE benchmark for natural language understanding, the state-of-the-art performance has increased from 81.6% in 2018 to 90.4% in 2021 [2].

However, the increase in parameter count has also raised concerns about the computational cost and environmental impact of training and running these models. For example, training the GPT-3 model on a single accelerator can consume up to 1.2 GWh of electricity [3]. To address these concerns, in addition to exploring various techniques for reducing the neural network parameter counts while maintaining performance, researchers are building more efficient hardware. The market demand for high performance per watt chips is incredibly high, resulting in large companies to join in developing powerful hardware applicable to even future model iterations, e.g., Cerebras AI lining up to tackle GPT-4's estimated trillion-parameter workload.

Such AI-based applications primarily base their functionality in tensor arithmetic. NVIDIA, a pioneer in high performance computing (HPC), takes advantage of this by building efficient electronic processors labeled Graphics Processing Units (GPUs). These GPUs are able to accelerate artificial intelligence processes by enabling high-speed matrix multiplication to take place through tensor cores or unique applications of the Single Instruction-Stream, Multiple Data-Stream (SIMD) architecture. Google developed a digital Application-Specific Integrated Circuit (ASIC) specialized in AI processing, labeled a Tensor Processing Unit (TPU). The TPU specializes in computing important activation functions and matrix multiplication for NN inference [4]. The Edge TPU is able to perform 4 trillion operations per second using 2 watts of power (2 TOPS/watt) [5]. It is apparent that many significant hardware advancements have been made over the decades.

However, while these advancements have proven to be advantageous with regard to computational efficiency, they are fundamentally limited by the decline of trends such as Moore's Law and Dennard Scaling. Moore's Law asserts that the number of transistors on a digital electronic microchip will approximately double every two years, directly correlating to the increase in computing power [6]. However, it has been predicted by experts that this trend will no longer be viable by 2036 due to an inability to further reduce the size of transistors. For this reason, coupled with the end of Dennard Scaling, their computational ability will stagnate unless matrix processing systems are reinvented to function without the usage of digital electronics.

To account for this problem, the use of analog computing for AI has been investigated; such a platform eliminates the reliance on transistors. Years of research and production have confirmed that while having lower accuracy and higher loss, the analog approach is generally more efficient for operations such as matrix multiplication [7, 8]. Mythic AI's M1076 analog matrix processor achieves the AI compute performance of a digital electronic desktop GPU while consuming 10% of the energy, the typical power consumption for running complex neural networks being approximately 3–4 W [9]. Digital-analog electronic hybrid processors have also been investigated in an attempt to maximize speed and accuracy together [8].

However, most electronics-based systems are subject to interference and temperature problems. Even recently created analog architectures, such as Mythic's AMPs, possess limitations from electromagnetic crosstalk and Joule heating. In addition, the computing power required for functional operation scales dramatically with NN feature size. Such problems are sidestepped through silicon photonics, with matrix-vector multiplication (MVM) able to be done cleanly and passively. While such circuits are themselves inherently analog, they are significantly faster than their analog electronic equivalents. In this paper, we explore a hybrid silicon photonic-digital electronic hardware accelerator. We first go over the accelerator design and implementation. We then discuss the programming interface used to run tensor arithmetic on the accelerator. We finish by discussing results and implications.

2 Accelerator Design

2.1 Overview

The Tachyon accelerator contains an optical matrix processing unit (MPU), an engine where MVM sequences are executed at high speed to power important linear algebra-based tasks such as NN inference. When an operation is run within an application, relevant aspects of it are sent to the MPU (Fig. 1).

Tachyon's instruction set architecture (ISA) is given in Table 1. When instructed to a) write a matrix or vector to the MPU, or b) compute an MVM, the host platform sends instructions to a master microcontroller within the Tachyon accelerator via serial communications. The master relays corresponding digital signals to digital-to-analog converters (DACs); corresponding analog signals are then used to drive a grid of laser beams that send analog optical signals within the MPU. These signals are then manipulated to perform an MVM as outlined in Sect. 2.2. Output optical signals are read through photodetectors and converted into corresponding analog current streams, which are then converted to digital signals through analog-to-digital converters (ADCs). The bitstreams are sent back to the master and results are reported back to the host through serial.

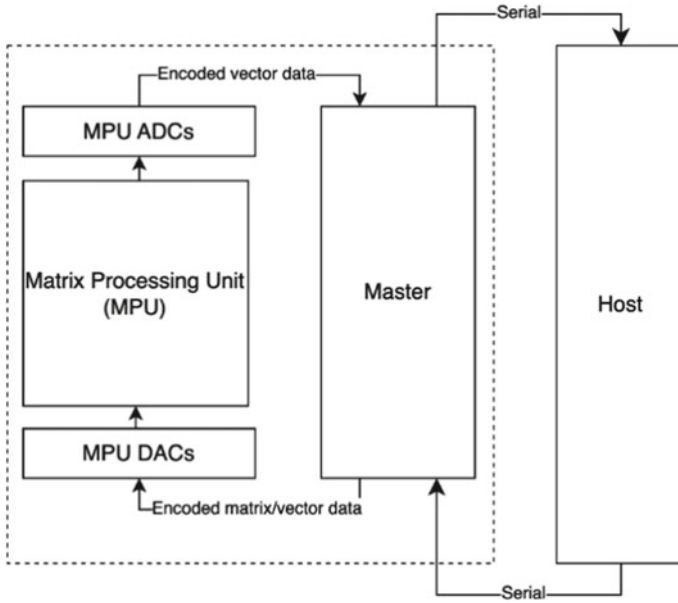


Fig. 1 System on-chip diagram of the Tachyon hardware accelerator

Table 1 Instruction set

Name	Location
mwrite m	MPU
vsend v	MPU
mvmul a b	MPU

2.2 Matrix Processing Unit

The MPU, shown in Fig. 3, performs a matrix-vector multiplication (MVM) using analog optical signals. A single laser toggles on and the output beam is split into n channels through an even beam splitter network. A modulator is integrated into each splitter arm to alter incoming light discretely, inherently allowing unique elements in the input vector \vec{x} to the MPU. The grid of laser light is then altered by components within the MPU that correspond to a matrix transformation $\vec{x} \mapsto \mathbf{M}\vec{x}$, where \mathbf{M} is determined by the specific component settings. The results are sent as optical analog signals to a grid of photodetectors, each of which reports back an analog current stream. This concept is shown in Fig. 2.

To configure the components within the MPU to represent any $\mathbf{M} \in \mathbb{R}^{m \times n}$, we take advantage of the fact that any such matrix \mathbf{M} can be represented as $\mathbf{U}\Sigma\mathbf{V}^*$ through the singular value decomposition (SVD), where $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V}^* \in \mathbb{R}^{n \times n}$ are unitary, and $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix. The MPU performs $\vec{x} \mapsto \mathbf{M}\vec{x}$ by splitting \mathbf{M} into these 3 smaller matrix transformations, as shown in Fig. 4. Light passes through each input port; the corresponding Stokes vector is transformed by \mathbf{V}^* , Σ , and \mathbf{U} ,

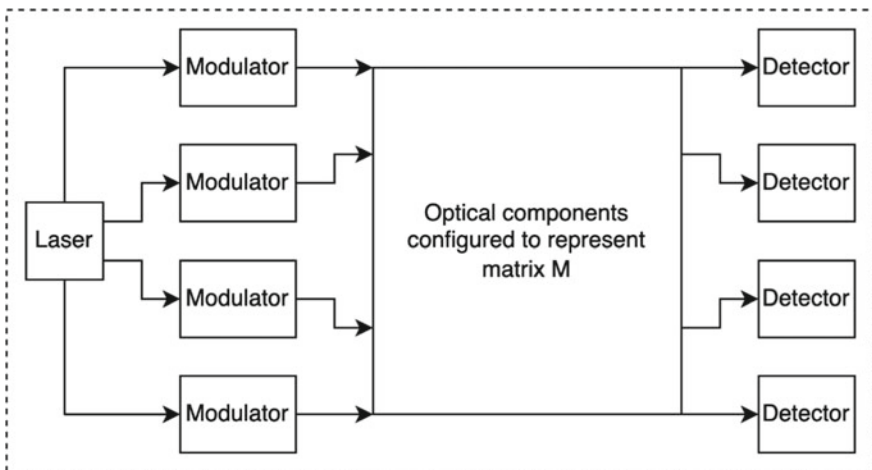


Fig. 2 $\vec{x} \mapsto \mathbf{M}\vec{x}$ transformation pipeline. On-chip driving laser split across beam splitter network, tuned according to input vector \vec{v} , sent to optical components implementing \mathbf{M} , and read out via on-chip photodetectors

respectively. This essentially means that rather than explicitly manipulating each entry in \mathbf{U} , Σ , and \mathbf{V}^* , the MPU manipulates a set of phase angles parameterizing each matrix. Doing so allows the MVM computation to be done on hardware in $\mathcal{O}(1)$, but extra work is required for setting up calculations (writing matrices and vectors).

2.2.1 Optical Unitary Matrix Transform

The MPU uses silicon photonic sub-circuits to implement the unitary matrix transforms given by \mathbf{U} and \mathbf{V}^* . A lattice of Mach-Zehnder Modulators (MZMs) are laid out in the pattern proposed by Clements et al. [10]. Each MZM is 50:50 directional-coupler based, with the following $\mathbb{R}^{2 \times 2}$ transfer matrix:

$$\mathbf{U}^{(2)} = \frac{1}{2} \begin{bmatrix} e^{j\alpha^{(2)}}(e^{j\theta^{(2)}} - 1) & je^{j\alpha^{(2)}}(1 + e^{j\theta^{(2)}}) \\ je^{j\beta^{(2)}}(1 + e^{j\theta^{(2)}}) & e^{j\beta^{(2)}}(1 - e^{j\theta^{(2)}}) \end{bmatrix} \in \mathbb{R}^{2 \times 2}. \quad (1)$$

The phase angles θ , α , β are shifts induced by modulators incorporated in rib waveguides within the MZM. The full MZM lattice then represents the unitary matrix $\mathbf{U}^{(n)} \in \mathbb{R}^{n \times n}$ as

$$\mathbf{U}^{(n)} = \prod_{i=1}^n \prod_{j=1}^{n-i} \mathbf{T}_{n-j, n-j}(\theta_{i,j}^{(n)}, \alpha_{i,j}^{(n)}, \beta_{i,j}^{(n)}) \in \mathbb{R}^{n \times n}, \quad (2)$$

where $\mathbf{T}_{i,j} \in \mathbb{R}^{n \times n}$ contains the top left of a unitary matrix $\mathbf{U}^{(2)} \in \mathbb{R}^{2 \times 2}$ at location (i, j) , a diagonal of 1s (excluding $\mathbf{U}^{(2)}$), and the remaining entries as 0s [10].

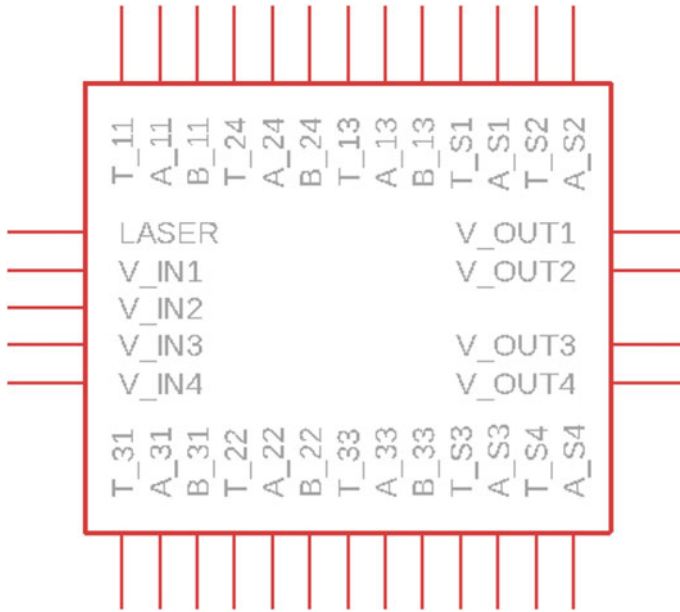


Fig. 3 Example MPU accepting an \mathbb{R}^4 vector as input. One optical input from the LASER port is tuned by analog voltage signals from 30 other ports

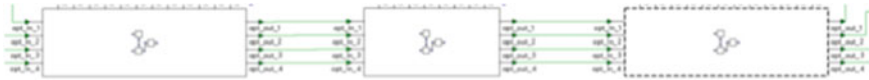


Fig. 4 Optical V^* , Σ , and U cores making up the MPU, respectively

2.2.2 Optical Diagonal Matrix Transform

Left-multiplying a vector $\vec{v} \in \mathbb{R}^n$ by a diagonal matrix Σ effectively scales each entry in \vec{v} by a corresponding entry in Σ . We therefore use n single ports of dual-modulator MZMs arranged in a column, where each MZM scales the input electric field:

$$E_{out} = \frac{E_{in}}{2} e^{j\omega} (e^{j\phi} - 1). \tag{3}$$

2.3 Supporting Electronics

Voltaire is the interconnect enabling the optical Tachyon chip to communicate with external systems. The MPU operates using analog voltage signals, which are used to produce and manipulate laser light. We use a series of off-the-shelf microcontrollers

and other components such as digital-to-analog converters (DACs) in a master-slave system to organize, distribute, and convert digital signals from the host hardware. The host interfaces with these components through serial communications protocols. After the MPU completes its operation, it outputs an analog differential current signal which is converted to a digital signal using analog-to-digital converters (ADCs). These digital signals are sent back to the host hardware, and the next operation is ready to be carried out.

2.3.1 Microcontroller Master-Slave System

Information from the host is communicated to a Teensy 4.1 development board through a hardware serial communications protocol [11]. The Teensy is connected to an additional pair of microcontrollers: Texas Instrument's Arm Cortex based RM57L843 [12]. The devices are set up in a master-slave system using the Serial Peripheral Interface (SPI) protocol with the Teensy as the master and the 2 RM57L843 microcontrollers as slaves. The vector and matrix data from the host is distributed from the Teensy to each RM57L843. The slaves then distribute the digital signals representing element data to each DAC. The output analog signals from the MPU are routed directly to the slaves which then convert them back to digital signals using onboard ADCs. The resultant element data is communicated back to the host via the Teensy through SPI and hardware serial.

2.3.2 Signal Converter System

The digital signals from the host need to be converted to analog signals to be compatible with the optical components in the MPU. The digital signals from the RM57L843 slave microcontrollers are converted to analog signals using Analog Devices' AD9748 DAC [13]. The converter accepts 8 digital signals as input to represent an 8 bit digital representation of the signal. The DAC then outputs a differential current as the analog representation of the 8 bit digital signal. The AD9748 includes an option for scaling the output to fit the MPU's input parameters. This differential current is then converted to an analog voltage signal using Analog Device's AD8047 operational amplifier (op amp) [14]. Figure 5 shows this DAC assembly. This analog voltage is passed to MPU for calculation. There are a total of 36 DACs for every input the MPU requires, including the laser, the initial 4 dimensional vector, and the initial 4×4 matrix. Figure 6 shows a schematic of the DACs connected to the MPU.

The MPU's photodetectors output a differential current to represent each output vector element. This is converted to a single analog voltage signal and routed to the ADCs on each slave microcontroller. Each slave has 2 ADC modules with up to 12 bits of resolution, with the option of reducing resolution for higher conversion speeds. The slave ADCs convert the 4 output analog signals into digital form to be sent back to the host.

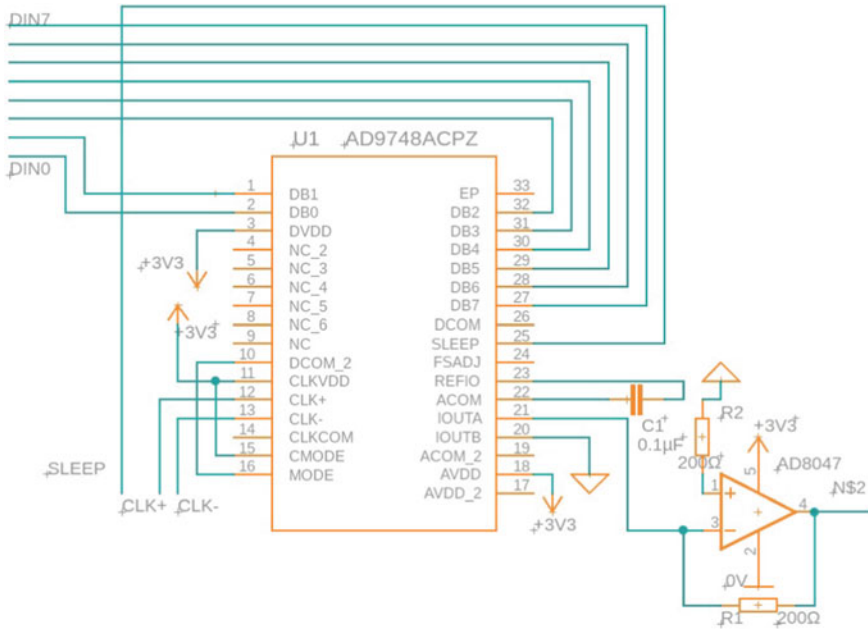


Fig. 5 AD9748 current-driven DAC extended to output single-ended voltage with an AD8047 operational amplifier. Both resistors have a 200 resistance, whereas the capacitor is rated at 0.1. Connections for digital signal inputs, analog signal output, power, timing signals, and the power down function are noted

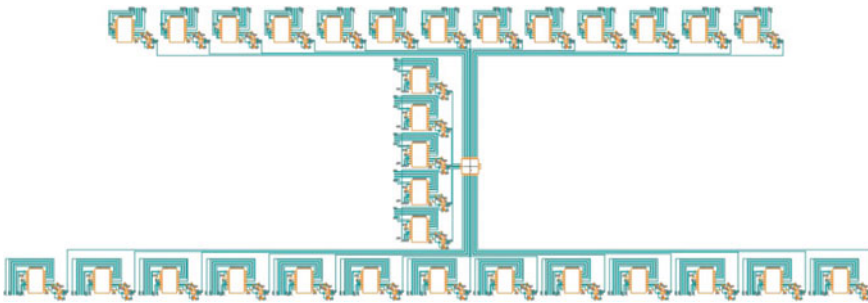


Fig. 6 Schematic of DACs connected to the MPU

3 Accelerator Implementation

Chip fabrication took place at the Maryland NanoCenter. We used the Raith eLINE electron-beam lithography (EBL) machine operated at 100 keV with an 8 nA beam current on a 20 nm process node. The Matrix Processing Unit (MPU) itself is on a diced 605 μm × 410 μm × 0.22 μm silicon wafer.

3.1 Matrix Processing Unit

3.1.1 Passive Circuitry

We use strip waveguides for passive structures throughout the Matrix Processing Unit, each with a height of 220 nm and a width of 20 nm. The waveguide cladding is composed of Silicon Dioxide (SiO₂). The waveguide bend radius is 5, the smallest bend radius practical for transverse electric (TE) polarization. The waveguide mode profile, along with a corresponding wavelength sweep, is shown in Fig. 7. Through the obtained data, the compact model was calculated:

$$n_{\text{eff(undoped)}}(\lambda) = 2.56862 - 0.729403(\lambda - 1.55) + 0.0995556(\lambda - 1.55)^2 + jn_{\text{eff(imag)}}. \quad (4)$$

The last term is included to convert the loss to the imaginary part of the effective index. Here, $n_{\text{eff(imag)}}$ is equal to the standard $\frac{\lambda[\text{cm}]}{4\pi} \ln(10^{\text{loss}[\frac{\text{dB}}{\text{cm}}]}/10)$.

3.1.2 Phase Shifters

We use rib waveguides within our phase shifters. To achieve the phase shifts θ , α , β , we modulate the light using electro-optic modulators. Electro-optic modulators induce a phase shift by altering the effective index of the rib waveguide composing it. Forward-biased PIN diodes are able to achieve a larger effective index change Δn_{eff} . However, their speed is limited by the carrier recombination rate. In addition, they consume a significant amount of power and have a large absorption α . Reverse-biased PN diodes, on the other hand, consume less power and are able to achieve higher speeds, but have a smaller effect on Δn_{eff} and a greater loss.

Taking these factors into consideration, we proceed to use reverse-biased PN diodes in our electro-optic modulators, a Computer Aided Design (CAD) model of which is shown in Fig. 8. The cathode and anode are made from aluminum, and we use the standard carrier recombination models. An increase in applied voltage to one of these diodes corresponds to a decrease in the width of the depletion layer, directly modulating the light. The rib waveguide after carrier injection is shown in Fig. 9.

3.1.3 Mach-Zehnder Modulator

To implement a Mach-Zehnder Modulator (MZM), we place an electro-optic modulator on each arm of a Mach-Zehnder Interferometer. Our undoped MZM's compact model is given by

$$n_{\text{eff(undoped)}}(\lambda) = 2.54606 - 0.872858(\lambda - 1.55) + 0.196977(\lambda - 1.55)^2 + jn_{\text{eff(imag)}}. \quad (5)$$

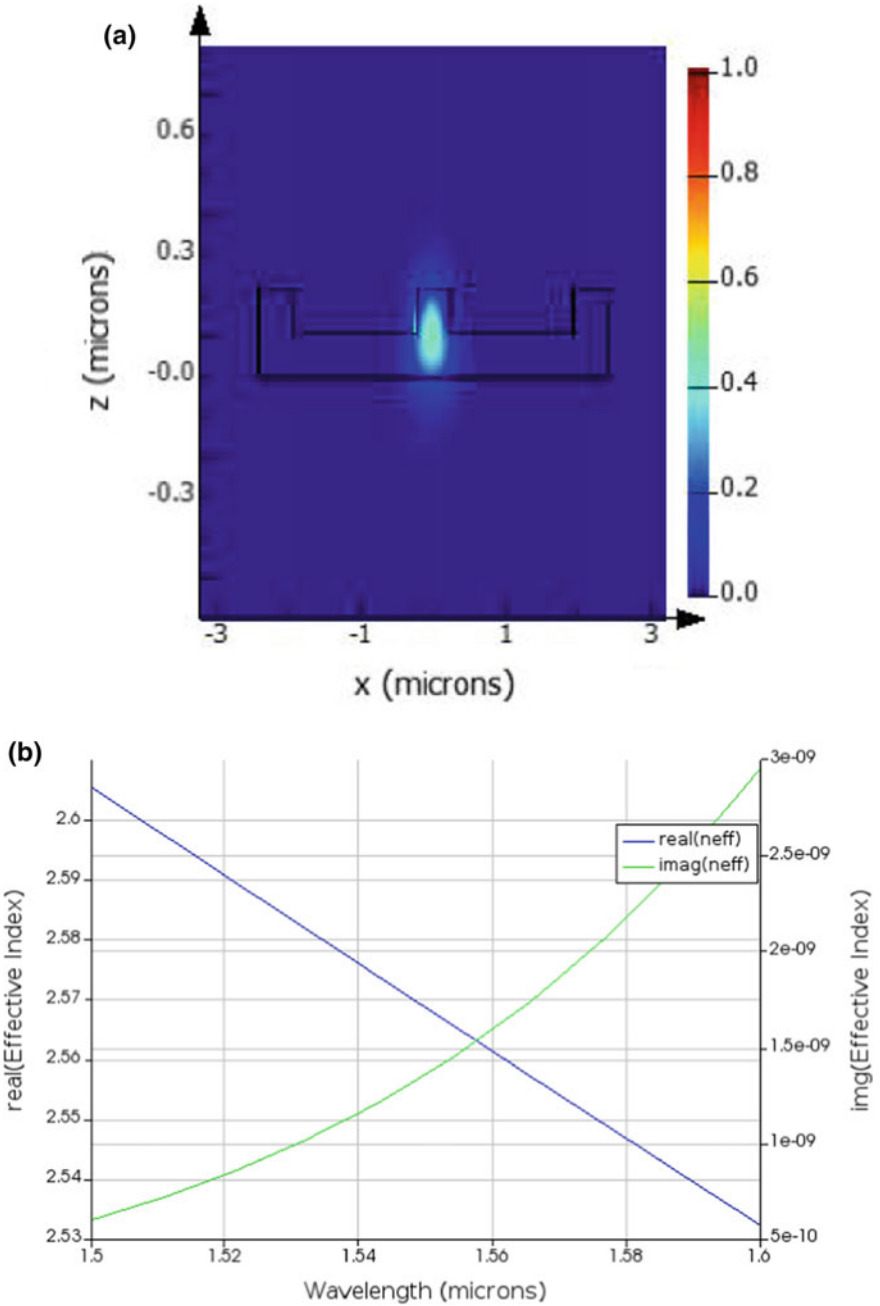


Fig. 7 a Mode profile and b wavelength versus effective index in our undoped rib waveguides

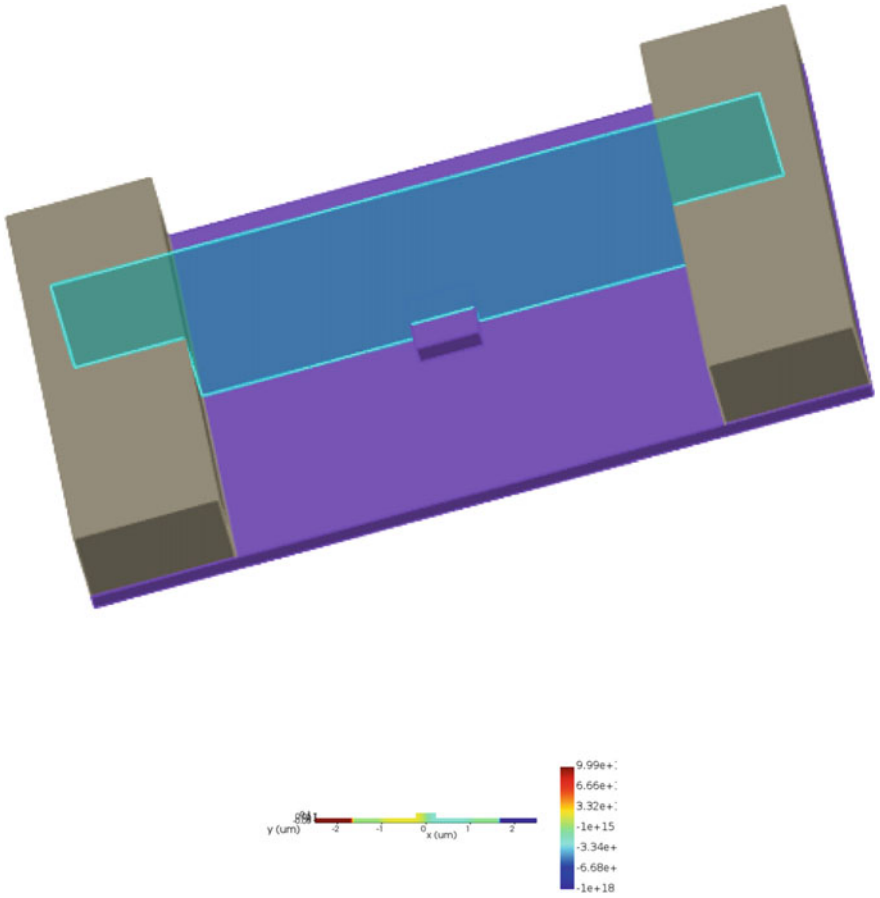


Fig. 8 **a** We integrate a PN junction into the center of a rib waveguide. **b** N and P doped regions are placed to the left and right, respectively, increasing in doping concentration as shown

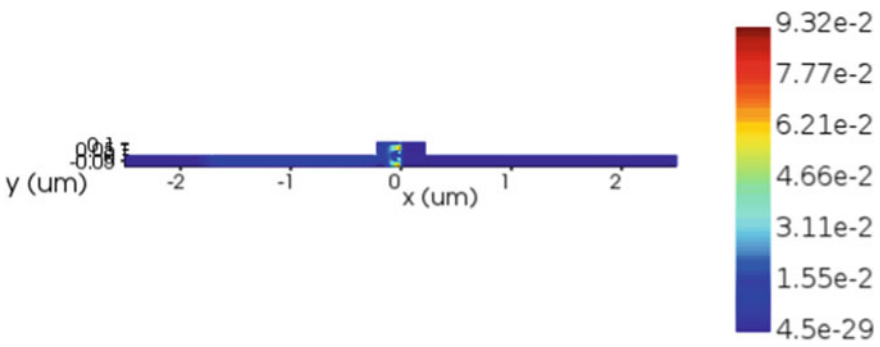


Fig. 9 Waveguide charge profile

We now let the phase angles $\theta, \alpha, \beta = \frac{2\pi}{\lambda}L(n_{\text{eff}(\text{undoped})_{1,2}}(\lambda) + \Delta n_{\text{eff}1,2}(V))$, noting that we can halve the driving voltage needed by using the push-pull operation, where we apply opposite voltages to each electro-optic modulator on the MZM arm.

We integrate broadband directional couplers and extra modulators to our MZMs when needed. Note that our directional couplers each have a transmission coefficient $\tau = \frac{1}{\sqrt{2}}$ and a coupling coefficient $\kappa = \frac{1}{\sqrt{2}}$.

3.1.4 Optical I/O

Our design makes use of on-chip quantum lasers and photodetectors for optical I/O.

During our fabricated chip tests, we used an off-chip HP81680A continuous wave (CW) diode laser and HP81635A InGaAs photodetectors.

3.2 Supporting Electronics

All electronic and optical components are to be placed on a custom designed Printed Circuit Board (PCB). Most off-the-shelf and all optical components are surface mounted. The Teensy 4.1 development board is inserted into surface mounted female headers and connected to the host using a USB connection. The MPU is packaged into a Ceramic Pin Grid Array (CPGA) format which is inserted into a socket on the PCB. An example of the CPGA package is shown in Fig. 10. The RM57L843 slave microcontrollers are manufactured into 337-pin Ball Grid Array (BGA) packages, the AD9748 DACs are 32-lead Lead Frame Chip Scale Packages (LFCSP), and the AD8047 op amps are 8-lead Small Outline Integrated Circuit (SOIC) packages.

Listing 1 Initiating a tensor algebra request on the host platform through Apollo.

```
// initialize vector v and rank-3 tensor T
let tensor v = {2, 9, 5};
let tensor T =
{
  {
    {3, 4, 2},
    {0, 8, 9},
    {1, 9, 4}
  },
  {
    {31, 3, 5},
    {0, 0, 3},
    {43, 0, 16}
  }
};
// compute tensor dot product and store resultant tensor R
let tensor R = T * v;
```

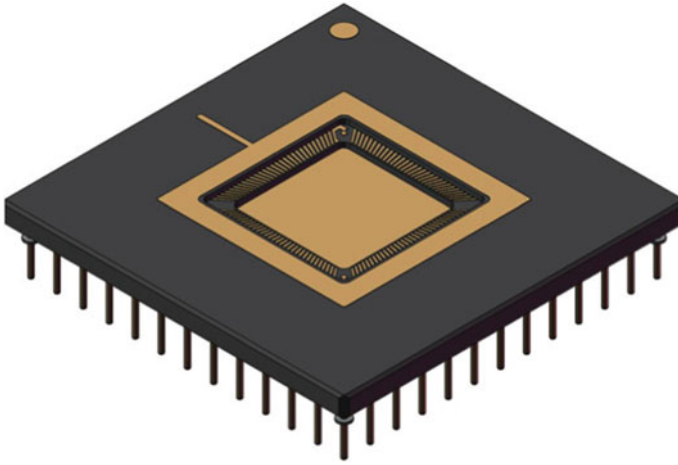


Fig. 10 CPGA packaging used for the MPU

4 Programming the Accelerator

4.1 Overview

We intend for the Tachyon accelerator to be programmable from an external host platform. Therefore, we developed a custom Domain-Specific Language (DSL) oriented toward tensor algebra computations, coined Apollo [15]. Apollo supports the scalar-tensor product, rank- n Kronecker product, tensor inner product, tensor dot product, Khatri-Rao product, face-splitting product, and the \mathbb{R}^3 vector cross product. The language also supports writing tensors and vectors directly to the MPU for ease of future calculations [15].

As indicated in Fig. 11, the accelerator programming pipeline starts when user creates a program on a host platform. The Apollo compiler front end generates an optimized program tree given any input code sequence.¹ The compiler back end then takes the optimized tree and traverses it to generate appropriate Virtual Machine (VM) instructions. The VM then determines if the selected instruction involves MVM. If so, it accepts each leftover matrix or vector from the tensor computation, (1) mapping into corresponding phase angles/voltages and (2) scheduling groups of instructions to be sent to the on-accelerator microcontroller. If not, it keeps it within the host platform and executes it traditionally.

¹ The specifics of the compiler front end are discussed in [15].

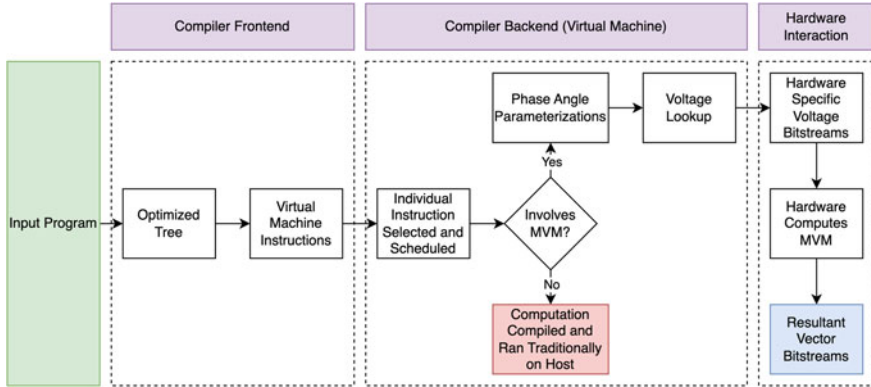


Fig. 11 General program flow within the Apollo DSL compiler

4.2 Compiler Back End

The compiler back end takes the generated Virtual Machine (VM) instructions as input. For any selected instruction, it schedules it to be run and determines whether or not it is MVM-based. If it is MVM-based, the VM runs a process unique to the Tachyon setup. In the case where a rank- n tensor-related request (where $n > 2$) is initialized as in Listing 1, the compiler implicitly allocates memory in the Binary Sparse Tensor Tree format [15]. The tensor is recursively unrolled into matrices as shown in Fig. 12. Each matrix \mathbf{M} is then converted into a set of voltages which, when sent to the modulators, would induce phase shifts collectively representing \mathbf{M} . This is done through the phase angle parameterizations discussed in Sect. 4.3. The assembly instruction in Table 1 corresponding to the request is then executed. If $n < 2$, the unrolling step is skipped, and the remaining steps stay the same. Results are sent back to the host computer and ripple up the compiler back to the user.

4.3 Phase Angle Parameterization

It is important to calculate the required phase angles as parameterizations of input matrices and vectors to allow the compiler to send the appropriate signals. Since the Matrix Processing Unit (MPU) represents a matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ through the SVD, parameterizing \mathbf{M} is done by parameterizing each individual SVD term.

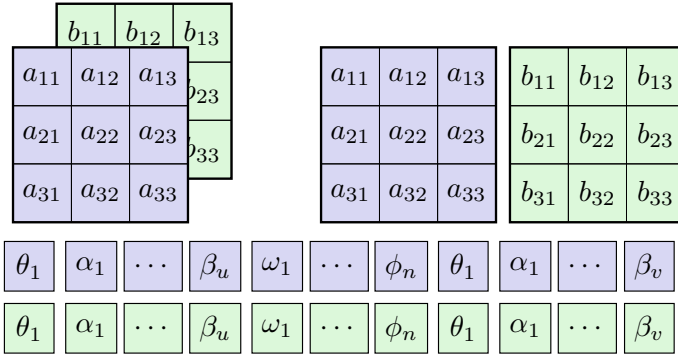


Fig. 12 Writing a tensor to the MPU. **a** An tensor is initialized and **b** unrolled into matrices by the Apollo compiler; **c** each matrix is then translated into a set of angles parameterizing it. The angles are mapped into the voltages needed to achieve them and are then sent to the MPU

4.3.1 Unitary Matrix Phase Angles

We first go over phase angle parameterizations of unitary matrices $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V}^T \in \mathbb{R}^{n \times n}$. Recall that the MPU uses a Mach-Zehnder Modulator (MZM) to implement (1). The values of the phase angles $\theta^{(2)}, \alpha^{(2)}$, and $\beta^{(2)}$ in terms of the desired matrix constants can be calculated through a system of equations. For any given unitary matrix $\mathbf{U}^{(2)} \in \mathbb{R}^{2 \times 2}$, with values $u_{11}^{(2)}, u_{12}^{(2)}, u_{21}^{(2)}, u_{22}^{(2)}$, it is true that

$$u_{11}^{(2)}(\alpha^{(2)}, \beta^{(2)}, \theta^{(2)}) = \frac{1}{2} e^{-j\alpha^{(2)}} (e^{-j\theta^{(2)}} - 1) \tag{6a}$$

$$u_{12}^{(2)}(\alpha^{(2)}, \beta^{(2)}, \theta^{(2)}) = \frac{j}{2} e^{-j\alpha^{(2)}} (1 + e^{-j\theta^{(2)}}) \tag{6b}$$

$$u_{21}^{(2)}(\alpha^{(2)}, \beta^{(2)}, \theta^{(2)}) = \frac{j}{2} e^{-j\beta^{(2)}} (1 + e^{-j\theta^{(2)}}) \tag{6c}$$

$$u_{22}^{(2)}(\alpha^{(2)}, \beta^{(2)}, \theta^{(2)}) = -\frac{1}{2} e^{-j\beta^{(2)}} (e^{-j\theta^{(2)}} - 1). \tag{6d}$$

We vary the phase angle values and store $\vec{\mathbf{f}}_{u^{(2)}}(\theta^{(2)}, \alpha^{(2)}, \beta^{(2)}) = u_{11}^{(2)} \hat{\mathbf{e}}_1 + u_{12}^{(2)} \hat{\mathbf{e}}_2 + u_{21}^{(2)} \hat{\mathbf{e}}_3 + u_{22}^{(2)} \hat{\mathbf{e}}_4$ in a lookup table, where $\vec{\mathbf{f}}_{u^{(2)}} : \{0 \leq \alpha^{(2)}, \beta^{(2)}, \theta^{(2)} \leq 2\pi\} \rightarrow \mathbb{C}^4$. We read phase angle values backwards from the stored $\mathbf{U}^{(2)}$ constants; if they are not available during lookup, we compute them in real time. Solving (6) yields the following phase angle equations under constraints $\{u_{11}^{(2)} = u_{22}^{(2)} = 0, u_{12}^{(2)} \neq 0, u_{21}^{(2)} u_{12}^{(2)} = -1\}$:

$$\theta^{(2)} = 2\pi C_1 \tag{7a}$$

$$\alpha^{(2)} = 2\pi j C_2 + \ln \left[-j u_{22}^{(2)} \right] \tag{7b}$$

$$\beta^{(2)} = 2\pi C_3 - j \ln \frac{j}{u_{12}^{(2)}}. \tag{7c}$$

or, under constraints $\{u_{12}^{(2)2} - u_{11}^{(2)2} \neq 0, u_{11}^{(2)} \neq 0, u_{11}^{(2)} - j u_{12}^{(2)} \neq 0, u_{21}^{(2)} = -\frac{u_{12}^{(2)}}{u_{12}^{(2)2} - u_{11}^{(2)2}}, u_{22}^{(2)} = -\frac{u_{11}^{(2)}}{u_{11}^{(2)2} - u_{12}^{(2)2}}\}$:

$$\theta^{(2)} = 2\pi C_1 - j \ln \frac{j u_{12}^{(2)} - u_{11}^{(2)}}{u_{11}^{(2)} + u_{12}^{(2)}} \tag{8a}$$

$$\alpha^{(2)} = 2\pi C_2 - j \ln \left[-u_{11}^{(2)} - j u_{12}^{(2)} \right] \tag{8b}$$

$$\beta^{(2)} = 2\pi C_3 - j \ln \frac{u_{11}^{(2)} + j u_{12}^{(2)}}{u_{12}^{(2)2} - u_{11}^{(2)2}}. \tag{8c}$$

where C_1, C_2, C_3 are constants. These give the required phase angles directly. We set them to zero for simplicity.

Recall that on the MPU, a unitary matrix $\mathbf{U}^{(n)} \in \mathbb{R}^{n \times n}$ is represented as (2). The angles $\theta, \alpha,$ and β then build up to $\frac{3}{2}n(n-1)$ angles used to parameterize $\mathbf{U}^{(n)}$:

$$\theta_1^{(n)}, \alpha_1^{(n)}, \beta_1^{(n)}, \theta_2^{(n)}, \dots, \alpha_{\frac{1}{2}n(n-1)}^{(n)}, \beta_{\frac{1}{2}n(n-1)}^{(n)}. \tag{9}$$

We solve for these angles by equating each entry of $\mathbf{U}^{(n)}$ to a series of $u_{mij}^{(2)}$ products and substituting each vector field angle stored. In the 4×4 case,

$$\mathbf{U}^{(4)} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \mathbf{U}_1^{(2)} & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & \mathbf{U}_2^{(2)} & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} \mathbf{U}_3^{(2)} & & & \\ & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix} \tag{10}$$

$$\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \mathbf{U}_4^{(2)} & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & \mathbf{U}_5^{(2)} & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \mathbf{U}_6^{(2)} & \\ & & & 1 \end{bmatrix} \in \mathbb{R}^{4 \times 4}.$$

Note that the constraint $\det(\mathbf{U}^{(4)}) = \det(\mathbf{U}_{1,2,\dots,6}^{(2)}) = 1$ must still hold. Since there are 16 $u_{ij}^{(4)}$ values and 24 $u_{mij}^{(2)}$ values, many possible solutions exist for this system. The Jacobian matrix \mathbf{J} of a function $\vec{\mathbf{f}}(u_{111}^{(2)}, u_{112}^{(2)}, \dots, u_{622}^{(2)}) : \mathbb{R}^{24} \rightarrow \mathbb{R}^{16}$ mapping vectors of $\mathbf{U}^{(2)}$ values to $\mathbf{U}^{(4)}$ values is highly sparse and rank-deficient, so we solve using a tensor-Krylov method rather than a multivariate form of Newton’s method [16]. Convergence plots are shown for various trials in Fig. 13.

We store $\vec{\mathbf{f}}_{u^{(n)}} : \mathbb{C}^{\frac{3}{2}n(n-1)} \rightarrow \mathbb{C}^{n^2}$ in a secondary lookup table for efficient access.

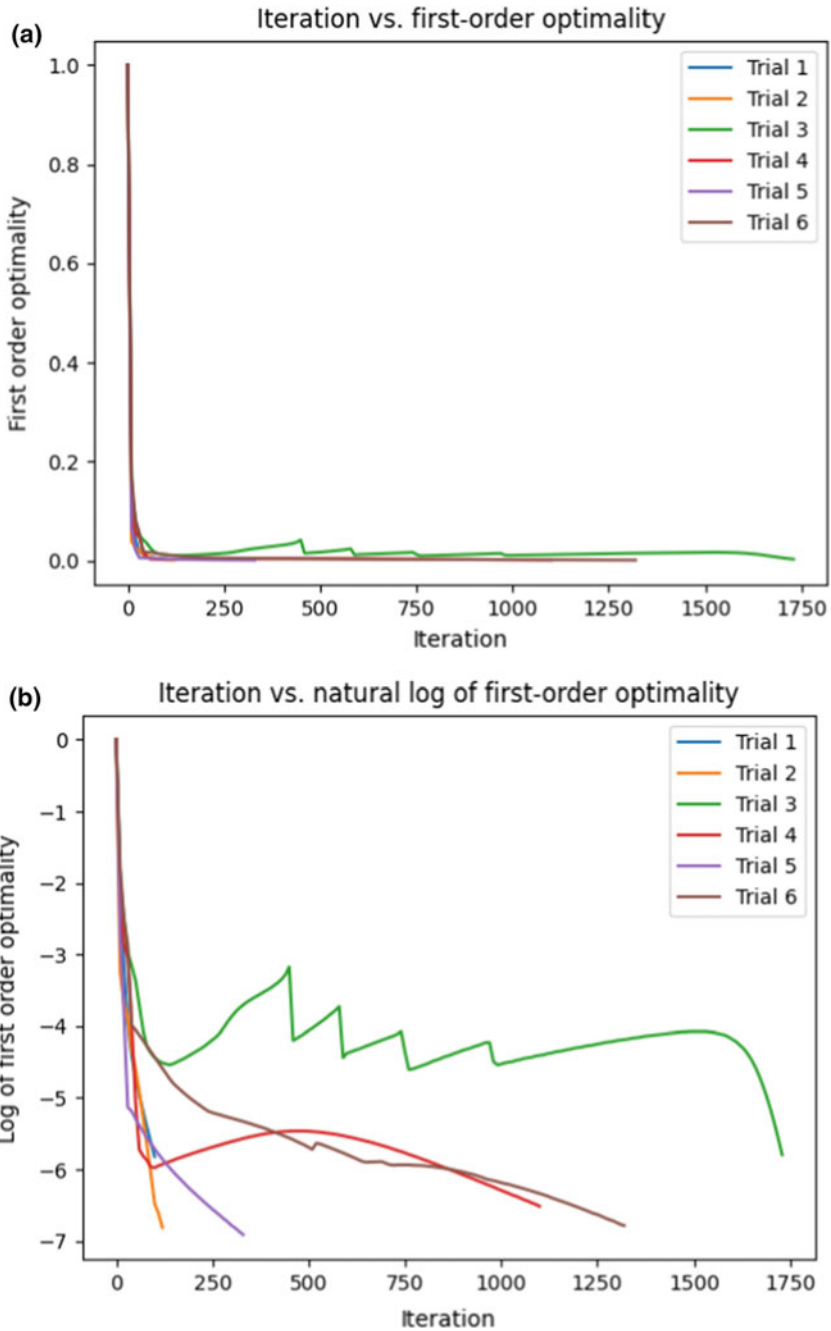


Fig. 13 Convergence plots for various tensor-Krylov runs. **a** Iteration versus first-order optimality. **b** Same algorithm iteration versus $\ln(\text{first-order optimality})$

$$\gamma_k = -j \ln \frac{2v_k}{E_{in}}. \quad (15)$$

E_{in} is calculated by averaging the input vector’s electric field amplitudes.

These phase angles and amplitude values are converted to distinct voltages to sent to the Tachyon accelerator through a lookup table.

5 Results

We demonstrate the efficacy of our core on matrix multiplication with matrices of various sizes, and we measure the efficacy for each size. The MPU stats are given in Table 2. Unlike most traditional digital electronic processors, the MPU performs an MVM in a single pass. Raw clock speed is 3 orders of magnitude higher than Google’s Tensor Processing Unit (TPU). The performance per watt is 245 TFLOPS/watt for the MPU’s optical core. Power draw does not scale dramatically with matrix size, leading to an estimated 261 TFLOPs/watt rating for a 256×256 optical core (the dimensions currently used within Google’s TPU v4).

The MPU core comes with several benefits. It can perform an entire ANN forward pass in a single clock cycle. In addition, it expels negligible heat due to its silicon photonic architecture. At scale, it is faster and more energy-efficient than specialized ML hardware: Google’s TPU v4 and GPUs across the board (NVIDIA’s A100, all GTX/RTX chips, AMD’s Radeon chips, etc.).

However, the overall Tachyon accelerator’s performance limiting factor is the surrounding electronic interconnect. We estimate a performance drop to approximately 13.4 KFLOPS and 1.6 KFLOPS/W with the cheap, off-the-shelf components used in research; the drop would be much lower with high quality, custom analog integrated circuits (ICs). In addition, due to possessing a very specialized circuit, the accelerator is not applicable (advantageously) to training neural networks without further component optimization.

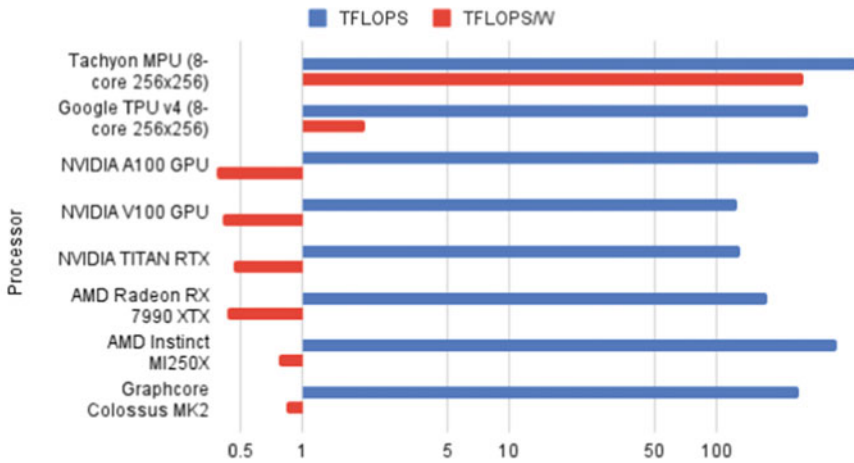
Table 2 Optical MPU core stats

Specification	Value (4×4)	Value (256×256)
Instructions	1 MVM/cycle	1 MVM/cycle
Clock	220 GHz	3.7 GHz
PPW	245 TFLOPS/watt	261 TFLOPS/watt
Performance	7 TFLOPS	478 TFLOPS

6 Conclusion

In this paper, we introduce Tachyon, a highly scalable optical hardware accelerator targeted toward matrix-vector multiplication (MVM). We go over the accelerator design, cost-efficient implementation, and programming interface. We then go over accelerator statistics, benefits, and scalability. We show that the accelerator is able to

256x256 MPU vs. Cutting-edge Deep Learning Hardware



4x4 MPU vs. Cutting-edge Deep Learning Hardware (Energy Efficiency)

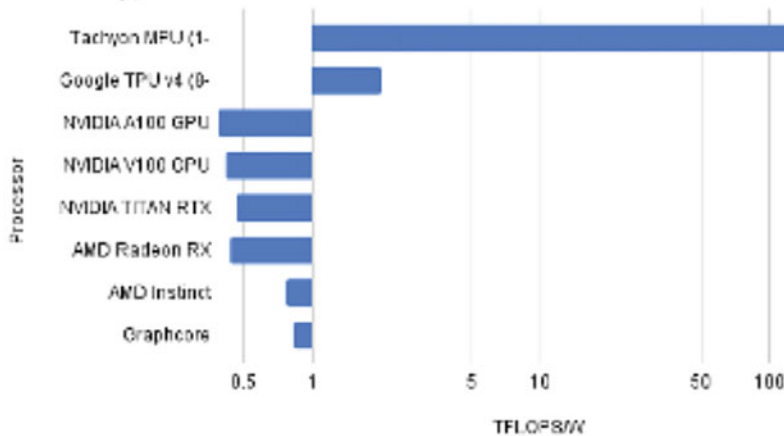


Fig. 14 **a** Scaled 256×256 MPU core outperforms TPU v4 and GPUs in FLOPS and FLOPS/W [17, 18]. **b** Tested 4×4 MPU core has an energy-efficiency similar to scaled version. However, the FLOPS rating is significantly below most deep learning hardware (7.0 TFLOPS) and is not shown

perform an MVM in a single cycle with a clock speed 3 orders of magnitude higher than Google's TPU v4 and a PPW 2 orders of magnitude higher, while still scaling very efficiently to higher matrix dimensions (Fig. 14).

The Tachyon accelerator, in using configured phase shifter settings to carry out computations, is best suited for MVM situations where the matrix is held constant and vectors are multiplied sequentially. This is the case in neural network inference, making the Matrix Processing Unit (MPU) highly tuned toward forward passes. Future research from a software perspective will focus on creating efficient compilers to map large models onto Tachyon to take advantage of this capability. Integration with packages in popular languages will also be investigated to allow an end user to easily access the accelerator's capabilities.

From a hardware perspective, future research will focus on developing integrated circuits along with optimizing the bottleneck between optics and electronics. In addition, further optical enhancements will be investigated, such as parallelization boosts through wavelength division multiplexing. Such enhancements are stepping stones toward our end goal of having the Tachyon accelerator as a product that consumers can use to advantageously accelerate their machine learning workloads.

Acknowledgements We thank the following individuals and organizations for their contributions to this project: Mr. Jonathan Hummel from the Maryland NanoCenter for printing initial designs with the Raith e-LiNE; Dr. Lukas Chrostowski and Dr. Iman Taghavi from the University of British Columbia for printing later designs with the JEOL JBX-6300F and performing our requested tests; Thorlabs Quantum Electronics for advice regarding design optimization; Dr. Yanne Chembo from the Institute for Research in Electronics and Applied Physics for advice on future development methodologies; Sherry Yu and Tarini Basireddy for fabrication advice; Mr. Mark Hannum for providing mathematical assistance and helpful resources; Vishal Nandakumar for recommending Newton-class approximation methods; Shihao Cao for support and useful discussion about the project's future; Mr. Hari Krishnan from NVIDIA for insightful discussion; and finally, our families for extended support and patience.

References

1. Garcia-Gasulla D, Peres F, Vilalta A. On the behavior of convolutional nets for feature extraction
2. Li C (2022) OpenAI's GPT-3 language model: a technical overview. Lambda, Inc. <https://lambdalabs.com/blog/demystifying-gpt-3>
3. Saul J, Bass D (2023) How much energy do AI and CHATGPT use? no one knows for sure. Bloomberg. <https://www.bloomberg.com/news/articles/2023-03-09/how-much-energy-do-ai-and-chatgpt-use-no-one-knows-for-sure>
4. Jouppi NP, Young C, Patil N, Patterson DA, Agrawal G, Bajwa R, Bates S, Bhatia S, Boden N, Borchers A, Boyle R, Cantin P, Chao C, Clark C, Coriell J, Daley M, Dau M, Dean J, Gelb B, Ghaemmaghami TV, Gottipati R, Gulland W, Hagmann R, Ho CR, Hogberg D, Hu J, Hundt R, Hurt D, Ibarz J, Jaffey A, Jaworski A, Kaplan A, Khaitan H, Koch A, Kumar N, Lacy S, Laudon J, Law J, Le D, Leary C, Liu Z, Lucke K, Lundin A, MacKean G, Maggiore A, Mahony M, Miller K, Nagarajan R, Narayanaswami, R, Ni, R, Nix, K, Norrie, T, Omernick, M, Penukonda, N, Phelps, A, Ross, J, Salek, A., Samadiani, E, Severn, C, Sizikov, G, Snelham, M, Souter J, Steinberg D, Swing A, Tan M, Thorson G, Tian B, Toma H, Tuttle E, Vasudevan V, Walter R,

- Wang W, Wilcox E, Yoon DH (2017) In-datacenter performance analysis of a tensor processing unit. CoRR [arXIV:abs/1704.04760](https://arxiv.org/abs/1704.04760) 1704.04760 (2017)
5. Cloud tensor processing units (TPUs) nbsp; nbsp; google cloud. Google. <https://cloud.google.com/tpu/docs/tpus>
 6. Moore's Law or how overall processing power for computers will double every two years. <http://www.moorelaw.org/>
 7. Cowan GER, Melville RC, Tsividis YP (2006) A VLSI analog computer/digital computer accelerator. IEEE J Solid-State Circ 41(1):42–53. <https://doi.org/10.1109/JSSC.2005.858618>
 8. Guo N, Huang Y, Mai T, Patil S, Cao C, Seok M, Sethumadhavan S, Tsividis Y (2016) Energy-efficient hybrid analog/digital approximate computation in continuous time. IEEE J Solid-State Circ 51(7):1514–1524. <https://doi.org/10.1109/JSSC.2016.2543729>
 9. M1076 analog matrix processor (2022). <https://mythic.ai/products/m1076-analog-matrix-processor/>
 10. Clements WR, Humphreys PC, Metcalf BJ, Kolthammer WS, Walmsley IA (2017) An optimal design for universal multiport interferometers (2017)
 11. PJRC: Teensy@4.1 Development Board. PJRC. <https://www.pjrc.com/store/teensy41.html>
 12. Texas Instruments (2016) 16/32 Bit arm cortex-R5F flash MCU, RISC, EMAC. Texas Instruments. Rev. C. <https://www.ti.com/lit/ds/symlink/rm571843.pdf?>
 13. Analog Devices (2005) 8-Bit, 210 MSPS TxDAC@D/A converter. Analog Devices. Rev. B. <https://www.analog.com/media/en/technical-documentation/data-sheets/AD9748.pdf>
 14. Analog Devices (1995) 250 MHz, General purpose voltage feedback Op amps gain 1 stable. Analog Devices. Rev. A. https://www.analog.com/media/en/technical-documentation/data-sheets/AD8047_8048.pdf
 15. Redrouthu S, Athavale R (2022) Tensor algebra on an optoelectronic microchip. <https://arxiv.org/abs/2208.06749>
 16. Bader BW, Pawlowski RP, Kolda TG (2005) Robust large-scale parallel nonlinear solvers for simulations. <https://www.osti.gov/servlets/purl/876345>
 17. Jouppi NP, Kurian G, Li S, Ma P, Nagarajan R, Nai L, Patil N, Subramanian S, Swing A, Towles B, Young C, Zhou X, Zhou Z, Patterson D (2023) TPU v4: an optically reconfigurable supercomputer for machine learning with hardware support for embeddings (2023)
 18. Index. <https://gadgetversus.com/>

Performance Analysis of Multi-user Cooperative Non-orthogonal Multiple Access on Time-Sharing Basis



Smriti Dubey, Neelesh Mehra, and Devendra Kumar Shakya

Abstract Non-orthogonal multiple access (NOMA) has taken place as one of the key empowering technique in 5G networks and beyond. For multiple users in the network, the work in this paper presented a cooperative non-orthogonal multiple access (C-NOMA) system based on the time-sharing model. An amalgamation of cooperative communication with NOMA alleviates the system's performance by improving diversity gains leading to superior reception reliability and user fairness. When the number of users increases within a network, then power distribution become challenging with C-NOMA system. In this study, a complete power allocation method using user clustering and one of the conventional multiple access technique that is time division multiple access (TDMA) techniques in the scenario of downlink C-NOMA networks is examined. While TDMA is used between different clusters, C-NOMA allows users within the same cluster to receive signals concurrently. Hybrid TDMA C-NOMA (HTCN) is the name given to the proposed user clustering with time slot allocation systems. Using the aforementioned methodologies, the outage probability and system throughput are thoroughly evaluated. Matlab simulation is used to put it into practice. In the current network, the proposed method outperforms simple NOMA and simple TDMA.

Keywords Non-orthogonal multiple access · Cooperative non-orthogonal multiple access · User clustering · Time division multiple access · Outage probability · System throughput

S. Dubey (✉)

Department of Information Technology, Samrat Ashok Technological Institute, Vidisha, Madhya Pradesh, India

e-mail: smritidubey.ec@satiengg.in

N. Mehra · D. K. Shakya

Department of Electronics and Communication, Samrat Ashok Technological Institute, Vidisha, Madhya Pradesh, India

1 Introduction

The multiple access technique is the core of any communication system. Using this method, numerous users can connect to and use the wireless network simultaneously. From generation to generation there are substantially diverse MA approaches have been used [1, 2]. MA has been classified into orthogonal and non-orthogonal multiple access. Two signals are said to be orthogonal when the dot product between them is zero. Orthogonal signaling is done in multiple access techniques like time division multiple access (TDMA), frequency division multiple access (FDMA), and orthogonal frequency division multiple access (OFDMA), etc. When the dot product between the signals are non-zero, then the signals are said to be non-orthogonal [3].

Among the orthogonal MA techniques, OFDMA has been widely adopted within wireless communication networks during 4G. NOMA is required to maintain the pace of growth of next-generation demands [4]. Massive connectivity, extremely low latency, and tremendous throughput must all be supported by the next-generation network. To improve system spectral efficiency, all the users are multiplexed with various power levels using the same time, frequency, and code. Multi-user interference can be avoided by successive interference cancellation (SIC) at the receiver end [5]. Hence, higher spectral efficiency can be achieved by the system as compared to the conventional multiple access techniques [6, 7].

Less power has been allocated to the user with good channel conditions and more power to the user with poor channel conditions. The SIC is done to decode its information [8]. Since the signal is broadcasted to the same frequency, consequently, the user with good channel conditions will decode the message intended for the other users in the network; such prior information can be utilized for the improvement in the diversity gain of the users [9]. The use of decoded information by the near user has been done in a cooperative NOMA (C-NOMA) transmission scheme [10]. The near user will act as a relay for the user with poor channel conditions. It increases achievable diversity [11].

Even while C-NOMA maximizes the diversity of users connected within the network, the complexity of the system may increase when the number of users increases in the network. It provides simultaneous service to several users at the same frequency, although the frequency's user capacity is finite. The network's overall throughput decreases as a result. Hybrid C-NOMA can be one of the solutions for the above-mentioned problem. It is a combination of C-NOMA with any OMA techniques like TDMA, FDMA, OFDMA, etc. In this work, C-NOMA is combined with TDMA. Hybrid TDMA C-NOMA (HTCN) divides users into different time slots to resolve the system complexity and reduces the processing time when the number of users exceeds within the network and is difficult to accommodate within the same frequency. To perform HTCN, first clustering of users is done to perform C-NOMA and then TDMA is applied to serve different user clusters. Network management can be done with clustering. According to this, all the users are divided into groups based on pre-defined criteria [12–14]. In [15], the effect of user pairing has been shown, where the ergodic sum capacity in two NOMA schemes is analyzed, keeping

the power allocation fixed and cognitive radio-inspired NOMA [16]. The reported C-NOMA scheme uses cell-centered users paired on a time-shared basis with two clustered cell edge users through a relay [9]. Energy harvested NOMA combined with TDMA has been presented in [17]. A hybrid TDMA NOMA in the cellular-enabled machine-to-machine networks has been discussed in [18]. The time allocation which is optimal for increasing the spectrum efficiency of TDMA-based wireless-powered communication networks and NOMA-based wireless-powered communication networks were compared in [19]. In downlink hybrid NOMA networks, a thorough user clustering, and power allocation technique has been examined [20], where TDMA strategies are used between various clusters, while NOMA techniques are used within the same cluster to allow users to receive signals simultaneously. To enhance the sum throughput of the network, a method has been described in [21] that makes use of the channel allocation coefficient difference across users in a NOMA cluster and organizes them into a single cluster or many clusters. The main contributions of the proposed work are summed up as follows:

1. The proposed HTCN scheme can accommodate more users without degradation in power allocation coefficient, sum throughput, and processing time. C-NOMA results increase in diversity gain.
2. Outage probability's closed-form expression of users for the proposed scheme has been derived and compared with Hybrid TDMA NOMA (HTN) and simple NOMA.
3. The proposed scheme's sum throughput has been derived and compared with the above-mentioned existing scheme.
4. Remarkable improvement has been noticed in the proposed scheme simulation results.

2 System Model

2.1 Architecture

As shown in Fig. 1, a broadcast channel with a single base station (BS) and N users has been considered.

Following are the assumptions made in the architecture:

1. BS represents the base station.
2. $P_1, P_2, P_3 \dots P_i$ are near users within the disc D_P with radius R_{D_P} .
3. $Q_1, Q_2, Q_3 \dots Q_j$ are the far users within the ring D_Q with radius R_{D_R} and R_{D_Q} (assuming $R_{D_R} \gg R_{D_P}$).
4. N represents the total number of users, such that $P_i + Q_j = N$.
5. Φ_κ ($\kappa \in \{P, Q\}$) with densities λ_{Φ_κ} , represents the modeled homogeneous poison point processes.

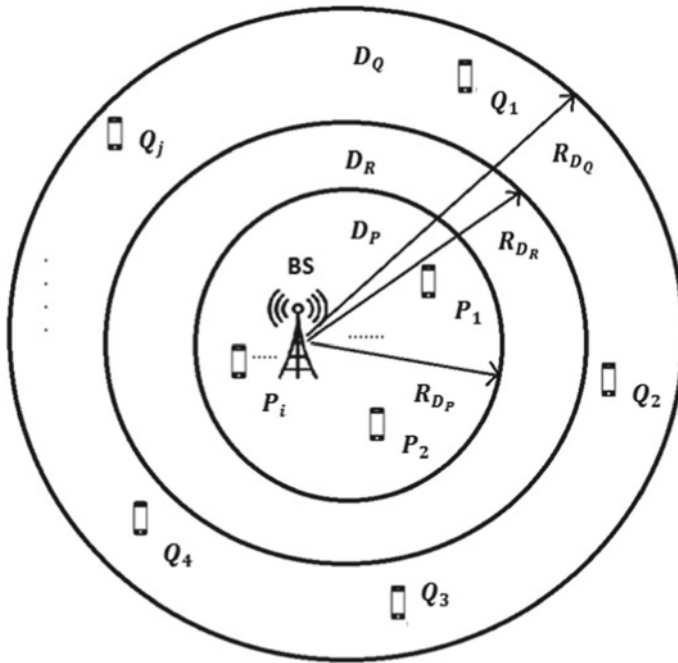


Fig. 1 The architecture of randomly deployed users, that is near users and far users

- 6. $|h_N|^2 \leq |h_{N-1}|^2 \leq \dots \leq |h_2|^2 \leq |h_1|^2$, where h denotes the channel coefficient of users.
- 7. g_1, g_2, \dots, g_N are the channel gain of users.

The randomly deployed users form two groups, i.e., near users $\{P_i\}$ and far users $\{Q_j\}$ such that $P_i + Q_j = N$. All the users within the group $\{P_i\}$ are assumed to be positioned within the disc D_P whose radius is R_{D_P} . The far users $\{Q_j\}$ are located within the circle D_Q whose radius is R_{D_R} and R_{D_Q} (assuming $R_{D_R} \gg R_{D_P}$). The basestation is located at the origin of both the disc D_P and the circle D_Q as shown in Fig. 1. The homogeneous poison point processes Φ_κ ($\kappa \in \{P, Q\}$) with densities λ_{Φ_κ} have been employed for the location of the users.

2.2 User Clustering

User clustering on the HTCN system has necessitated the employment of a less sophisticated and less effective technique. In [22], three clustering scheme has been discussed comprising random users, near users, and far users. In the first scheme

random users are selected from the two groups. In the second scheme, one nearest near and one nearest far has been selected from the two groups. It is abbreviated as NNNF. The third scheme comprises one nearest near and one farthest far from the two groups. It is abbreviated as NNFF. This user clustering system can be expanded to include numerous network users.

It is assumed that all the users are ordered-based. If h denotes the channel coefficient of the users present in the network. As per their distance from the base station, the ordered-based channel coefficient is assumed to be, $|h_N|^2 \leq |h_{N-1}|^2 \leq \dots \leq |h_2|^2 \leq |h_1|^2$. The nearest user has better channel condition due to its shortest distance from the base station, i.e., h_1 . Similarly, the farthest user has the lowest channel coefficient, i.e., h_N . The proposed scheme makes use of the channel gain differences, i.e., g_1, g_2, \dots, g_N .

Before user clustering, the total number of clusters and the total number of users per cluster are chosen. User clustering is carried out after the cluster size is chosen.

Let α denote the number of users in the group $\{P_i\}$ (near uses). As compared to the users in the group $\{Q_j\}$ (far users), it has much higher channel gains.

An algorithm for the user clustering is given as follows:

1. **Sort Users:**

$$g_1 > g_2 > \dots > g_\alpha > g_{\alpha+1} > g_{\alpha+2} > \dots > g_N$$

2. **Select no. of Clusters:**

- if $(\alpha < N/2)$.
then the number of clusters, $n = \alpha$;
- else if $(\alpha \geq N/2)$.
then the number of clusters, $n = N/2$.

3. **Pair Users into Clusters:**

- 1st cluster = $\{g_1, g_{n+1}, g_{2n+1}, \dots, g_N\}$,
- 2nd cluster = $\{g_2, g_{n+2}, g_{2n+2}, \dots, g_{N-1}\}, \dots$,
- n -th cluster = $\{g_n, g_{2n}, g_{3n}, \dots, g_{N-n-1}\}$.

4. **Cluster Size:**

- If $((N \bmod n) == 0)$.

If this is the case, cluster sizes will be uniform; otherwise, they will not.

Assuming 12 active users per cell, Fig. 2 illustrates the nearest near and nearest far user clustering with two, three, and four user clusters, respectively. Two users per cluster of NNNF users have been taken into consideration in this proposed study [5].

2.3 Time Allocation

In the proposed scheme, the TDMA has been applied to the proposed NNNF user clustering. The error propagation and system complexities seem to be reduced by the use of SIC [23]. Much research work has been carried out on TDMA and NOMA

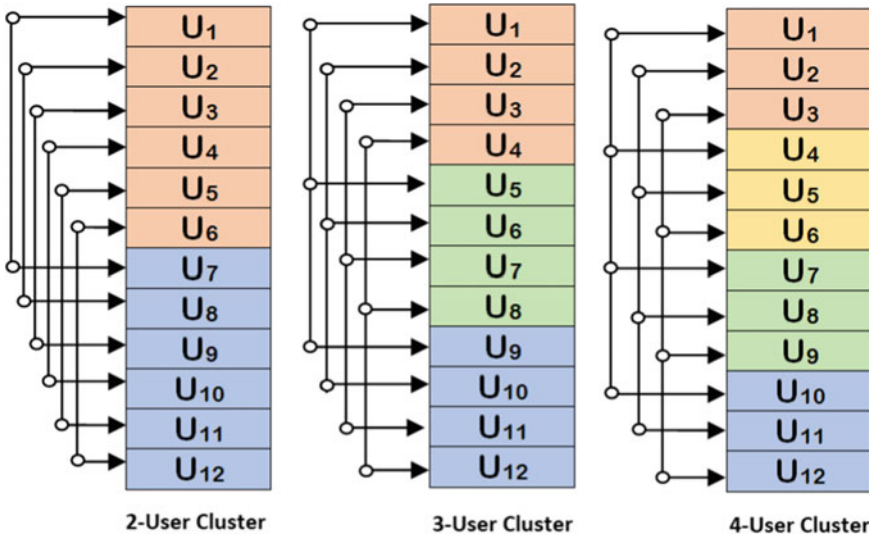


Fig. 2 A representation of 12 active users in a cell with C-NOMA clustering of 2 users, 3 users, and 4 users

[17, 24, 25]. Since the C-NOMA is integrated with TDMA, time allocation has been done.

Let the total time be T and it is distributed evenly among the clusters as demonstrated on the horizontal axis of Fig. 3. The figure also illustrates the power allocation on the vertical axis, where less power has been assigned to near users and more power to far users as per clusters. The number of users per cluster is given by m . For example in the b th cluster, m_b represents the total number of users.

$$\forall b \in \{1, 2, \dots, n\},$$

$$\text{such that } N = \sum_{b=1}^n m_b$$

In the figure, $P_{1,1}$ represents the first user of the first cluster and $P_{m_1,1}$ represents m_1 user of the first cluster. Note that t_b represents the time slot allocated to serve each cluster, such that $T = \sum_{b=1}^n t_b$ and $t_b = \frac{T}{n} \cdot t_b = \frac{T}{n}$.

2.4 Multiple Access

In the proposed network of C-NOMA, it has been considered that the near users, i.e., $\{P_i\}$ applied decode and forward (DF) strategy. In the DF strategy, after decoding

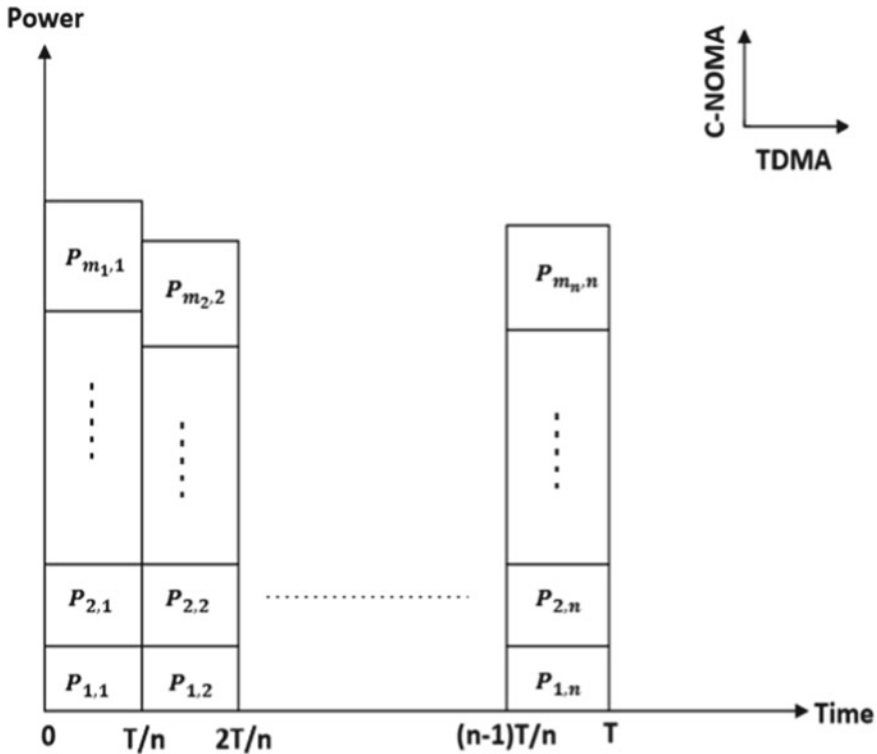


Fig. 3 Hybrid TDMA C-NOMA system time allocation

its own and another user’s (far user) information received from the BS, the near user will forward the signal to the far users [10, 22, 26].

It has been considered that all the nodes have a single antenna and will transfer the information in half-duplex mode. The channel coefficients are assumed to be ordered based and taken as follows:

$$|h_{Q_N}|^2 \leq |h_{Q_{N-1}}|^2 \dots \leq |h_{P_2}|^2 \leq |h_{P_1}|^2$$

where h_{Q_N} represents the farthest user in the given network and h_{P_1} as the nearest user as shown in Fig. 1. There are two phases in which cooperative NOMA works, which are direct phase and the cooperative phase. The BS broadcast signal $\sqrt{P_s a_1} x_P + \sqrt{P_s a_2} x_Q$ for all the users present per cluster. Where the total transmitted power has been taken as P_s , x_P , and x_Q , are the signals intended for near and far users, respectively. The power allocation coefficients are given by a_1 and a_2 , such that $a_1 > 0$, $a_2 > 0$, $a_2 > a_1$ and the $a_1 + a_2 = 1$. The received signals y_P and y_Q of each cluster at the near user and the far user can be expressed by Eqs. 1 and 2, respectively.

$$y_P = h_P \left(\sqrt{P_s a_1} x_P + \sqrt{P_s a_2} x_Q \right) + w_P \tag{1}$$

$$y_Q = h_Q \left(\sqrt{P_s a_1} x_P + \sqrt{P_s a_2} x_Q \right) + w_Q \tag{2}$$

where w_P and w_Q represents the additive white Gaussian noises. These noises have a variance σ^2 and the zero mean. The channel allocation coefficients are h_P and h_Q for the near user and the far users, respectively.

In the direct phase (d), considering near user signal x_P as noise, the received signal-to-interference noise ratio (SINR) $\gamma_Q^{x_Q}(d)$ for decoding x_Q at far user is given below in Eq. 3.

$$\gamma_Q^{x_Q}(d) = \frac{|h_Q|^2 \rho a_2}{|h_Q|^2 \rho a_1 + 1} \tag{3}$$

where $\rho = \frac{P_s}{\sigma^2}$ denotes the input SNR. During SIC the SINR of the near user is given by Eq. 4. Considering perfect SIC, the near user SNR is given by Eq. 5,

$$\gamma_P^{x_Q} = \frac{|h_P|^2 \rho a_2}{|h_P|^2 \rho a_1 + 1} \tag{4}$$

$$\gamma_P^{x_P} = |h_P|^2 \rho a_1 \tag{5}$$

In the cooperative phase (c), the near user will relay the signal of the far user as shown in Eq. 4. The channel coefficient between them is $|h_{PQ}|$. Equation 6 gives the signal-to-noise ratio (SNR) $\gamma_{PQ}^{x_Q}(c)$ at the far user.

$$\gamma_{PQ}^{x_Q}(c) = |h_{PQ}|^2 \rho \tag{6}$$

For comparing the proposed work with NOMA (no clustering), the channel coefficients are considered to be ordered base. Power allocation strategy differ in NOMA and C-NOMA. If the far user power allocation coefficient is a_N then its predeccessive user's power is given below

$$\begin{aligned} a_{N-1} &= \left(\frac{N-1}{N} \right) \times (1 - a_N); \\ a_{N-2} &= \left(\frac{N-1}{N} \right) \times (1 - (a_N + a_{N-1})); \\ a_1 &= (1 - (a_N + a_{N-1} + \dots + a_2)) \end{aligned}$$

where

$$a_1 + a_2 + \dots + a_N = 1 \tag{7}$$

Now, for all the users present in the network, the SNR can be given by β_{m_b} as shown in Eq. 8

$$\beta_{m_b} = \frac{|h_{m_b}|^2 \rho a_{m_b}}{|h_{m_b}|^2 \rho \sum_{b=1}^{b-1} a_{m_b} + 1} \quad (8)$$

Here, the number of clusters are b the number of users are m_b in b th cluster as given in eq. (a). Similarly, h_{m_b} and a_{m_b} represent the channel coefficient and power allocation coefficient of m_b th user, respectively.

3 Outage Probability and System Throughput

3.1 Outage Probability of the Far User in HTC/N

When the received power falls below the threshold level then the receiver is said to be in the outage [10]. To check the system performance, an outage probability is an important parameter. In the proposed model, let the desired data rate for near and far users be R_1 and R_2 , respectively. In direct and cooperative transmission, $R_Q(d)$ and $R_Q(c)$ represents the achievable data rates of far user respectively by Eqs. (9) and (10)

$$R_Q(d) = t_b * \log_2(1 + \gamma_Q^{x_Q}(d)) \quad (9)$$

$$R_Q(c) = t_b * \log_2(1 + \gamma_{PQ}^{x_Q}(c)) \quad (10)$$

where t_b is time per cluster as illustrated by Fig. 3. $\gamma_Q^{x_Q}(d) = \left(\frac{|h_Q|^2 \rho a_2}{|h_Q|^2 \rho a_1 + 1} \right)$ and $\gamma_{PQ}^{x_Q}(c) = |h_{PQ}|^2 \rho$ are the SINR and SNR of the far user respectively, as given by Eqs. 3 and 6. Now apply the selection combining at the far user receiver. Here the receiver selects the signal that was received with high SNR. Equation 11 represents the achievable data rate of the far user

$$R_Q = \log_2(1 + \max(R_Q(d), R_Q(c))) \quad (11)$$

The far user's outage probability can be given by

$$P_{\text{out}}^Q = P(R_Q < R_2)$$

3.2 Outage Probability of the Near User in the HTC� and HTN

The achievable data rate of the near user is given by Eq. 12, where the near user decodes the signal of the far user. Equation 13 shows the achievable data rate of the near user when it detects its signal.

$$R_p^{x_Q} = t_b * \log_2 \left(1 + \frac{|h_p|^2 \rho a_2}{|h_p|^2 \rho a_1 + 1} \right) \quad (12)$$

$$R_p^{x_p} = t_b * \log_2 (1 + |h_p|^2 \rho a_1) \quad (13)$$

Hence, the outage probability of the near user is given by Eq. 14

$$P_{\text{out}}^P = P \{ (R_p^{x_Q} < R_1) \cup (R_p^{x_p} < R_1) \} \quad (14)$$

3.3 Outage Probability of the Far User in HTN

Since there is no cooperative phase involved in HTN, Eq. 15 represents the achievable data rate of the far user through direct transmission and Eq. 16 represents its outage probability, respectively.

$$R_Q = t_b * \log_2 (1 + R_Q(d)) \quad (15)$$

$$P_{\text{out}}^Q = P(R_Q < R_2) \quad (16)$$

3.4 Outage Probability of the Far User in NOMA

The achievable data rate of the far user and its outage probability are given below as Eqs. 17 and 18, respectively

$$R_{Q_{m,n}} = \log_2 (1 + \beta_{m_b}) \quad (17)$$

$$P_{\text{out}}^{Q_{m,n}} = P(R_{Q_{m,n}} < R_2) \quad (18)$$

3.5 System Throughput

System throughput is the amount of data received by the destination [25]. The average throughput is the throughput per unit of time. If the desired data rate is given, then the system throughput of the proposed scheme is evaluated as the summation of data successfully received at the near users and the far user.

The sum rate of HTCN per cluster can be evaluated as

$$\text{mean} \left(t_b * \log_2(1 + \max(R_Q(d), R_Q(c))) + t_b * \log_2(1 + |h_P|^2 \rho a_1) \right)$$

The sum rate of HTN per cluster can be evaluated as

$$\text{mean} \left(t_b * \log_2 \left(1 + \frac{|h_Q|^2 \rho a_2}{|h_Q|^2 \rho a_1 + 1} \right) + t_b * \log_2(1 + |h_P|^2 \rho a_1) \right)$$

Similarly, the sum rate of NOMA per user can be evaluated as

$$\text{mean}(\log_2(1 + \beta_{m_b}))$$

4 Numerical Results and Discussion

To justify the analysis done in the above points and explore the operation of various system parameters, such as the desired data rate pair, input SNR, and path loss exponent, on the system performance has been carried out. The system parameter setup for four users comprising two clusters is as follows: path loss exponent (ε) = 2, d_1, d_2, d_3 , and d_4 , are the distances of the links BS $\rightarrow P_1$, BS $\rightarrow P_2$, BS $\rightarrow Q_1$ and BS $\rightarrow Q_2$, respectively. The distance of the links has been set to $d_1 = 1$ m, $d_2 = 3$ m, $d_3 = 10$ m, and $d_4 = 11$ m. The distance between the link $P_1 \rightarrow Q_1$ has been calculated by $\sqrt{(d_1^2 + d_3^2 - 2d_1d_3 \cos \theta)}$. Where θ represents the angle between the two users and BS is the vertex. Similarly with link $P_2 \rightarrow Q_2$. The power allocation coefficients for HTCN and HTN are set to $a_1 = 0.2$ and $a_2 = 0.8$, i.e., fixed power allocation, but in NOMA the power allocation coefficient for the far user is set to $a_4 = 0.8$ and Eq. 7 represent the remaining coefficients which are successively decreasing.

Bit per channel use (BPCU) is considered as a unit for the desired data rate. In Fig. 4, an outage probability of far users is plotted with the desired data rate $R_1 = 0.2$ and $R_2 = 0.8$ and $\varepsilon = 2$. The performance of HTCN is much better than the existing scheme of HTN and NOMA. At 15 dB SNR, all the techniques are giving almost the same outage probability. As the SNR increases, the outage probability of HTCN techniques is decreasing considerably. As Table 1 describes, it is around 10^{-4}

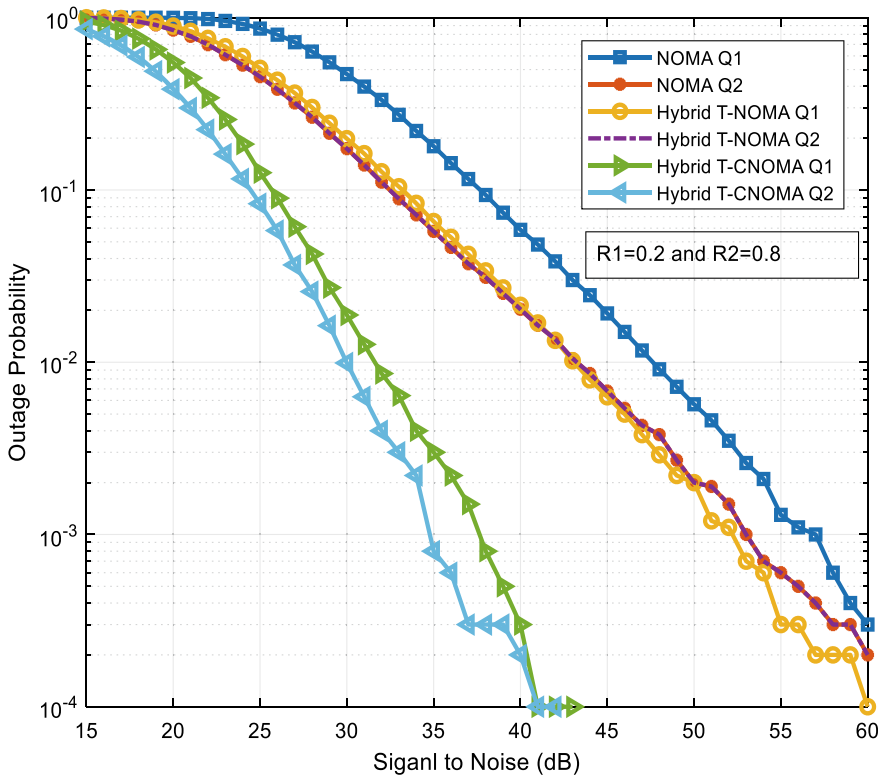


Fig. 4 An outage probabilities of the far user in case of HTCN, HTN, and NOMA with target data

after 40 dB but other techniques are after 60 dB. Hence, in the proposed method, the outage probability is less even with less signal power.

In Fig. 5, on increasing the desired data rate to $\{R_1 = 0.5 \text{ and } R_2 = 1.5\}$ the outage probability performance of HTCN is still better than HTN and NOMA. Table 2 describes the comparison of SNR and an outage probability of different schemes for increasing the desired data rate. The outage probability of HTCN near user goes to 10^{-4} at 43 dB. Similarly, HTCN far user at 54 dB. On other existing schemes of HTN and NOMA, outage increases considerably.

In Fig. 6, the system throughput of different schemes with $\epsilon = 2$ is compared. It is increasing in HTCN, as compared to the existing scheme. From Table 3, it can be seen that at 60 dB HTCN's throughput is 28.17, which is higher than HTN and NOMA.

Table 1 The comparison of outage probabilities of the far user in the case of HTCN, HTN, and NOMA with $R_1 = 0.2$ and $R_2 = 0.8$

SNR (dB)	NOMA Q_1	NOMA Q_2	HTN Q_1	HTN Q_2	HTCN Q_1	HTCN Q_2
15	0.9967	0.9967	0.9967	0.9967	0.9967	0.9967
20	0.9967	0.9967	0.9967	0.9967	0.7874	0.9055
25	0.9927	0.9927	0.9927	0.9927	0.4692	0.6095
30	0.9476	0.7664	0.8203	0.7664	0.1107	0.1755
35	0.609	0.375	0.419	0.375	0.0199	0.0304
40	0.2543	0.1361	0.1583	0.1361	0.0018	0.0036
45	0.0898	0.0444	0.0548	0.0444	0.0001	0.0005
50	0.0313	0.0137	0.0184	0.0137	0.0001	0.0002
55	0.0114	0.0033	0.0066	0.0033	0.0001	0.0001
60	0.0035	0.0015	0.002	0.0015	0.0001	0.0001

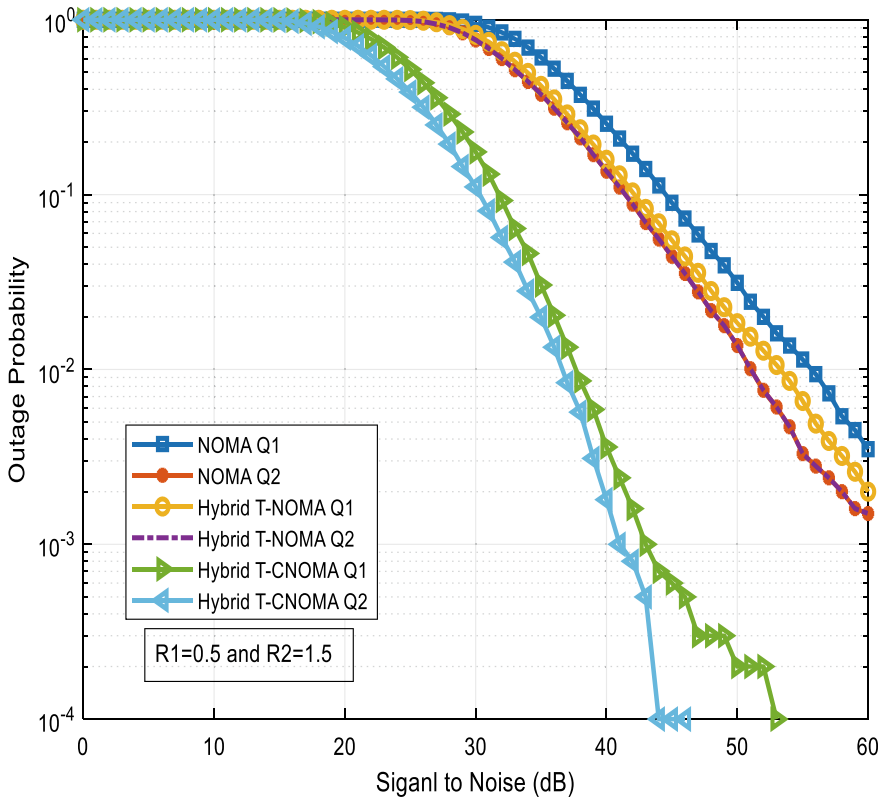


Fig. 5 An outage probability of HTCN, HTN, and NOMA with an increased target data rate

Table 2 The comparison of outage probabilities of the far user in the case of HTCEN, HTN, and NOMA with $R_1 = 0.5$ and $R_2 = 1.5$

SNR (dB)	NOMA Q_1	NOMA Q_2	HTN Q_1	HTN Q_2	HTCN Q_1	HTCN Q_2
15	0.9994	0.9994	0.9994	0.9994	0.9499	0.8588
20	0.9983	0.8474	0.8944	0.8474	0.5493	0.3847
25	0.8647	0.4549	0.5084	0.4549	0.126	0.0832
30	0.4707	0.1737	0.1985	0.1737	0.0188	0.0099
35	0.1794	0.0575	0.0656	0.0575	0.003	0.0008
40	0.0587	0.0203	0.0214	0.0203	0.0003	0.0002
45	0.0192	0.0068	0.0063	0.0068	0.0001	0.0001
50	0.0057	0.002	0.002	0.002	0.0001	0.0001
55	0.0013	0.0006	0.0003	0.0006	0.0001	0.0001
60	0.0003	0.0002	0.0001	0.0002	0.0001	0.0001

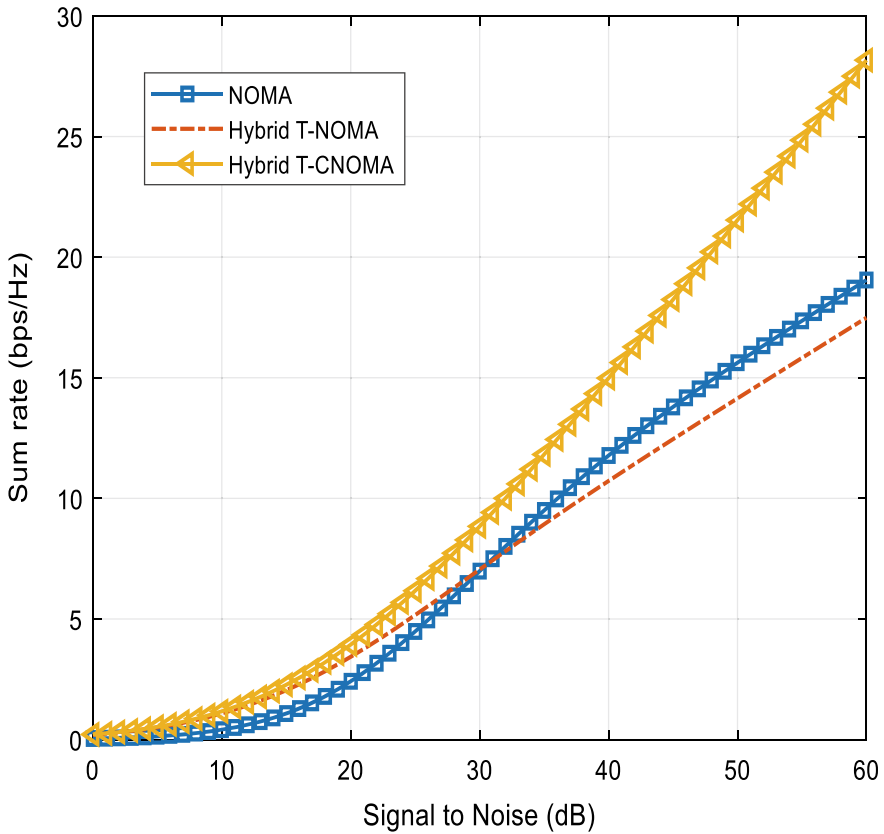


Fig. 6 The sum rate comparison of HTCEN, HTN, and NOMA with different path loss exponents (ϵ)

Table 3 The sum rate comparison of HTCN, HTN, and NOMA

SNR (dB)	NOMA	HTN	HTCN
10	0.4103	1.171	1.171
20	2.417	3.432	3.94
30	7.051	7.051	8.84
40	11.78	10.72	14.98
50	15.62	14.15	21.54
60	19.05	17.49	28.17

5 Conclusion

In the work, a new model of HTCH has been proposed. The multiple users are arranged by a stochastic geometric approach. The complete framework which represents the location of near users and far users has been shown. The performance has been evaluated. A closed-form expression of the outage probabilities and system throughput has been derived. Numerical results validate the analysis and manifest that the performance of the proposed scheme is better. The achievement of higher system throughput is considerable and there is a decrease in outage probability. The parameters of the networks are chosen carefully like the coefficient of power allocation, the desired data rate, path loss exponent, etc. Now, when the number of users increases within the network, the proposed model can perform well. This work can be extended with relay-aided TDMA C-NOMA communications and multicell TDMA C-NOMA communications. As energy harvesting is a major concern of communication nowadays, this work can also be extended with Simultaneous Wireless Information and Power Transfer, i.e., SWIPT.

References

1. Del Peral-Rosado JA, Raulefs R, López-Salcedo JA, Seco-Granados G (2018) Survey of cellular mobile radio localization methods: from 1G to 5G. *IEEE Commun Surv Tutor* 20(2):1124–1148. <https://ieeexplore.ieee.org/document/8226757>
2. Kebede T, Wondie Y, Steinbrunn J, Belay H, Kornegay K (2022) Multi-carrier waveforms and multiple access strategies in wireless networks: performance, applications, and challenges. *IEEE Access* 1–1. <https://doi.org/10.1109/access.2022.3151360>
3. King C (2014) Fundamentals of wireless communications. In: 67th annual conference for protective relay engineers, CPRE, pp 470–474. <https://ieeexplore.ieee.org/document/6799022>
4. Islam SMR, Avazov N, Dobre OA, Kwak KS (2017) Power-domain non-orthogonal multiple access (NOMA) in 5G systems: potentials and challenges. *IEEE Commun Surv Tutor* 19(2):721–742. <https://dl.acm.org/doi/10.1109/COMST.2016.2621116>
5. Liu Y, Ding H, Shen J, Xiao R, Yang H (2019) Outage performance analysis for SWIPT-based cooperative non-orthogonal multiple access systems. *IEEE Commun Lett* 23(9):1501–1505. <https://ieeexplore.ieee.org/document/8744551>

6. Dai L, Wang B, Yuan Y, Han S, CLI, Wang Z (2015) Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends. *IEEE Commun Mag* 53(9):74–81. <https://ieeexplore.ieee.org/document/7263349>
7. Vaezi M, Ding Z, Vincent Poor H (2018) Multiple access techniques for 5G wireless networks and beyond
8. Ding Z et al (2017) Application of non-orthogonal multiple access in LTE and 5G networks. *IEEE Commun Mag* 55(2):185–191. <https://ieeexplore.ieee.org/document/7842433>
9. Huang R et al (2020) Performance analysis of NOMA-based cooperative networks with relay selection. *China Commun* 17(11):111–119. <https://ieeexplore.ieee.org/document/9267801>
10. Ding Z, Peng M, Poor HV (2015) Cooperative non-orthogonal multiple access in 5G systems. *IEEE Commun Lett* 19(8):1462–1465. <https://ieeexplore.ieee.org/abstract/document/7117391>
11. Abbasi O, Ebrahimi A, Mokari N (2019) NOMA inspired cooperative relaying system using an AF relay. *IEEE Wirel Commun Lett* 8(1):261–264. <https://ieeexplore.ieee.org/abstract/document/8462722>
12. Al-Hussaibi WA, Ali FH (2019) Efficient user clustering, receive antenna selection, and power allocation algorithms for massive MIMO-NOMA systems. *IEEE Access* 7:31865–31882. <https://ieeexplore.ieee.org/document/8654616>
13. Prabha Kumaresan S, Tan CK, Ng YH (2020) Efficient user clustering using a low-complexity artificial neural network (ANN) for 5G NOMA systems. *IEEE Access* 8:179307–179316. <https://ieeexplore.ieee.org/document/9208731>
14. Rajasekaran AS, Maraga O, Sokun HU, Yanikomeroglu H, Al-Ahmadi S (2020) User clustering in mm wave-NOMA systems with user decoding capability constraints for B5G networks. *IEEE Access* 8:209949–209963. <https://ieeexplore.ieee.org/document/9264161>
15. Ding Z, Yang Z, Fan P, Poor HV (2014) On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. *IEEE Signal Process Lett* 21(12):1501–1505. <https://ieeexplore.ieee.org/document/6868214>
16. Ding Z, Fan P, Poor HV (2016) Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions. *IEEE Trans Veh Technol* 65(8):6010–6023. <https://ieeexplore.ieee.org/document/7273963>
17. Al-Obiedollah H et al (2019) On energy harvesting of hybrid TDMA-NOMA systems. *IEEE global communication conference, GLOBECOM 2019—proceedings*, pp 0–5. <https://ieeexplore.ieee.org/document/9013805>
18. Li Z, Gui J (2019) Energy-efficient resource allocation with hybrid TDMA-NOMA for cellular-enabled machine-to-machine communications. *IEEE Access* 7:105800–105815. <https://ieeexplore.ieee.org/document/8778641>
19. Wu Q, Chen W, Ng DWK, Schober R (2018) Spectral and energy-efficient wireless powered IoT networks: NOMA or TDMA? *IEEE Trans Veh Technol* 67(7):6663–6667. <https://ieeexplore.ieee.org/document/8275025>
20. Wang K, Liang W, Yuan Y, Liu Y, Ma Z, Ding Z (2019) User clustering and power allocation for hybrid non-orthogonal multiple access systems. *IEEE Trans Veh Technol* 68(12):12052–12065. <https://ieeexplore.ieee.org/document/8873605>
21. Ali MS, Tabassum H, Hossain E (2016) Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems. *IEEE Access* 4:6325–6343. <https://ieeexplore.ieee.org/document/7557079>
22. Liu Y, Ding Z, Elkashlan M, Poor HV (2016) Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer. *IEEE J Sel Areas Commun* 34(4):938–953. <https://ieeexplore.ieee.org/document/7445146>
23. Zafar A, Shaqfeh M, Alouini MS, Alnuweiri H (2013) On multiple users scheduling using superposition coding over rayleigh fading channels. *IEEE Commun Lett* 17(4):733–736. <https://ieeexplore.ieee.org/document/6464495>
24. Wei X et al (2020) Resource allocation technique for hybrid TDMA-NOMA system with opportunistic time assignment. *IEEE international conference communication work. ICC Work 2020—proceedings*, pp 1–6. <https://ieeexplore.ieee.org/document/9145342>

25. Zhang D, Wu Q, Cui M, Zhang G, Niyato D (2021) Throughput maximization for IRS-assisted wireless powered hybrid NOMA and TDMA. *IEEE Wirel Commun Lett* 10(9):1944–1948. <https://ieeexplore.ieee.org/document/9448351>
26. Ding Z, Krikidis I, Sharif B, Poor HV (2014) Wireless information and power transfer in cooperative networks with spatially random relays. *IEEE Trans Wirel Commun* 13(8):4440–4453. <https://ieeexplore.ieee.org/document/6779694>

TL-Shaped Circular Parasitic Compact Planar Antenna for 5G Microwave Applications



Ridhi Gupta, Manish Varun Yadav, and Swati Varun Yadav

Abstract TL-Shaped Circular Parasitic Compact Planar Antenna for 5G microwave applications is proposed and investigated. The current structure is complex and size efficient. The antenna's front side consists of rectangular patch in center that consists of L-shaped, T-shaped and inverse T-shaped elements, respectively, with four parasitic rings on the top and microstrip line in center fed up by 50Ω strip line. Backside of antenna consists of three parasitic rings on the top, four horizontal strips arranged in increasing order and ground plane with small rectangular slot in it. Using CST Microwave Studio, antenna simulations are performed on the FR-4 substrate with the overall dimensions of $13 * 15 * 1.5 \text{ mm}^3$. The antenna offers an impedance bandwidth of 60.8% and a return loss of -28 dB , with a frequency range of 3.2–6 GHz and a central frequency of 4.6 GHz. Radiation patterns from the proposed structure are constant throughout the operational range. Proposed antenna having stable polar patterns and efficient performance. The antenna is appropriate for 5G microwave communication applications since it has a wide gain of 4.81dBi and an efficiency of 82%.

Keywords 5G antenna · Microwave application · High efficiency · Impedance bandwidth · High gain

R. Gupta · M. V. Yadav (✉)

Department of Aeronautical and Automobile Engineering, Manipal Institute of Technology,
Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
e-mail: yadav.manish@manipal.edu

R. Gupta

e-mail: ridhi.gupta@learner.manipal.edu

S. V. Yadav

Department of Instrumentation and Control Engineering, Manipal Institute of Technology,
Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
e-mail: yadav.swati@manipal.edu

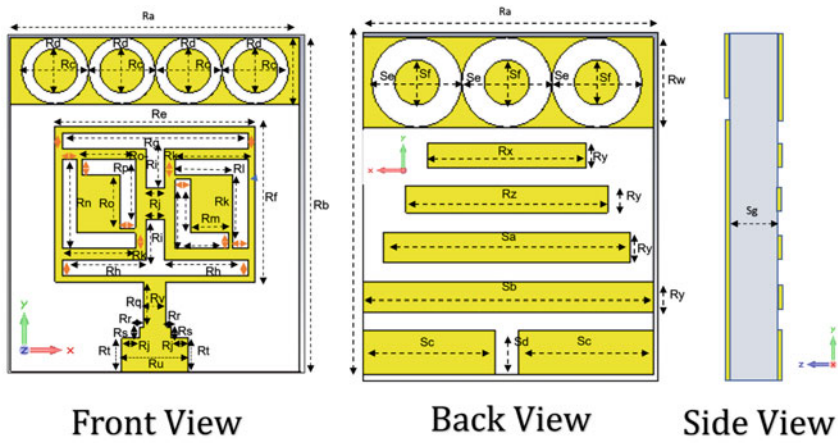
1 Introduction

Antennas play a crucial role in aerospace communication by transmitting and receiving radio signals between aircraft, satellites, and ground stations. A wide range of antennas, including semi-circular, L-shaped, triangular, multi-slot, and U-shaped antennas, are used in aerospace communication [1–6]. Techniques like changing the ground plane with parasitic elements, adding a small fractal component, U-shaped or circular slots, and circles are discussed to enhance their performance [7–11]. Polar radiation patterns can be created by altering the circular patch or by cutting a rectangular piece [12–14]. Previous research has shown that a wideband aerospace band can be created by changing a patch with a defective backside and semi-circular slots [15, 16]. As stated in references, altering the backside by modifying slots leads to a patch element with a defective plane [2, 17]. According to studies [18, 19], one method involved the use of a flexible antenna and “dumbbell-shaped” radiator. A long strip patch with cut rear slots is another design modification noted in reference [20], which improves current flow. The literature presents various design options for planar antennas. To generate balanced radiation patterns, one suggestion is to use a modified backside with precisely trimmed circular holes, as shown in references [21, 22]. The addition of a rectangular patch to the edges increases oscillations, as per reference [23, 24]. A fractal antenna design has been demonstrated to achieve high gain, as stated in reference [25]. Impedance matching reported through the design of multi-slots [26], The L-shaped design has been shown to reduce band resonance, as stated in reference [27]. More information on the antenna design, development, and simulated outcomes is provided in other sections of the paper.

2 Geometry and Design Principle

Figure 1 illustrates the geometry and concept for the design of an antenna. Substrate used in the antenna design is FR-4 inscribed on copper annealed with overall dimensions as 13 mm * 15 mm * 1.5 mm. The R_a and R_b represent the width and length of antenna, respectively. The parameter values for the design are displayed in Table 1 (given in mm). The length ‘ R_q ’ and width ‘ R_v ’ of the microstrip line are 2 mm and 1.5 mm, respectively, and are connected to the substrate’s thickness, denoted by ‘ S_g ’. The front part of the antenna is created by parasitic circular rings, rectangular patch, T, and inverse T with microstrip line. The outer and inner diameters of circular parasitic circular ring are denoted by ‘ R_c ’ and ‘ R_d ’. The length and width of inner rectangular patch are defined as ‘ R_e ’ and ‘ R_f ’. With a partial ground plane and a rectangular slot, the antenna’s back portion has ‘ S_d ’ and ‘ S_c ’ as its length and width, respectively. Backside also consists of four horizontal strips arranged in increasing order with width as ‘ R_x ’, ‘ R_z ’, ‘ S_a ’, and ‘ S_b ’, respectively, with width denoted by ‘ R_y ’. Three parasitic rings are placed on the top of the back plane and it is denoted by

“Se” and “Sf” as outer and inner radii of the circular rings. Parasitic elements help to resonant the lower order band.



Parameters	Ra	Rb	Rc	Rd	Re	Rf	Rg	Rh	Ri	Rj	Rk	l1	Rm	Rn	Sc	Sd	Se
Values (mm)	13	15	3	2	9	7	8.4	3.8	1.8	0.8	3.3	2.6	1.7	4	6	2	4
Parameters	Ro	Rp	Rq	Rr	Rs	Rt	Ru	Rv	Rw	Rx	Ry	Rz	Sa	Sb	Sf	Sg	↔
Values (mm)	2.4	3.1	2	0.2	0.5	1.5	3	1.4	4	7	1.3	9	11	13	2	1.5	0.7

Fig. 1 Proposed antenna structure

Table 1 Comparison of published planar antennas

Refs.	Band obtained (GHz)	Peak Gain (dBi)	Fractional B/W (%)	Peak (η) (%)	Overall volume (in λ)
[2]	3.1–22	1.7	150	NA	$0.28\lambda * 0.25\lambda * 0.016\lambda$
[2]	3.1–11	5.1	110	89	$0.20\lambda * 0.25\lambda * 0.015\lambda$
[6]	3.9–14	3.5	142	75	$0.26\lambda * 0.26\lambda * 0.019\lambda$
[8]	2–9	4.5	127	62	$0.33\lambda * 0.22\lambda * 0.1\lambda$
[12]	3.5–19	3.2	145	81	$0.23\lambda * 0.23\lambda * 0.015\lambda$
[15]	3.1–11	2	109	60	$0.55\lambda * 0.41\lambda * 0.022\lambda$
[17]	2.9–16	5.2	139	87	$0.33\lambda * 0.24\lambda * 0.014\lambda$
[22]	2.8–12	2.79	122	72	$0.18\lambda * 0.14\lambda * 0.15\lambda$
[25]	2.7–7.3	2.3	108	78.3	$0.32\lambda * 0.2\lambda * 0.014\lambda$
[26]	3.1–11	2.2	110	69	$0.14\lambda * 0.18\lambda * 0.015\lambda$
[27]	2.3–11	2.1	129	70	$0.2\lambda * 0.3\lambda * 0.014\lambda$
Presented	3.2–6	4.81	60.8	82	$0.13\lambda * 0.16\lambda * 0.016\lambda$

3 Result and Discussion

Based on an analysis of their impacts on the S11 parameter and other simulated results, the proposed structure optimizes the result.

Figure 2 is showing simulated return loss of the proposed antenna. The antenna's gain is another important parameter that indicates the strength of the transmitted signal. The maximum radiation effectiveness is shown within 3.3–5.2 GHz frequency range. The proposed design's simulated antenna efficiency and gain vs. frequency curve is depicted in Fig. 3. Peak gain, with a value of 4.81 dB, is seen at 4.5 GHz. Maximum radiation efficiency of antenna is observed at 4.25 GHz having a value of 0.82 or an efficiency of 82%. As the frequency rises, ohmic losses increase, leading to a gradual decline in antenna efficiency at higher frequencies.

Fig. 2 Simulated return loss

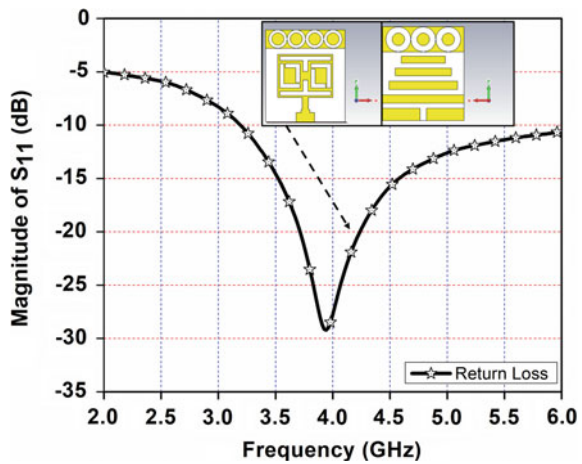


Fig. 3 Gain and antenna efficiency curve

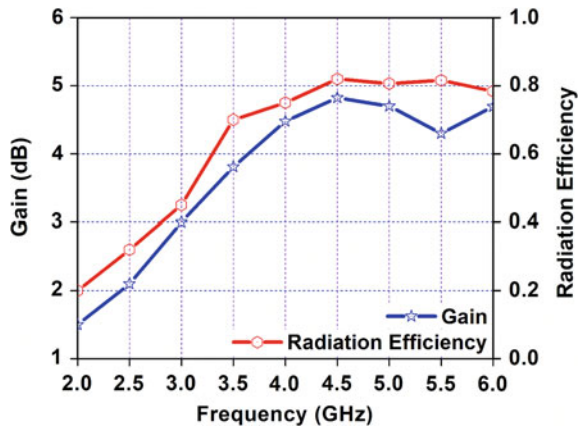


Fig. 4 I/P impedance curve

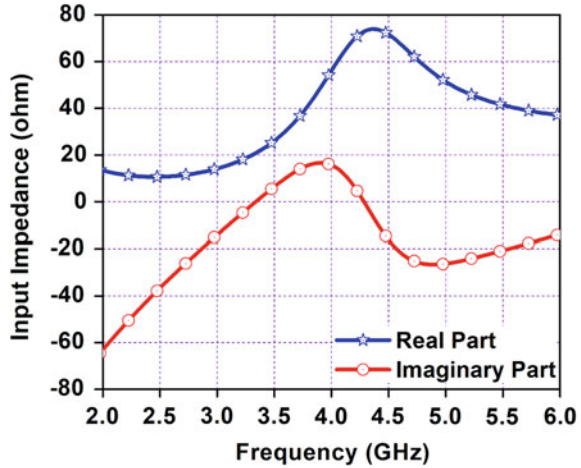


Fig. 5 Radiation pattern

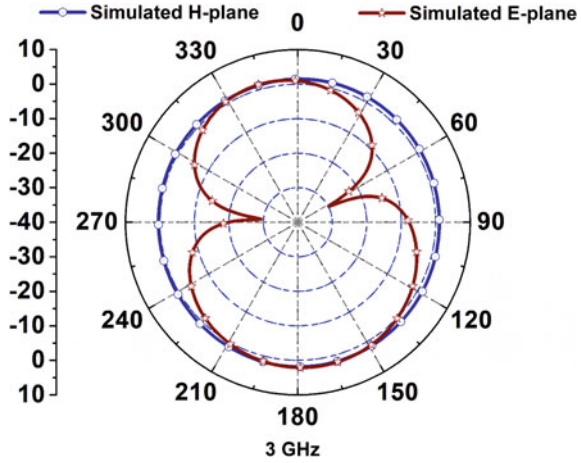


Figure 4 displays an I/P impedance curve that reflects the frequency curve for the proposed design’s simulated real and imaginary components. The behavior is inductive within the 3.2–4.2 GHz frequency range, followed by a capacitive behavior. The inductive behavior is represented by a positive polarity, whereas a negative polarity indicates capacitive behavior. The curve illustrates the impedance normalized to 50 Ω.

Figure 5 depicts the 5G antenna’s 3.5 GHz polar pattern, in two orthogonal coordinates: both H- and E-plane patterns. The radiation pattern improves the ability to identify the radiation pattern at a specific frequency by displaying measured and simulated H- and E-plane radiation patterns. Two perpendicular planes depict the radiation pattern, one of which represents the E-field and is oriented at a 90-degree angle to the YOZ plane, and the other of which represents the H-field and is oriented

at a 0-degree angle to the XOZ plane. The radiation pattern remains stable, and the antenna is coherent.

The current field distribution of the 5G antenna is presented in the paper. Figure 6 displays the current distributions from a front-to-rear perspective. The equivalent surface of the antenna represents its radiated current field and highlights its signal strengths, with the ability to generate both E-field and H-field, indicating significant signal strength in the design. Depicts the simulated radiation lobe pattern of the 5G antenna. The front plane of the structure and the frequency distribution of the 2D-vector surface current (at 3 GHz) are shown in Fig. 7. The radiated field will closely mirror the surface current of the original antenna by substituting similar surface currents for the antenna. This indicates that even in the absence of the main current source, the surface current can help to detect radiation.

Comparison table with previously published antenna is shown below, and our proposed TL-Shaped Circular Parasitic Compact planar antenna for 5G microwave applications is satisfy the requirement of all antenna parameters.

4 Conclusion

This research proposes a circular TL-shaped parasitic planar antenna for 5G applications. The proposed planar antenna has been measured, and the findings were subjected to a comprehensive analysis having dimensions of $13 * 15 * 1.5 \text{ mm}^3$. Efficiency for the suggested radiator is 82% and a high gain of 4.81 dBi. Impedance bandwidth of design is 60.8%, a 4.5 GHz center frequency, and a 3.2–6 GHz frequency range. There is an intense radiation signal visible in the E-/H-field current distribution. Suggested design has totally stable radiation pattern. The distribution of surface currents shows a significant signal. This ‘TL’-shaped antenna is suitable for 5G microwave communication applications.

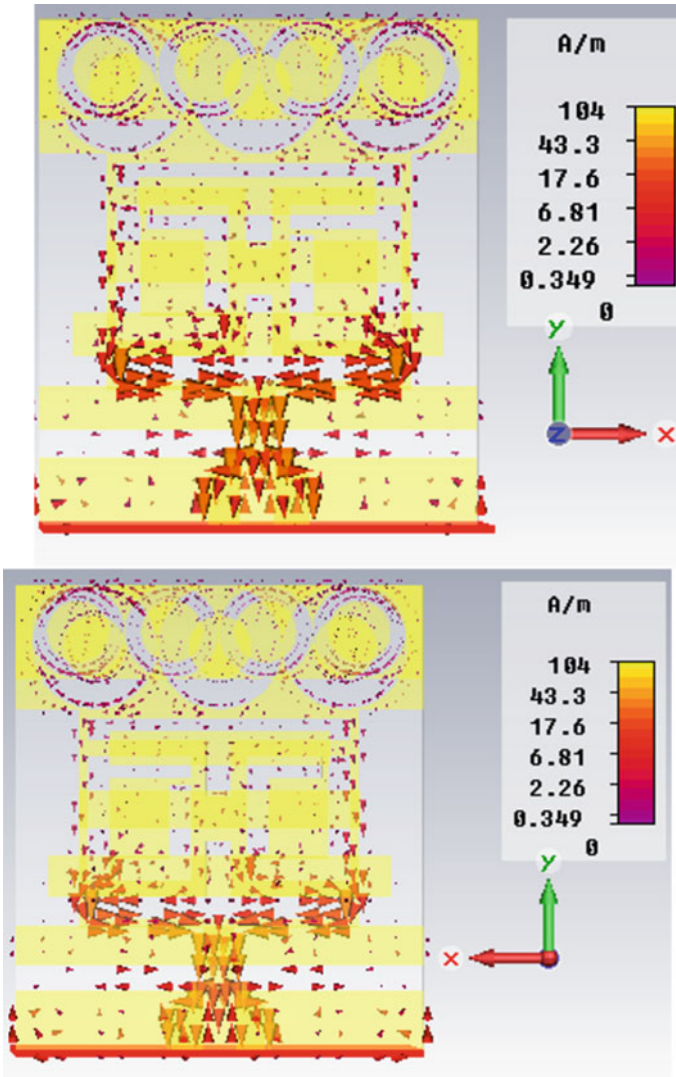


Fig. 6 Surface current distribution

References

1. Mishra B, Verma RK, Yashwanth N, Singh RK (2021) A review on microstrip patch antenna parameters of different geometry and bandwidth enhancement techniques. *Int J Microwave Wirel Technol* 1–22
2. Baudha S, Yadav MV (2019) A novel design of a planar antenna with modified patch and defective ground plane for ultra-wideband applications. *Microw Opt Technol Lett* 61(5):1320–1327

3. Yash B, Yadav MV, Baudha S (2020) A compact mace shaped ground plane modified circular patch antenna for ultra-wideband applications. *Telecommun Radio Eng* 79(5):363–381
4. Khidre A, Lee KF, Elsherbeni AZ, Yang F (2013) Wide band dual-beam U-slot microstrip antenna. *IEEE Trans Antennas Propag* 61(3):1415–1418
5. Srivastava I, Baudha S, Yadav MV (2020) A novel approach for compact antenna with parasitic elements aimed at ultra-wideband applications. In: 2020 14th European conference on antennas and propagation (EuCAP), vol 14(4). IEEE, pp 1–5
6. Shagar A, Wahidabanu S (2011) Novel wideband slot antenna having notch-band function for 2.4 GHz WLAN and UWB applications. *Int J Microwave Wirel Technol* 3(4):451–458
7. Kapoor K (2019) U-shaped microstrip patch antenna with partial ground plane for mobile satellite services (MSS). In: 2019 URSI Asia-Pacific radio science conference (AP-RASC). IEEE, pp 1–5
8. Kurniawan A, Mukhlisin S (2013) Wideband antenna design and fabrication for modern wireless communications systems. *Procedia Technol* 11:348–353
9. Ghosh A, Ghosh SK, Ghosh D, Chattopadhyay S (2016) Improved polarization purity for circular microstrip antenna with defected patch surface. *Int J Microw Wirel Technol* 8(1):89–94
10. Baudha S, Yadav MV (2019) A compact ultra-wide band planar antenna with corrugated ladder ground plane for multiple applications. *Microw Opt Technol Lett* 61(5):1341–1348
11. Abdelraheem AM, Abdalla MA (2016) Compact curved half circular disc-monopole UWB antenna. *Int J Microw Wirel Technol* 8(2):283–290
12. Gupta S (2019) Parasitic rectangular patch antenna with variable shape ground plane for satellite and defence communication. In: 2019 URSI Asia-Pacific radio science conference (AP-RASC). IEEE, pp 1–4
13. Awad N, Abdelazeez MK (2018) Multislot microstrip antenna for ultra-wide band applications. *J King Saud Univ Eng Sci* 30(1):38–45
14. Basak A, Manocha M, Baudha S, Yadav MV (2020) A compact planar antenna with extended patch and truncated ground plane for ultra wide band application. *Microw Opt Technol Lett* 62(1):200–209
15. Yadav MV, Baudha S (2021) A miniaturized printed antenna with extended circular patch and partial ground plane for UWB applications. *Wirel Pers Commun* 116(1):311–323
16. Deshmukh AA, Singh D, Zaveri P, Gala M, Ray KP (2016) Broadband slot cut rectangular microstrip antenna. *Procedia Comput Sci* 93:53–59
17. Hota S, Mangaraj BB (2019) Miniaturized planar ultra-wideband patch antenna with semi-circular slot partial ground plane. In: 2019 IEEE Indian conference on antennas and propagation (InCAP). IEEE, pp 1–4
18. Garg H et al (2019) Dumbbell shaped microstrip broadband antenna. *J Microwaves Optoelectron Electromagnet Appl* 18:33–42
19. Lakshmanan R, Sukumaran SK (2016) Flexible ultra wide band antenna for WBAN applications. *Procedia Technol* 24:880–887
20. Mangaraj BB, Hota S, Varun Yadav M (2020) A novel compact planar antenna for ultra-wideband application. *J Electromagnet Waves Appl* 34(1):116–128
21. Hota S, Mangaraj BB (2019) A compact, ultrawide band planar antenna with modified circular patch and a defective ground plane for multiple applications. *Microw Opt Technol Lett* 61(9):2088–2097
22. Golait M, Varun Yadav M, Patil BH, Baudha S, Bramhane LK (2021). A compact ultra-wideband square and circular slot ground plane planar antenna with a modified circular patch. *Int J Microwave Wirel Technol* 1–6
23. Varun Yadav M, Baudha S, Sanghi V (2023) A 5G rotated frame radiator for ultra wideband microwave communication. *Int J Microwave Wirel Technol* 1–9. <https://doi.org/10.1017/S1759078722001453>
24. Mazinani SM, Hassani HR (2009) A novel broadband plate-loaded planar monopole antenna. *IEEE Antennas Wirel Propag Lett* 8:1123–1126
25. Thukral S (2021) A compact hammer shaped printed antenna with parasitic elements for defense and mobile satellite applications. In: 2021 IEEE Indian conference on antennas and propagation (InCAP). IEEE, pp 521–524

26. Bansal Y (2020) A compact slot antenna for ultra-wideband applications. *Telecommun Radio Eng* 79(3)
27. Kim GH, Yun TY (2013) Compact ultrawideband monopole antenna with an inverted-L-shaped coupled strip. *IEEE Antennas Wirel Propag Lett* 12:1291–1294

Systematic Literature Review on the Machine Learning Techniques for UAV-Assisted mm-Wave Communications



M. S. Sugesh and G. Vairavel

Abstract The application of machine learning (ML) solutions in UAV-assisted 5G communication can bring significant benefits to 5G and beyond 5G communication. There is little elementary, secondary, and tertiary study on machine learning applications in UAV-assisted 5G communication. The apparent paucity of such investigation makes it hard to develop precise solutions for UAV-assisted 5G communication. Therefore, it is essential to study and comprehend how to use machine learning in UAV-aided 5G communication. In this article, we deliver a systematic study of all-important research activities, in which machine learning (ML) methods have been used on the wireless communication based on UAV for improving several design and functional characteristics such as beamforming, resource allocation, dynamic deployment, and trajectory prediction. The studies were clustered into four themes: the main machine learning algorithms applied in UAV-assisted wireless communication; UAV-assisted 5G communication process in which machine learning processes and/or frameworks are applied; main application categories of machine learning algorithms and/or frameworks; and the results of this review specify that the main machine learning algorithms/framework applied in UAV-assisted wireless communication are: Q-Learning, MARL, K-means, AMSSA, genetic algorithm, support vector machine (SVM), support vector regression, artificial neural network (ANN), LSM, cross-entropy algorithm, DL algorithm, and reinforcement learning algorithm.

Keywords Machine learning (ML) · UAV · Beamforming · 5G · Millimeter wave · MIMO

M. S. Sugesh (✉)

Sreepathy Institute of Management and Technology, Vavanoor, Palakkad, India

e-mail: sugesh.ms@simat.ac.in

G. Vairavel

SRM Institute of Science and Technology, Kattankulathur, Chennai, India

e-mail: vairaveg@srmist.edu.in

1 Introduction

The maneuverability of UAVs offers extensive coverage, capacity, energy efficiency, reliability, and on-demand connectivity in envisioned areas for 5G millimeter-wave communication. As per the 3GPP-releases, the deployment of 5G will be in full-fledged operation by the year 2030, which provides low latency, high data rate and reliability, high connection density of the order of 1 million/km², and long battery life. Leveraging the UAV-assisted 5G communication network in exigency situations is an economically feasible solution instead of deploying the network infrastructure. UAV-assisted 5G communication is viewed as a good potential paradigm to support three centralized future wireless network usage scenarios, i.e., enhanced mobile broadband (eMBB) with bandwidth-consuming, ultra-reliable low-latency communication (URLLC), and massive machine-type communications (mMTC).

Owing to the dynamicity of UAV position and altitude, low altitude platform (LPA) UAVs are extensively used as aerial base stations and for communication relaying in terrestrial communications and high traffic areas such as stadiums and crowded places. The technical challenges in the use of UAVs instead of terrestrial networks include payload and time constraints, trajectory optimization of UAV, channel acquisition, and backhauling. The main attributes of UAVs for the communications service focus on three aspects [1], energy-aware deployment and operation of UAV, multiple input multiple output (MIMO) communications, and path planning in UAV systems. And the characteristics that a UAV must possess for terrestrial communication [2] are Line-of-Sight (LoS) Links, Dynamic Deployment Ability, and UAV-Based Swarm Networks. There are two types of energy-efficient operation schemes for UAVs [3], energy-efficient mobility and energy-efficient communication.

The primary concern in the realization of 5G using UAVs in the under covered areas is the energy efficiency, and in [3], Koumaras et al. proposed softwarization of the possible hardware components of the communication systems in the UAV base station. In device-to-device communication, having a massive MIMO antenna array, there should be a trade-off among energy efficiency (EE) and average sum rate (ASR). The high mobility of UAVs would make it harder to align the transmitter/receiver beam for mm-wave directional communications. In [4–7] different compatible antenna structures are presented. Multiple antennas in UAVs are quite expensive due the obvious high computational cost, the circuit, and the energy consumption cost. In [8] it is showed that there is a great potential for massive MIMO technique in UAV base stations. The challenging task of obtaining channel state information (CSI) accurately bounds the practical gain of UAV-MIMO systems. The gain and reliability of the UAV-assisted network also rely on the interference effect of the downlink channel (Fig. 1).

The beamforming in massive MIMO UAV systems is an optimal solution for the interference and position problem. Since energy efficiency is of major concern in UAV aerial base stations, the introduction of machine learning techniques is very essential to overcome the beam selection and energy optimization issues. Although

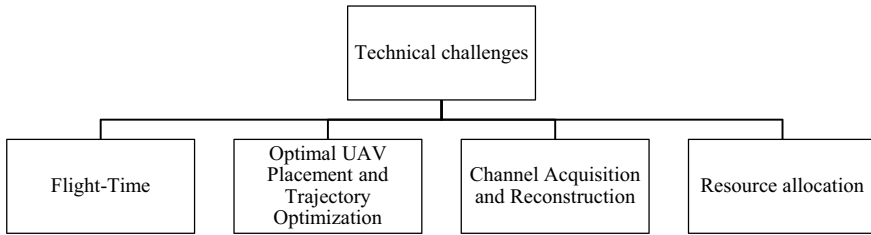


Fig. 1 Technical challenges in the use of UAVs instead of terrestrial networks

there are many research works for UAV-assisted 5G millimeter-wave communication, the use of ML for beamforming in UAV-assisted 5G network is still in the initial stage.

In this article, we review the machine learning techniques and algorithms used in UAV-assisted 5G mm-wave communication.

2 Existing Surveys and Tutorials

In light of the advantages in UAV-assisted mm-wave communication, recently there are numerous research works which are progressing. This activity also includes number of surveys and tutorials on UAV communications, such as [2, 3, 8–10]. There are few survey papers published over the last few years, including network architecture, channel modeling, resource allocation, and technical challenges, in UAV-assisted mm-wave 5G communication. Even though these surveys discussed relevant issues in UAV-assisted communication, there are many limits.

In [9], Xiao et al. investigated the hierarchical beamforming (BF) codebook structure along with the codebook design via sub-array technique for fast BF training and tracking. Channel propagation characteristics and the authors identified the correlation between the location of UAV and the user discovery at mm-wave frequencies. In most of the reviewed surveys [1, 2, 8–11], the channel modeling and channel characteristics were discussed. The areas, in UAV-assisted mm-wave communication, explored and the challenges identified in survey papers include inference, channel modeling, routing, seamless handover, UAV placement, etc. One-line summary of the review papers in the field of UAV-assisted 5G mm-wave communication is given in Table 1. From the review of survey papers, it is evident that there are no or only few works which are concentrated in the application of machine learning in UAV-assisted mm-wave communication. To fulfill the gap in the aforementioned area, we are focusing on the application of machine learning in the beamforming (BF), three-dimensional deployment, and UAV trajectory planning in the application of UAV in mm-wave communication.

Table 1 One-sentence summary of the existing survey papers in UAV-assisted wireless communication

Review paper	One-sentence summary	Focused areas
[9]	Millimeter-wave communication for UAV cellular networks	<ul style="list-style-type: none"> • Key challenges in mm-wave UAV cellular • Channel propagation characteristics • Hierarchical beamforming codebook structure
[1]	Wireless communication through the use of UAV	<ul style="list-style-type: none"> • Basic networking architecture • Characteristics of the channel • The key design aspects of the network
[12]	Survey of significant issues in UAV mm-wave communication links	<ul style="list-style-type: none"> • Self-organization and distraction tolerance • Control of SDN • Unified handover and energy efficiency
[13]	Significant hurdles and fundamental trade-offs in UAV-assisted wireless links	<ul style="list-style-type: none"> • Three-dimensional deployment and performance analysis • Air-to-ground channel modeling • Energy efficiency
[2]	Study of UAV communication in the context of 5G/B5G	<ul style="list-style-type: none"> • The space-air-ground interconnect networks • UAV platform-based 5G approaches • Future research directions
[14]	UAV millimeter-wave communications research challenges and opportunities	<ul style="list-style-type: none"> • Channel characteristics and its modeling concerns • UAV mm-wave channel acquisition and pre-coder design • Communications and spectrum sharing
[15]	An review of recent developments in UAV communications	<ul style="list-style-type: none"> • LoS-dominant UAV-ground channels • QoS consideration for UAV control messages versus payload data • SWAP limitations of UAVs
[16]	UAV-assisted wireless networks with 5G mm-wave communications	<ul style="list-style-type: none"> • Important technical benefits and obstacles • Potential uses for UAV-assisted wireless in 5G mm-wave communications • Current issues, solutions, and unsolved problems
[17]	Survey on UAV cellular communications	<ul style="list-style-type: none"> • Types of UAV and standardization • Aerial base stations and prototyping and field tests • Regulation and security
[11]	5G's prospective beneficiaries and use cases	<ul style="list-style-type: none"> • Vehicle-to-everything (V2X) communication • Drone • Health care
[18]	Machine learning in the design of a 5G RAN based on UAVs	<ul style="list-style-type: none"> • Supervised and reinforcement learning strategies • Application scenarios • Technical challenges and requirements compared

(continued)

Table 1 (continued)

Review paper	One-sentence summary	Focused areas
[19]	mm-wave UAV communications aided by beamforming techniques	<ul style="list-style-type: none"> • The structure of the antenna and the channel model UAV communication terminal • UAV access point and UAV-supported backbone link
[8]	Massive MIMO for UAVs with cellular connectivity	<ul style="list-style-type: none"> • Challenges in serving UAVs with massive MIMO • Pilot decontamination • Three-dimensional beam tracking and hybrid beamforming
[20]	UAVs and mm-wave in 5G	<ul style="list-style-type: none"> • UAV relay architectures • Optimal positioning of UAVs as a relay • The scope of artificial intelligence and machine learning techniques
[10]	Study of UAV-assisted 5G and beyond wireless networks	<ul style="list-style-type: none"> • Optimized 3D placement resource allocation • UAVs in wireless networks • Deployment of UAV in 5G and beyond 5G networks
[21]	mm-wave beamforming enabled UAV communication and networking	<ul style="list-style-type: none"> • Antenna architecture and channel modeling in the mm-wave band • UAV-connected mm-wave cellular networks technologies and solutions • Problems and hopeful directions for future research
[22]	Path planning algorithms for UAVs	<ul style="list-style-type: none"> • Model-based path planning • Conventional path planning • Learning-based path planning • Cell-based path planning
[23]	ML architectures and designs	<ul style="list-style-type: none"> • Optimizing communication, computation, and resource distribution

3 Methodology

A systematic evaluation procedure was explained according to the guiding principle of [24]. In this practice, the work heading, the investigation framework, the motive for doing the studies, the key research question, the minor questions, exploration process definition, and the exploration string are assessed.

Four research inquiries were created, and the latter three were being derived from the primary question. These research inquiries have the subsequent aims: to characterize the main ML algorithms and/or frameworks applied in 5G UAV-assisted wireless communication; identify the main UAV-based communication where these algorithms and/or frameworks are applied; define the key types of application of these algorithms and/or frameworks in UAV-assisted wireless communication; and

describe the key results and benefits of applying these algorithms and/or frameworks in UAV-assisted wireless communication (Table 2).

About the addition and elimination criteria, those main studies that presented ML methods and applications in UAV-assisted wireless communication, with qualitative and quantitative main studies, were considered qualified.

Articles published in English throughout the period 2018–2022 were taken into consideration. Main papers that did not discourse the focal objective of this work were excluded, as were documents that were unsuccessful in the quality valuation process, case studies, and papers published outside the specified period.

Regarding the research repositories considered, the following research libraries were chosen:

- Hindawi.
- IEEE-Xplore.
- Elsevier-ScienceDirect.
- Springer Link.
- MDPI.
- Google Scholar.

The article assortment process was divided according to the phases stated in Fig. 2. The search procedure in the sources used the descriptors below to generate the exploration thread.

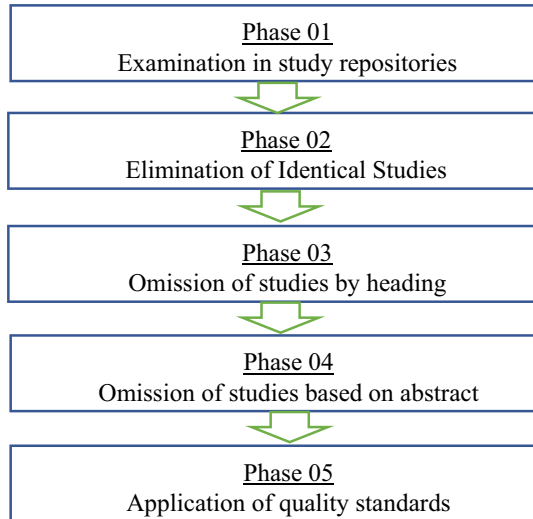
- UAV-assisted 5G communication and machine learning.
- UAV AND wireless communication and machine learning.
- 5G AND machine learning and UAV.
- Reinforcement learning and UAV.
- Deep learning AND UAV and 5G.
- Artificial intelligence and UAV communication.

Once the above descriptors have been defined, they were joint using the Boolean operator “OR” to form a query thread that give back documents that at least comprised one of the terms specified in the exploration thread.

Table 2 Research questions

Research questions	Description
RQ	What are the major ML algorithms or frameworks used in UAV-enabled wireless communication?
RQ	What are the UAV-assisted 5G communication process in which ML algorithms or frameworks are functional?
RQ	What are the major application categories of ML algorithms or frameworks in UAV-assisted 5G communication?
RQ	What are the major outcomes accomplished of applying ML algorithms or frameworks in UAV-assisted 5G communication?

Fig. 2 Selection phases of systematic review papers



4 Results

The number of publications has declined considerably in recent years compared to 2018. This could be associated with the impact of the COVID-19 pandemic, which has reduced the number of conferences as a result of health restrictions.

(a) **What are the major ML algorithms or frameworks used in UAV-enabled wireless communication?**

This investigation query is characterizing the major ML algorithms or frameworks applied in UAV-assisted 5G communication. Incorporating machine learning techniques into 5G wireless networks can leverage intelligence to solve various problems which assures time complexity reduction and unique performance gains, in terms of energy efficiency and memory usage. Machine learning is the essential technology that enables supervisory capabilities to achieve the proper deployment and trajectory of UAVs. A number of researches have suggested the use of ML to improve wireless networks performance. The main areas, in UAV-assisted wireless communication, where ML is applied include UAV swarm communication [21, 22], resource allocation [23–25], trajectory prediction [26], energy optimization [27–32], target detection [33], etc. The reinforcement learning and unsupervised learning are the two main machine learning categories which are applied in UAV-assisted wireless communication. K-means algorithm, Q-learning, SALP algorithm, Deep Q-network (DQN), liquid state machine (LSM), and convolutional neural network (CNN) are few learning algorithms and frameworks applied to tackle the problems in aforesaid wireless communication areas in UAV scenarios. Table 3 shows the ML algorithms and process in UAV-assisted wireless communication.

Table 3 Machine learning algorithms and process in UAV-assisted wireless communication

Paper	Algorithms	Process
[27]	Hierarchical Q-learning-based power allocation algorithm (HQLA)	Best power distribution plan for the transmitter
[30]	Multi-agent reinforcement learning (MARL) algorithm	Decision-making in UAV swarm confrontation
[31]	Particle swarm optimization- (PSO-) K-means algorithm	Optimization of number and location of UAV
	Adaptive mutation salp swarm algorithm (AMSSA)	Optimizes the energy efficiency (EE) of D2D pairs
[32]	Dinkelbach method	Energy efficiency maximization
[33]	Artificial neural network (ANN) and support vector regression (SVR)	Prediction of RSSI and path loss characteristics
[34]	Support vector machine (SVM)	Multi-target detection
[35]	Block coordinate descent (BCD)	Maximize the system worst-case secrecy energy efficiency (SEE)
	Dinkelbach's algorithm	
	Successive convex approximation (SCA)	
[36]	Genetic algorithm	Minimize total UAV energy consumption
[37]	Reinforcement learning	Optimization of multi-UAV communication network
[25]	Liquid state machine (LSM) algorithm	Cache contents, user association, and spectrum distribution UAVs
[26]	Multi-agent Q-learning-based placement algorithm	Optimization of UAVs' trajectories and power
[38]	Genetic algorithm-based K-means (GAK-means) Q-learning-based deployment algorithm	Complexity optimization
[39]	Deep reinforcement learning-based trajectory control algorithm (RAT)	Minimize energy consumption of all UEs
[40]	Salp swarm algorithm (SSA)	RSS at UAV
[41]	Hybrid EA-LM algorithm	Prediction of received signal strength (RSS)
	DE-LM algorithm	
[42]	Unsupervised learning	Foresee the quality of air-to-ground links' performance for UAV-BS
[43]	CNN, LSTM	Automatic modulation classification
[44]	Load prediction algorithm (LPA)	Predicts congested macro-cells
	UAV BSS positioning and clustering algorithm	Optimization of number and location of UAV
[45]	Echo state network (ESN)-based prediction algorithm	Predicting the movement of users
	K-means-based algorithm	Dynamic 3D placement of the UAV

(continued)

Table 3 (continued)

Paper	Algorithms	Process
[46]	Recurrent neural networks (RNNs)-based arrival angle predictor	Predict the specific communication location of UAV
[47]	Cross-entropy (CE) algorithm	Lens array architecture with efficient and energy-saving hybrid precoding algorithm
[48]	DL algorithm	Deep autoencoder and a neural network
[49]	Voting regression	Combine several ML learners in an ensemble learning method,
[50]	Asynchronous advantage actor-critic (A3C)-AFL algorithm	Resource management, UAVs placement, and joint device selection
[51]	Reinforcement learning algorithm	MAB approach
[52]	Q-learning	Beam tracking scheme

(b) **What are the UAV-assisted 5G communication process in which ML algorithms or frameworks are functional?**

(i) *Machine Learning in beam-forming*

One of the typical applications of UAV-assisted mobile networks is that UAVs are deployed as aerial access points (APs) to establish temporary connectivity with the ground users in critical and hotspot regions. The beamforming enables massive connectivity for different application scenarios of UAV-assisted cellular communication. The energy constraints in UAV lead to the investigation of the more efficient communication strategies at mm-wave MIMO communication. Beamforming enables the highly directive narrow beam transmission and reception possible in MIMO systems. In [28], Yu et al. proposed a computationally simple hybrid beamforming algorithm for mm UAV communication which utilizes the predefined analog codebook for selecting the beam with high channel matrix correlation. Though it provides better channel detection performance, it can only be utilized in a system that serves a small number of users.

To increase the coverage and reliability of 5G communication networks in hard-to-reach areas, terrestrials, and natural calamity affected areas, the deployment of UAVs as the aerial base station is a viable solution. In [29], Boschiero et al. studied the feasibility of deploying the UAV-assisted network with the help of a stochastic geometry approach. Analytical expressions for coverage probability and different system models are being proposed. Even though the optimum altitude of UAV is obtained, performance of antenna array plays a significant role in reliable communication. The key fundamental of attaining minimum path loss is beamforming. Susarla et al. [53] proved that a deep reinforcement learning method can be used to guide the UAV in reaching its destination from a point within the BS coverage area without losing the 5G communication. The author proposed a DQN algorithm for planning the UAV path and assuring the connectivity quality. The proposed algorithm is considered only for a single UAV environment without considering the energy and

channel constraints. Lens array antenna with a cross-entropy optimization algorithm for multi-UAVs is proposed in [41] by Ren et al. In the multi-UAV environment, the computational complexity and energy consumption need to be addressed.

Salman et al. [50] introduced cost-effective beam steering for the UAV signal reception. The author proposes a single wideband signal with a null steering technique to mitigate the computation complexity of the smart antenna array. In this, the beam from the UAV antenna array is aligned in the direction of the incoming signal by multiplying the weighted vector with the generated phase conjugate signal. The system technique utilizes the fact that the UAV is always within the wideband main lobe. But in the UAV to multi-UE communication, the interference problem is severe. A constellation division multiple access (CoDMA) technique is proposed, by Wang et al. [51], to manage the concurrent signal interference problem of single wideband UAV-aided dynamic networks. Multi-UAV system requires optimum beam width for better interference performance and is an open challenge in UAV-assisted 5G mm-wave communication.

In [54], Liu et al. put forward a technique, LSTM-based LRNet algorithm for predictive beamforming scheme, to track the moving UAV communication signal. The angular separation of the UAV and the UE is calculated from the foretold location. The deep learning method is used for facilitating the beam orientation in UAV-assisted communication. The advantages of using the ML tools in a 5G network with UAVs are described in [13]. The authors, Kouhdaragh et al., outlined the major application scenarios of UAVs, viz. UAV-Mounted Base Stations, UAV-Based Cooperation, and UAV-Based Software-Defined Networks. The challenges in deploying UAVs instead of the terrestrial network are delineated in the paper. Also, various applications and advantages of ML tools in UAV-RAN are presented. Although there are many research works for UAV-assisted 5G millimeter-wave communication, the use of ML for beamforming in UAV-assisted 5G network is still in the initial stage.

(ii) *Machine Learning in Resource Allocation*

In UAV-based millimeter-wave communication, the limited resources are shared between mobile UAV base stations and ground communication networks. To attain the optimum performance of the network in terms of network users and throughput, it is necessary to design an efficient system for resource allocation. Since the UAVs are battery powered with limited lifetime, the key factors in resource allocation are time slot allocation, bandwidth, power, and user associations. Chen et al. [25] introduced a LSM-based resource allocation which considers the distribution of content request of network users based on which the UAV can choose the spectrum allocation scheme. The locations of the UAVs must be planned to offer the optimum channel gains for communication services, and the power radio resources must be managed to increase the communication data rate. In [44], Helin Yang et al. proposed a reinforcement learning algorithm called A3C-AFL algorithm which completes its process in three steps, (1) Device Selection, UAVs Placement, and Resource Management (2) Local Training, and (3) Global Aggregation. A UAV-supported advanced computing resource allocation plan for 5G communication in the Internet of things environment

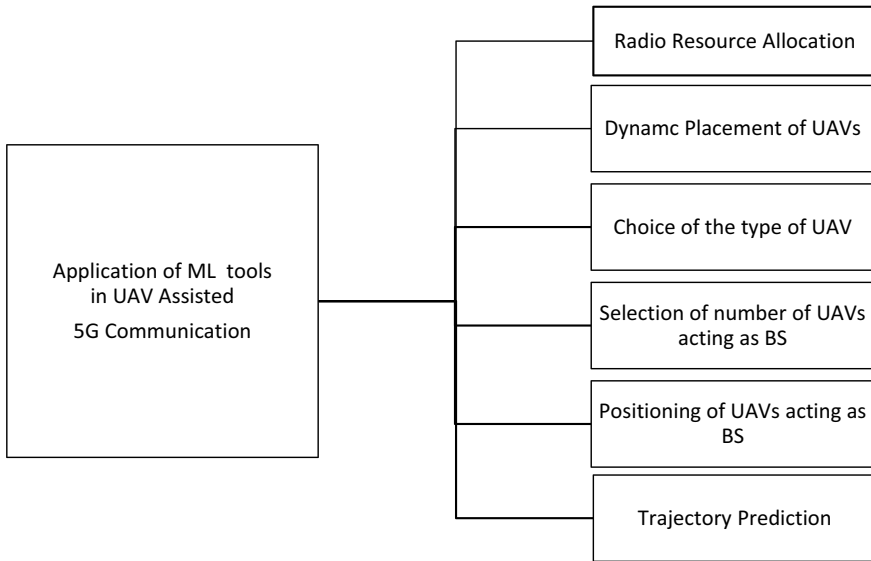


Fig. 3 Applications of machine learning tools in U-RAN

is anticipated in [29]. Hao Liu et al. used the genetic algorithm to obtain the possible strategy to allocate resources (Fig. 3).

(iii) *Machine Learning in Dynamic Placement of the UAV*

To support ground communication networks, the main mechanisms of UAV communication are (i) dynamic deployment of UAVs, (ii) LoS links, and (iii) UAV swarm networks [18]. In dynamic UAV deployment, UAVs can be deployed dynamically in a timely changing environment compared to fixed organization, making them more strong to be used in actual applications. The forecast of user location data and lively 3D UAV placement for upholding high service quality were simultaneously investigated in [39]. In [22] Enchang Sun et al. used particle swarm optimization K-means algorithm to decide number and deployment position of UAVs. And also, Q-learning algorithm is proposed to implement the master–slave follow-up strategy among the swarm members. The initial positioning of the UAVs is done by K-means clustering, and number of iterations required is determined by the PSO algorithm. The time complexity of PSO-K-means algorithm depends up on the number of users, UAVs, particle position, etc. In [31], Xiao Liu et al. proposed a Q-learning-based deployment algorithm for the dynamic location of both UAV with the assumption of the movement of users which are within the cluster. In [39], three-dimensional deployment of UAV based on predicted movement of bunch of user using ML algorithm based on echo state network is proposed by Peize Zhao et al. In [38, 39], K-means-based algorithm is used to place the UAV.

(iv) *Machine Learning in Trajectory Prediction*

In SWARM networks or UAV-assisted communication networks, it is very much important to predict the location in the next movement of the participating UAV for seamless and reliable communication. Zhu [48] proposed a deep learning-based trajectory prediction for UAV swarm network. GRU-based DL network is used to train the elevation and azimuth data of the front-runner UAV to forecast its position at the subsequent instant. In the conventional DL method, memory lessness makes the system less accurate. In [46], Xiao et al. proposed an incoming angle predictor based on the RNN to predict the position of the UAV at the subsequent moment. The elevation angle and horizontal angle are calculated by applying DOA algorithm and a model is created using four-layer RNN network to foresee the UAV trajectory. Optimization-based trajectory control algorithm is one of the effective methods for the efficient utilization of the available energy in a UAV. Wang et al. [39] proposed DRL-based RAT algorithm to predict the multi-UAVs' trajectories. In most of the trajectory prediction algorithms, it considers only the 2D movement of the UAV; hence, there is a scope for research in trajectory prediction of UAV considering three-dimensional movement (Table 4).

(c) **What are the major application categories of ML algorithms or frameworks in UAV-assisted 5G communication?**

This investigation question aims to describe the major classes of applications of ML algorithms in UAV-aided 5G communication. In UAV swarm network, the commonly used machine learning algorithms are reinforcement learning, Q-learning algorithm, and GAK-means algorithm [29, 31, 32]. For reliable communication, there should be line of site reliable communication between UAV and ground stations. For this purpose, BCD, Dinkelbach's, and SCA algorithms are proposed in [24, 35, 38]. The genetic algorithm facilitates resource allocation and UAV deployment in multi-UAV communication. For the signal recognition and signal identification purposes in 5G, Millimeter-wave communication is supported by CNN and LSTM algorithms (Table 5). For the trajectory tracking and prediction purpose, SVM and SVR algorithms are applied. Table 6 shows the article distribution by classes of application.

(d) **What are the major outcomes accomplished of applying ML algorithms or frameworks in UAV-assisted 5G communication?**

RQ-01

(a) What are the major ML algorithms or frameworks used in UAV-enabled wireless communication?

Answer: The major ML algorithms or framework applied in UAV-assisted wireless Communication are: Q-Learning, MARL, K-means, AMSSA, genetic algorithm,

Table 4 UAV-assisted 5G communication process

Paper	Process
[27]	Optimal power allocation strategy of the transmitter
[30]	Decision-making in UAV swarm confrontation
[55]	Optimization of number and location of UAV Maximizes the energy efficiency (EE) of D2D pairs
[32]	Energy efficiency maximization
[33]	Prediction of RSSI and path loss characteristics
[34]	Multi-target detection
[35]	Maximize the system worst-case secrecy energy efficiency (SEE)
[36]	Minimize total UAV energy consumption
[37]	Optimization of multi-UAV communication network
[25]	Cache contents, user association, and spectrum allocation UAVs
[26]	Optimization of UAVs' power and trajectories
[38]	Complexity optimization
[39]	Minimize energy consumption of all UEs
[40]	RSS at UAV
[41]	Prediction of received signal strength (RSS)
[42]	Predict the quality of air-to-ground links' performance for UAV-BS
[43]	Automatic modulation classification
[44]	Predicts congested macro-cells Optimization of number and location of UAV
[45]	Dynamic 3D deployment of UAV

Table 5 Application classes of machine learning algorithms

Application	Paper
UAV secure communication	[27, 30]
D2D communication in UAV-based networks	[55]
Emergency communication networks	[32]
Smart farming	[33]
Trajectory tracking	[34]
UAV to ground wireless communication	[35, 40, 44]
Resource allocation	[36, 43]
Multi-UAV communication	[26, 37, 38]
Flying mobile edge computing (F-MEC)	[39]
Channel modeling	[41, 42]

Table 6 Article distribution by algorithm application results

Algorithm	Result	Paper
PSO-K-means algorithm	Maximize the energy efficiency	[31]
AMSSA		
Dinkelbach method		[32]
SVR and ANN	Improved path loss prediction accuracy	[33]
Genetic algorithm	Computational cost minimization	[36]
LSM algorithm	Improved convergence time	[25]
Q-learning-based placement algorithm	Improved instantaneous sum transmit rate	[26]
GAK-means	Improved convergence rate after	[38]
Load prediction algorithm	Improved QoS and energy efficiency	[44]
ESN-based prediction algorithm	Improving the sum transmit rate	[45]
RNN-based predictor	Improved prediction accuracy	[46]
Cross-entropy (CE) algorithm	High sum rate and energy efficiency	[47]
DL algorithm	Throughput maximization	[56]
A3C-based (AFL) algorithm	Minimized execution time	[57]
RL algorithm	Fast beam section	[58]
Q-learning	Improved SINR	[52]

SVM, SVR, ANN, LSM, cross-entropy algorithm, DL algorithm, and reinforcement learning algorithm.

RQ-02

- (b) What are the UAV-assisted 5G communication process in which ML algorithms or frameworks are functional?

Answer: The major activities are beamforming, resource allocation, 3D deployment, and trajectory planning.

RQ-03

- (c) What are the major application categories of ML algorithms or frameworks in UAV-assisted 5G communication?

Answer: The main applications of machine learning algorithms in UAV-assisted 5G communication are in D2D communication, multi-UAV communication, trajectory planning, channel modeling, resource allocation, smart farming, and flying mobile edge computing (F-MEC).

RQ-04

- (d) What are the major outcomes accomplished of applying ML algorithms or frameworks in UAV-assisted 5G communication?

Answer: The major results' ML algorithms' application in UAV-assisted 5G communication is improved energy efficiency, less time complexity, improved accuracy in prediction, and efficient resource allocation.

5 Open Issues

Open issues identified during the systematic review process are listed below:

- Multi-UAV system requires optimum beam width for better interference performance and is an open challenge in UAV-assisted 5G mm-wave communication.
- In the multi-UAV environment, the computational complexity and energy consumption need to be addressed.
- Movement of UAVs has not been pointed based on the users' mobility in most present investigation articles, which mostly focus on the two-dimensional placement of multiple UAVs or the motion of single UAV, while ground users remain stationary.
- Research on three-dimensional placement of mobile multiple UAVs considering ground user mobility.
- In UAV swarm communication, the performance of the front-runner UAV has become the tailback to enhance the performance of the whole UAV cluster, which necessitates more research consideration.
- In most of the trajectory prediction algorithms, it considers only the 2D movement of the UAV; hence, there is a scope for research in trajectory prediction of UAV considering three-dimensional movement.

References

1. Zeng Y, Zhang R, Lim TJ (2016) Wireless communications with unmanned aerial vehicles: opportunities and challenges. *IEEE Commun Mag* 54(5):36–42. <https://doi.org/10.1109/MCOM.2016.7470933>
2. Li B, Fei Z, Zhang Y (2019) UAV communications for 5G and beyond: recent advances and future trends. *IEEE Internet Things J* 6(2):2241–2263. <https://doi.org/10.1109/JIOT.2018.2887086>
3. Koumaras H et al (2021) 5G-enabled Uavs with command and control software component at the edge for supporting energy efficient opportunistic networks. *Energies* 14(5):1–17. <https://doi.org/10.3390/en14051480>
4. Sreelakshmy R, Vairavel G (2019) Novel cuff button antenna for dual-band applications. *ICT Express* 5(1):26–30. <https://doi.org/10.1016/j.ict.2018.01.012>
5. Kolangiammal S, Balaji L, Vairavel G (2022) A compact planar monopole UWB MIMO antenna design with increased isolation for diversity applications. *Appl Comput Electromagn Soc J* 37(4):458–465. <https://doi.org/10.13052/2022.ACES.J.370411>
6. Kolangiammal S, Vairavel G (2021) Compact planar monopole UWB MIMO antenna for diversity applications. *Adv Smart Syst Technol* 281–291

7. Kittur A, Vairavel G (2021) Understanding textile antenna by reviewing and simulating it for high data rates applications. In: *Advances in smart system technologies*, pp 189–207
8. Huang Y, Wu Q, Lu R, Peng X, Zhang R (2021) Massive MIMO for cellular-connected UAV: challenges and promising solutions. *IEEE Commun Mag* 59(2):84–90. <https://doi.org/10.1109/MCOM.001.2000552>
9. Xiao Z, Xia P, Xia XG (2016) Enabling UAV cellular with millimeter-wave communication: potentials and approaches. *IEEE Commun Mag* 54(5):66–73. <https://doi.org/10.1109/MCOM.2016.7470937>
10. Shahzadi R, Ali M, Khan HZ, Naeem M (2021) UAV assisted 5G and beyond wireless networks: a survey. *J Netw Comput Appl* 189:103114. <https://doi.org/10.1016/j.jnca.2021.103114>
11. Ullah H, Gopalakrishnan Nair N, Moore A, Nugent C, Muschamp P, Cuevas M (2019) 5G communication: an overview of vehicle-to-everything, drones, and healthcare use-cases. *IEEE Access* 7(c):37251–37268. <https://doi.org/10.1109/ACCESS.2019.2905347>
12. Gupta L, Jain R, Vaszkun G (2016) Survey of important issues in UAV communication networks. *IEEE Commun Surv Tutor* 18(2):1123–1152. <https://doi.org/10.1109/COMST.2015.2495297>
13. Mozaffari M, Saad W, Bennis M, Nam YH, Debbah M (2019) A tutorial on UAVs for wireless networks: applications, challenges, and open problems. *IEEE Commun Surv Tutor* 21(3):2334–2360. <https://doi.org/10.1109/COMST.2019.2902862>
14. Zhang C, Zhang W, Wang W, Yang L, Zhang W (2019) Research challenges and opportunities of UAV millimeter-wave communications. *IEEE Wirel Commun* 26(1):58–62. <https://doi.org/10.1109/MWC.2018.1800214>
15. Zeng Y, Wu Q, Zhang R (2019) Accessing from the sky: a tutorial on UAV communications for 5G and beyond. *Proc IEEE* 107(12):2327–2375. <https://doi.org/10.1109/JPROC.2019.2952892>
16. Zhang L et al (2019) A survey on 5G millimeter wave communications for UAV-assisted wireless networks. *IEEE Access* 7:117460–117504. <https://doi.org/10.1109/ACCESS.2019.2929241>
17. Fotouhi A et al (2019) Survey on UAV cellular communications: practical aspects, standardization advancements, regulation, and security challenges. *IEEE Commun Surv Tutor* 21(4):3417–3442. <https://doi.org/10.1109/COMST.2019.2906228>
18. Kouhदारagh V, Verde F, Gelli G, Abouei J (2020) On the application of machine learning to the design of UAV-based 5G radio access networks. *Electron* 9(4):1–20. <https://doi.org/10.3390/electronics9040689>
19. Xiao Z, Zhu L, Xia XG (2020) UAV communications with millimeter-wave beamforming: potentials, scenarios, and challenges. *China Commun* 17(9):147–166. <https://doi.org/10.23919/JCC.2020.09.012>
20. Khan SK, Naseem U, Siraj H, Razzak I, Imran M (2021) The role of unmanned aerial vehicles and mmWave in 5G: recent advances and challenges. *Trans Emerg Telecommun Technol* 32(7):1–18. <https://doi.org/10.1002/ett.4241>
21. Xiao Z et al (2022) A survey on millimeter-wave beamforming enabled UAV communications and networking. *IEEE Commun Surv Tutor* 24(1):557–610. <https://doi.org/10.1109/COMST.2021.3124512>
22. Qadir Z, Ullah F, Munawar HS, Al-Turjman F (2021) Addressing disasters in smart cities through UAVs path planning and 5G communications: a systematic review. *Comput Commun* 168(January):114–135. <https://doi.org/10.1016/j.comcom.2021.01.003>
23. Nassef O, Sun W, Purmehdi H, Tatipamula M, Mahmoodi T (2022) A survey: distributed machine learning for 5G and beyond. *Comput Netw* 207(January):108820. <https://doi.org/10.1016/j.comnet.2022.108820>
24. Kitchenham B, Pearl Brereton O, Budgen D, Turner M, Bailey J, Linkman S (2009) Systematic literature reviews in software engineering—a systematic literature review. *Inf Softw Technol* 51(1):7–15. <https://doi.org/10.1016/j.infsof.2008.09.009>
25. Chen M, Saad W, Yin C (2019) Liquid state machine learning for resource and cache management in LTE-U unmanned aerial vehicle (UAV) networks. *IEEE Trans Wirel Commun* 18(3):1504–1517. <https://doi.org/10.1109/TWC.2019.2891629>

26. Liu X, Liu Y, Chen Y, Hanzo L (2019) Trajectory design and power control for multi-UAV assisted wireless networks: a machine learning approach. *IEEE Trans Veh Technol* 68(8):7957–7969. <https://doi.org/10.1109/TVT.2019.2920284>
27. Liu J, Sha N, Yang W, Tu J, Yang L (2020) Hierarchical Q-learning based UAV secure communication against multiple UAV adaptive eavesdroppers. *Wirel Commun Mob Comput* 2020. <https://doi.org/10.1155/2020/8825120>
28. Yu Q, Han C, Wang J, Bai L (2019) Low complexity hybrid beamforming for MmWave-UAV communication systems with a pre-defined codebook. *ACM Int Conf Proceeding Ser* 86:109–115. <https://doi.org/10.1145/3341016.3341036>
29. Boschiero M, Giordani M, Polese M, Zorzi M (2020) Coverage analysis of UAVs in millimeter wave networks: a stochastic geometry approach. <https://doi.org/10.1109/IWCMC48107.2020.9148550>
30. Wang B, Li S, Gao X, Xie T (2021) UAV swarm confrontation using hierarchical multiagent reinforcement learning. *Int J Aerosp Eng* 2021. <https://doi.org/10.1155/2021/3360116>
31. Sun E, Qu H, Yuan Y, Li M, Wang Z, Chen D (2021) A joint channel allocation and power control scheme for D2D communication in UAV-based networks. *Int Conf Commun Technol Proc ICCT 2021*:919–924. <https://doi.org/10.1109/ICCT52962.2021.9658058>
32. Niu H, Zhao X, Hou L, Ma D (2021) Energy efficiency maximization for UAV-assisted emergency communication networks. *Wirel Commun Mob Comput* 2021. <https://doi.org/10.1155/2021/7595347>
33. Duangsuan S, Juengkittikul P, Myint Maw M (2021) Path loss characterization using machine learning models for GS-to-UAV-enabled communication in smart farming scenarios. *Int J Antennas Propag* 2021. <https://doi.org/10.1155/2021/5524709>
34. Tang C, Chen L, Wang Y, Yang W, Chen R, Wang Z (2021) The role of 5G network image information based on deep learning in UAV prediction target trajectory tracking. *Wirel Commun Mob Comput* 2021. <https://doi.org/10.1155/2021/3097031>
35. Li A, Han H, Yu C (2021) Secrecy energy-efficient UAV communication via trajectory design and power control. *Wirel Commun Mob Comput* 2021. <https://doi.org/10.1155/2021/9969311>
36. Liu H (2022) An UAV-assisted edge computing resource allocation strategy for 5G communication in IoT environment. *J Robot* 2022. <https://doi.org/10.1155/2022/9397783>
37. Cao Z (2022) Optimization design of multi-UAV communication network based on reinforcement learning 2022
38. Liu X, Liu Y, Chen Y (2019) Reinforcement learning in multiple-UAV networks: deployment and movement design. *IEEE Trans Veh Technol* 68(8):8036–8049. <https://doi.org/10.1109/TVT.2019.2922849>
39. Wang L, Wang K, Pan C, Xu W, Aslam N, Nallanathan A, Trajectory control for UAV-assisted mobile edge computing, pp 1–15
40. Goudos SK, Athanasiadou G (2019) Application of an ensemble method to UAV power modeling for cellular communications. *IEEE Antennas Wirel Propag Lett* 18(11):2340–2344. <https://doi.org/10.1109/LAWP.2019.2926784>
41. Goudos SK, Tsoulos GV, Athanasiadou G, Batistatos MC, Zarbouti D, Psannis KE (2019) Artificial neural network optimal modeling and optimization of UAV measurements for mobile communications using the L-SHADE algorithm. *IEEE Trans Antennas Propag* 67(6):4022–4031. <https://doi.org/10.1109/TAP.2019.2905665>
42. Wang JL, Li YR, Adege AB, Wang LC, Jeng SS, Chen JY (2019) Machine learning based rapid 3D channel modeling for UAV communication networks. 2019 16th IEEE Annu. Consum Commun Netw Conf CCNC 2019:1–5. <https://doi.org/10.1109/CCNC.2019.8651718>
43. Zhang D et al (2018) Automatic modulation classification based on deep learning for unmanned aerial vehicles. *Sensors (Switzerland)* 18(3). <https://doi.org/10.3390/s18030924>
44. Alfaia RD, de F. Souto AV, Cardoso EHS, de Araújo JPL, Francês CRL (2022) Resource management in 5G networks assisted by UAV base stations: machine learning for overloaded Macrocell prediction based on users' temporal and spatial flow. *Drones* 6(6). <https://doi.org/10.3390/drones6060145>

45. Zhao P, Liu X, Liu Y, Chen Y, Chai KK (2019) Machine learning for position prediction and determination in aerial base station system. *IEEE Int Conf Commun* 2019:1–6. <https://doi.org/10.1109/ICC.2019.8761117>
46. Xiao K, Zhao J, He Y, Yu S (2019) Trajectory prediction of UAV in smart city using recurrent neural networks. In: *IEEE international conference on communications*, vol 2019. <https://doi.org/10.1109/ICC.2019.8761110>
47. Ren H, Li L, Xu W, Chen W, Han Z (2019) Machine learning-based hybrid precoding with robust error for UAV mmWave massive MIMO. *IEEE Int Conf Commun* 2019:1–6. <https://doi.org/10.1109/ICC.2019.8761112>
48. Zhu X, Qi F, Feng Y (2020) Deep-learning-based multiple beamforming for 5G UAV IoT networks. *IEEE Netw* 34(5):32–38. <https://doi.org/10.1109/MNET.011.2000035>
49. Karra D, Goudos SK, Tsoulos GV, Athanasiadou G (2019) Prediction of received signal power in mobile communications using different machine learning algorithms: a comparative study. *5th Panhellenic Conf. Electron Telecommun PACET* 2019:2–5. <https://doi.org/10.1109/PACET48583.2019.8956271>
50. Salman A, Almekdad S, Alhariri M (2019) Forming high gain wide beam with NULL steering ability to receive UAV signal based on phased array antenna. *J Commun Softw Syst* 15(2):186–192. <https://doi.org/10.24138/jcomss.v15i2.673>
51. Wang L, Che YL, Long J, Duan L, Wu K (2019) Multiple access MmWave design for UAV-aided 5G communications. *IEEE Wirel Commun* 26(1):64–71. <https://doi.org/10.1109/MWC.2018.1800216>
52. Chiang HL, Chen KC, Rave W, Khalili Marandi M, Fettweis G (2021) Machine-learning beam tracking and weight optimization for mmWave multi-UAV links. *IEEE Trans Wirel Commun* 20(8):5481–5494. <https://doi.org/10.1109/TWC.2021.3068206>
53. Susarla P et al (2020) Learning-based trajectory optimization for 5G mmWave uplink UAVs. In: *2020 IEEE international conference on communications workshops (ICC Workshops)*, vol 2020, pp 1–7. <https://doi.org/10.1109/ICCWorkshops49005.2020.9145194>
54. Liu C, Yuan W, Wei Z, Liu X, Ng DWK (2021) Location-aware predictive beamforming for UAV communications: a deep learning approach. *IEEE Wirel Commun Lett* 10(3). <https://doi.org/10.1109/LWC.2020.3045150>
55. Xin B, Wang F, Zhai Z (2021) Balwin-teaching-learning-based artificial raindrop algorithm for UAV route planning. *Math Probl Eng* 2021. <https://doi.org/10.1155/2021/8865403>
56. Munaye YY, Lin HP, Adege AB, Tarekegn GB (2019) Uav positioning for throughput maximization using deep learning approaches. *Sensors (Switzerland)* 19(12). <https://doi.org/10.3390/s19122775>
57. Yang H, Zhao J, Xiong Z, Lam KY, Sun S, Xiao L (2021) Privacy-preserving federated learning for UAV-enabled networks: learning-based joint scheduling and resource management. *IEEE J Sel Areas Commun* 39(10):3144–3159. <https://doi.org/10.1109/JSAC.2021.3088655>
58. Shafik W, Matinkhah SM, Afolabi SS, Sanda MN (2021) A 3-dimensional fast machine learning algorithm for mobile unmanned aerial vehicle base stations. *Int J Adv Appl Sci* 10(1). <https://doi.org/10.11591/ijaas.v10.i1.pp28-38>

Transmission of Wi-Fi Signals on 11 kV Power Transmission Lines—A Laboratory Study



B. Subba Reddy, L. Umanand, Vishram Pandit, Parthasarathy Ramaswamy, G. Nithin Reddy , M. Ramez Halloum, and Zoya Zaki

Abstract The paper addresses the low internet penetration in rural areas of India by proposing a novel technique to improve signal reach and quality using existing power lines. It highlights the collaborative efforts of the government and private sectors in bridging the digital divide. An innovative approach to extend internet connectivity in rural India by utilizing power lines for transmitting Wi-Fi signals is presented, where standard Wi-Fi signals are transmitted as surface waves through the power lines. A specialized launcher is developed and tested under different field conditions to ensure its reliability in practical scenarios. Additionally, the research investigates the influence of weather conditions on the quality of signal transmission. Based on the findings, the authors have provided recommendations for optimizing the technique and improving signal quality. Overall, the research provides valuable insights into the performance of the proposed technique, identifies challenges related to weather conditions, and suggests avenues for future exploration.

Keywords Wi-Fi · Radio · Launcher · Transverse electromagnetic waves · Surface wave propagation · Power line communication

B. Subba Reddy · L. Umanand · G. Nithin Reddy (✉) · M. Ramez Halloum · Z. Zaki
Indian Institute of Science, Bengaluru, Karnataka 560012, India
e-mail: nithinreddy@iisc.ac.in

B. Subba Reddy
e-mail: sreddy@iisc.ac.in

L. Umanand
e-mail: lums@iisc.ac.in

M. Ramez Halloum
e-mail: mramezh@iisc.ac.in

V. Pandit · P. Ramaswamy
Intel Corporation, Bengaluru, Karnataka 560103, India
e-mail: Vishram.pandit@intel.com

P. Ramaswamy
e-mail: ramaswamy.parthasarathy@intel.com

1 Introduction

With the rapid advancement of digitization, the demand for data and connectivity has seen exponential growth. Internet service providers have been meeting this demand by offering various types of connections, including the latest 5G technology. However, despite the phenomenal growth of digital infrastructure, a significant disparity persists between urban and rural areas regarding internet access. Even today, the rural parts of the country suffer from poor or non-existent internet connectivity. One of the primary reasons for this disparity is the lack of viability for service providers due to the low Average Revenue Per User (ARPU) in rural areas. India has approximately 450,000 villages and 150,000 g Panchayats. While fiber connections have been established in a substantial number of these Panchayats, the end users residing in the surrounding villages, typically 3–5 km away, remain unconnected. In response to this challenge, the Government of India has taken proactive steps to address the issue. They have developed a strategy that focuses on laying fiber connections in a significant number of Gram Panchayats. While this initiative helps solve some of the administrative aspects, it still leaves the end users in the surrounding villages without access to reliable internet connectivity. Despite the remaining challenges, the government's efforts demonstrate a commitment to bridge the digital divide in rural areas. Further measures and innovative solutions are required to extend the reach of internet services beyond the Gram Panchayats and ensure that the surrounding villages can also benefit from the advantages of digital connectivity.

The requirement to establish a viable and affordable connectivity solution for connecting Gram Panchayats (GPs) to surrounding villages remains unfulfilled. The solution needs to be cost-effective, resilient, and easily implementable. In order to address this challenge, the authors have proposed leveraging the existing infrastructure of power lines to transmit internet traffic from the GP to these villages, known as Wireless over Wire (WoW) technology. Implementing this solution requires the development of suitable technologies and equipment that can enable the transmission of internet traffic over power lines without interfering with the electrical power distribution. Additionally, careful planning and coordination with relevant stakeholders, including power companies and ISPs, will ensure successful implementation and widespread adoption. A pictorial view of the architecture is shown in Fig. 1.

Transmission lines serve as waveguides at the fundamental level. The conductor's surface wave propagation is made possible by a launcher, which transforms radio impulses into transverse electromagnetic (TEM) waves. Like other waves, anomalies like bends, taps, and discontinuities cause signal power losses, eventually lowering link performance. These losses are enhanced by the length of the line. To counteract the signal power losses and maintain consistent performance in the presence of these anomalies, the introduction of a repeater becomes necessary, as shown in Fig. 2. By amplifying the signal, the repeater ensures that the desired performance levels are sustained, overcoming the detrimental effects during signal transmission.



Fig. 1 Block diagram of Wireless over Wire (WoW) architecture

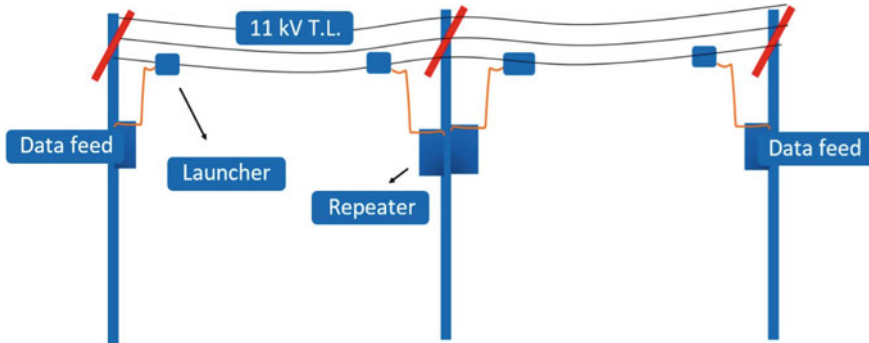


Fig. 2 Surface wave propagation along a power transmission line with repeaters

Theoretical and experimental work on surface wave propagation over single-conductor transmission line (SCTL) has been extensively reported by many researchers. Goubau introduced the single-conductor surface wave transmission line to lines having a dielectric-coated conductor [1]. He has also introduced the conductivity and dielectric losses in the line as a function of the thickness of the dielectric layer and the dependence of surface wave frequencies on the losses. Akalin et al. discussed using single-wire transmission lines for high-frequency applications in the terahertz range [2]. Traditional transmission lines are typically designed with multiple conductors, but the authors in [2] focused on the unique properties and advantages of single-wire configurations in the terahertz frequency range. The study presents design considerations and techniques for implementing efficient single-wire transmission lines at terahertz frequencies. The impact of different factors, such as wire diameter, dielectric materials, and impedance matching, on the performance

of the transmission lines has been explored. Elmore provided an overview of the propagating wave phenomenon on a single conductor in transmission line systems [3]. In the case of an unshielded single-conductor transmission line (SCTL), the surface wave fields originating from the wire axis spread out and experience attenuation in the transverse direction. Additionally, the wave also undergoes attenuation in the longitudinal direction due to several factors, including the finite resistivity of the wire, as well as dielectric losses and leakage in the surrounding medium [3, 4]. A non-intrusive solution for fault detection in power transmission systems using surface wave sensors was discussed in [5]. Challenges such as environmental influences, signal interference, and the need for calibration and maintenance of surface wave sensor technology were some of the limitations associated with it.

2 Overview of the Launcher for Surface Wave Transmission

The primary objective of the launcher in this technology is to establish transverse electromagnetic (TEM) waves, also known as surface waves, on the power transmission line. One of the applications of surface wave technology is being used in telemetry during the borehole drilling process. Amjadi et al. [6] presented a novel design and implementation of a compact launcher for a single-conductor transmission line to transmit data and signals from downhole sensors to the surface for real-time monitoring and control. The launcher efficiently converts radio frequency (RF) signals into transverse electromagnetic (TEM) waves, enabling the transmission of signals along the single-conductor line. A novel coplanar Vivaldi-style launcher designed for Goubau single-wire transmission lines is introduced in [7]. Wide bandwidth and directional radiation properties of Vivaldi-style antenna were utilized for designing the launcher specifically tailored for Goubau single-wire transmission lines. Several designs in the form of horn-shaped structures have been suggested as single-wire (SW) launchers [8–11]. These designs offer the advantage of field concentration along the transmission line and possess a symmetrical shape around the wire.

In general, the design of a standard launcher takes into consideration several important criteria, including:

- *Efficient TEM Signal Generation:* The launcher should efficiently convert radio (RF) signals into TEM signals to ensure adequate transmission along the power line.
- *Electrical Stress Endurance:* The launcher must be capable of withstanding high voltage, high current, voltage surges, current surges, and other electrical stresses commonly encountered on power lines.
- *Weather Resistance:* The launcher should be designed to operate reliably under various weather conditions, ensuring consistent performance regardless of environmental factors such as temperature, humidity, and precipitation.

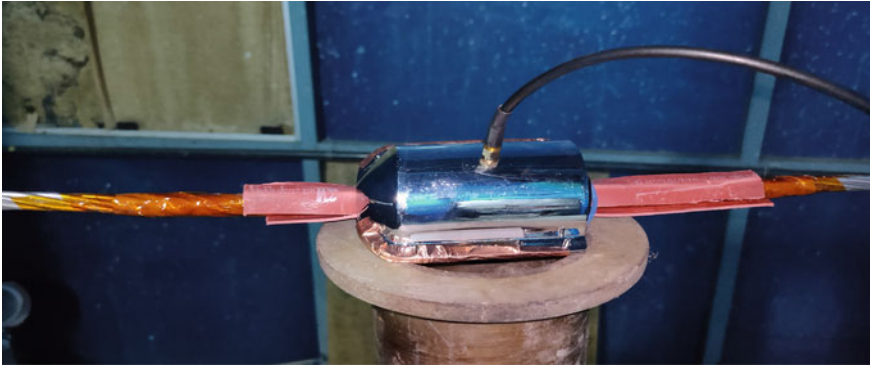


Fig. 3 Developed launcher used for WoW technology

- *Mechanical Requirements:* The launcher needs to meet specific mechanical requirements to enable its installation on a power line and ensure secure attachment. It should be designed to withstand vibrations, wind loads, and other mechanical stresses that may be present in the installation environment.
- *Quick and Easy Installation:* The launcher should be designed for quick and efficient mounting on the power line to minimize power disruptions during installation. It involves considering factors such as ease of access, simplicity of mounting mechanisms, and streamlined installation procedures.

Considering the above criteria in designing the launcher, the authors have considered a unique launcher profile for surface wave propagation of Wi-Fi signals under WoW technology, as shown in Figs. 3 and 4. Initially, different configurations of launchers have been evaluated to understand the failure of insulation under environmental conditions such as rain/fog. Initially, the launcher insulation comprised XPLE sleeves over a Kapton tape stuck on the conductor. After the insulation failure under moist conditions at 11 kV, the insulation was replaced with silicone rubber insulation at both transmitting and receiving side launchers. Further, SBRS heat-shrinking insulation above FRP rods was introduced for improved performance under moisture conditions. Under moist conditions, water droplets accumulate on the conductor changing the dielectric nature of the transmitting medium.

3 Experimental Investigations

The WoW configuration consisting of transmitter and receiver launchers on a aluminum rabbit-type conductor is evaluated for 11 kV AC at the high voltage laboratory, Indian Institute of Science. The supply to the launchers is provided by 12.8 V, 72 Ah battery. The transmitter and receiver launchers are connected to transmitting

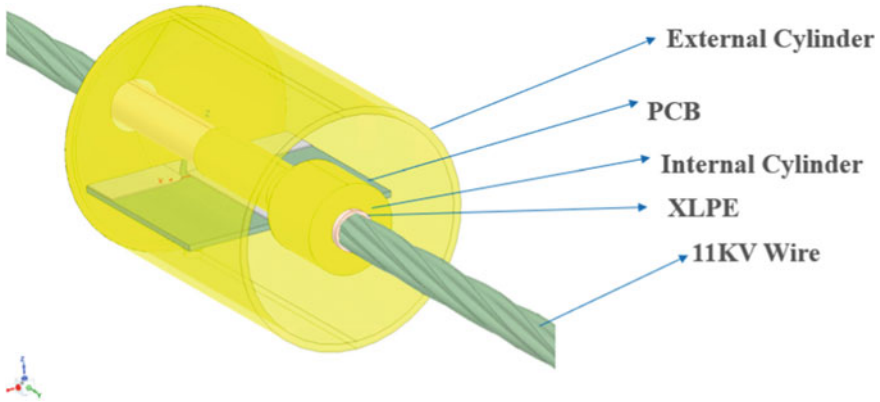


Fig. 4 Cross-section view of launcher used for WoW technology

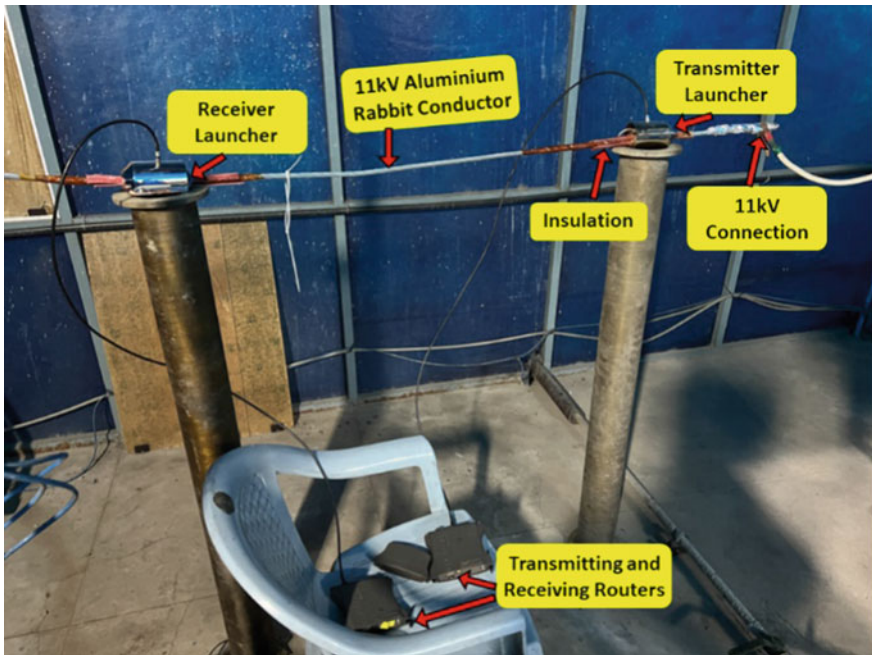


Fig. 5 Experimental setup for high voltage evaluation of WoW technology

and receiving routers. The Wi-Fi signal strength from the routers is observed wirelessly with the help of a laptop. The experimental setup for high voltage testing for the WoW technology is shown in Fig. 5.

3.1 Experimentation Under Normal Ambient Conditions

The entire setup was placed in a 5 m × 5 m × 5 m rain/fog chamber and energized for an 11 kV AC source. The high voltage source comprises a 150 kV/300 kVA transformer and a regulator unit. The regulator unit is designed to deliver a rated output of 400 V and 750 A, powered by a 400 V, two-phase power frequency input. The environmental parameters, including relative humidity and temperature, were measured and recorded during the experimentation using a Fluke-971 Temperature and Humidity Meter.

3.2 Experimentation with Simulated Fog Conditions

The same setup was placed in a 5 m × 5 m × 5 m rain/fog chamber and energized for 11 kV AC with the same AC source. The IEEE Std fog nozzles are operated by allowing water (conductivity of 440 $\mu\text{S}/\text{cm}$) and compressed air. The entire test setup for moisture application is shown in Fig. 6.

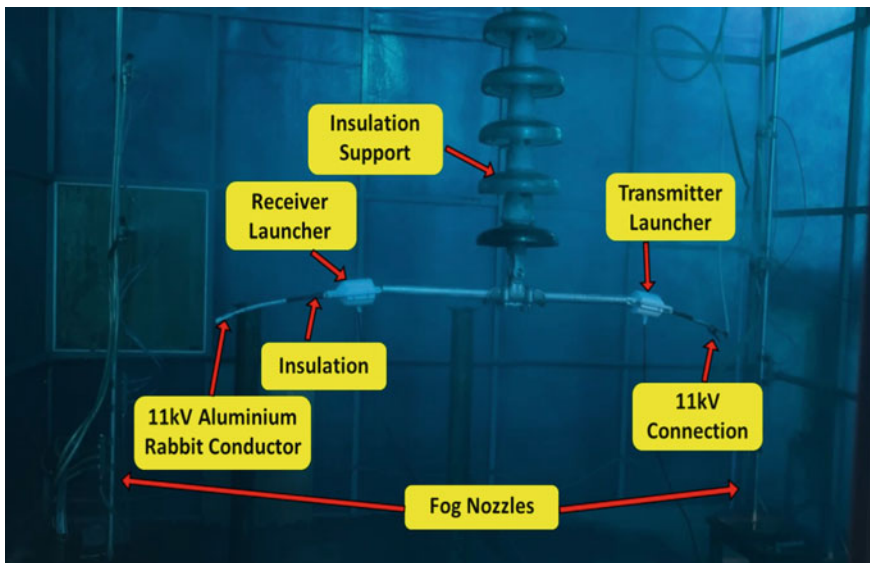


Fig. 6 Experimental setup for high voltage testing of WoW technology under moist/fog conditions

4 Preliminary Results and Discussion

The WoW technology setup comprised of transmitter and receiver launchers placed on an 11 kV aluminum rabbit-type conductor separated by different insulations is evaluated with charged line at 11 kV AC high voltage for normal dry conditions and with fog/ moist environment.

Some of the preliminary throughput results obtained Wi-Fi signals are shown in Table 1, and typical data signals are presented in Figs. 7, 8, 9, and 10.

Table 1 Preliminary throughput of WoW technology

Condition	Average current throughput (Mbps)
Without 11 kV (dry condition)	289.3
With 11 kV (dry condition)	294.8
Without 11 kV (moist condition)	320.9
With 11 kV (moist condition)	327.5

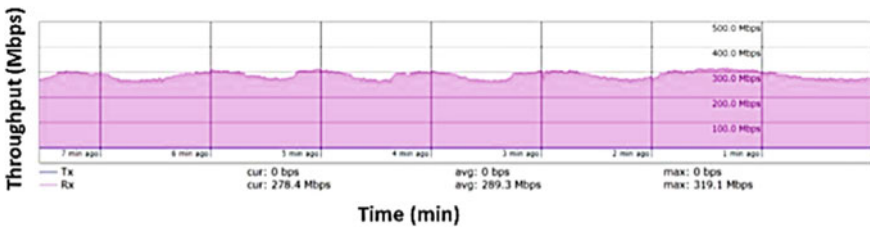


Fig. 7 Data signal without 11 kV (dry condition)

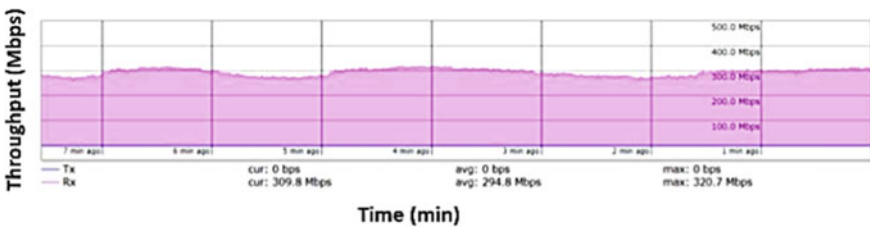


Fig. 8 Data signal with 11 kV (dry condition)

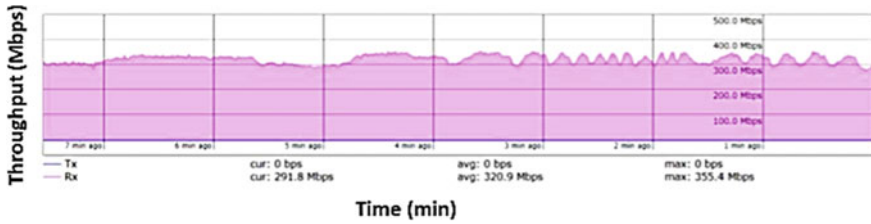


Fig. 9 Data signal without 11 kV (moist condition)

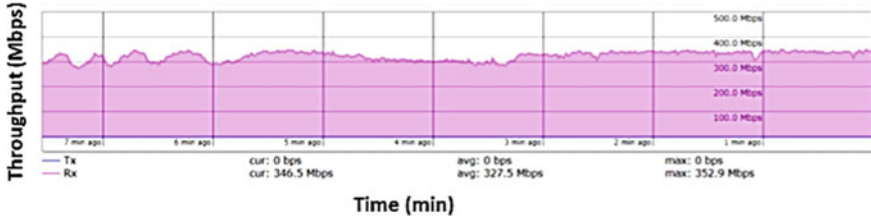


Fig. 10 Data signal with 11 kV (moist condition)

5 Conclusion

The paper introduces an innovative approach to enhance internet connectivity, particularly for rural areas and for the last mile connectivity residents of rural India by utilizing available power lines as a medium for transmitting Wi-Fi signals, known as WoW technology. The proposed technology involves the transmission of standard Wi-Fi signals as surface waves through the existing power line infrastructure. A specialized launcher is designed and subjected to extensive testing in various field conditions to ensure its reliability and performance in real-world scenarios. Launchers placed on an 11 kV conductor are evaluated for various test conditions for different insulations.

Preliminary results obtained under normal and under fog conditions are presented. Further work is in progress on evaluating the developed launchers for better insulation withstand capabilities, for long-term field performance, for different voltage/current loadings, etc.

Acknowledgements The authors are thankful to the Indian Institute of Science, Bangalore, and Intel Corporation Bengaluru, for the support and help.

References

1. Goubau G (1951) Single-conductor surface-wave transmission lines. *Proc IRE* 39(6):619–624
2. Akalin T et al (2006) Single-wire transmission lines at terahertz frequencies. *IEEE Trans Microw Theory Tech* 54(6):2762–2767
3. Elmore G (2009) Introduction to the propagating wave on a single conductor. <http://www.corridor.biz/FullArticle.pdf>
4. Elmore G (2012) A surface wave transmission line. <http://www.sonic.net/~n6gn/ELMORE.pdf>
5. Alam MN et al (2013) Design and application of surface wave sensors for nonintrusive power line fault detection. *IEEE Sens J* 13(1):339–347
6. Amjadi SM, Sarabandi K (2016) A compact single conductor transmission line launcher for telemetry in borehole drilling process. *IEEE international symposium on antennas and propagation (APSURSI) 2016, Fajardo, PR, USA*, pp 2075–2076
7. Sharma A et al (2017) A coplanar Vivaldi-style launcher for Goubau single-wire transmission lines. *IEEE Antennas Wirel Propag Lett* 16:2955–2958
8. Sharma A et al (2017) Long-range battery-free UHF RFID with a single wire transmission line. *IEEE Sens J* 17(17):5687–5693
9. Liu J et al (2020) A surface wave exciter adapting to different-diameter lines and its application to single power line communications. *Int J RF Microwave Comput Aided Eng* 30(3):e22102
10. Schaich T et al (2022) Advanced modeling of surface waves on twisted pair cables: surface wave stopbands. *IEEE Trans Microw Theory Tech* 70(5):2541–2552
11. Dinc E et al (2021) Launching surface waves with classical antennas. In: *ICC 2021—IEEE international conference on communications*, Montreal, QC, Canada, pp 1–6

An Intelligent LoRa-Based Wireless Sensor Network Mesh Architecture to Improve Precision Agriculture



Aman Shaikh, Maya Shelke, Satayush Rai, Pranil Ashok Rao,
and Ganesh Shinde

Abstract In today's world with evolving dynamics brought into the field of agriculture, there have been quite a few advancements unified into agrarian practices by the incorporation of modern technologies and methods. IoT, AI/ML, and other such technologies are sweeping the entire globe due to their ability to provide precise, reliable, consistent, and accurate outcomes to make data-driven decisions. Integration of IoT and ML models in order to perform data analytics was a natural choice made to have a rigorous improvement in the farm's yield quantity and quality with the intention to go above and beyond in terms of meeting the food demand. This survey intends to showcase how the culmination of various aspects of agrarian necessities cooperates for the synchronization of computational capabilities provided by the ML models and data generated via the use of IoT sensors, communicated over the LoRa-based WSN networks. This paper proposes an integrated system based on a three-tier architecture that incorporates ML and IoT systems to process data for yield prediction and is consequently further used to generate a visual representation of the outcome. A novel approach to ensure the sustainability of yield both in terms of production and grading standards has been presented.

Keywords Precision farming · Mesh networking · Machine intelligence · Embedded systems · IoT

A. Shaikh (✉) · M. Shelke · S. Rai · P. A. Rao · G. Shinde
Department of Information Technology, JSPM Rajarshi Shahu College of Engineering, Pune,
India
e-mail: amanshaikhw@gmail.com

M. Shelke
e-mail: mayabembde07@gmail.com

S. Rai
e-mail: raisatayush7232@gmail.com

P. A. Rao
e-mail: pranil3005@gmail.com

G. Shinde
e-mail: ganeshshinde72620@gmail.com

1 Introduction

In today's world where the population has expanded from 7.250 billion in 2013 to 8.045 billion, food demand has also steadily risen through the growth rate has dipped from 1.24 to 0.88% over the decade [1, 2]. However, the global trend in food production has seen better. It is projected that by the year 2050, the population shall be approaching the mark of 10 billion [3]. With so many mouths to feed the existing food production is nowhere near what is required. Though food production is being rapidly scaled, the quality of yield has been affected, delivering only in terms of calories and cosmetic perfection rather than nutrition. This has led to a phenomenon called hidden hunger. Apart from this, we are also facing a triple planetary crisis due to unsustainable patterns of production and consumption [4]. The COVID-19 crisis was a revelation of the fragility of our agricultural production which contributed to the global food crisis [5, 6]. Changing climatic conditions and natural calamities have coerced us to innovate. This led to a focused approach of precision agriculture which zeroes in on reducing production whilst maintaining quality. Precision agriculture stands on the pillars of technologies such as AI/ML, IoT, and Cloud Computing amongst others. Data gathered via the IoT sensors are transmitted over the network. The stored data is then fed into the ML model which processes it and provides us with appropriate outcomes. Optimal technological use will ensure better results and will be beneficial to agricultural production. In [7, 8] authors have discussed various ML algorithms that could possibly be used for yield prediction with the best fit discussed. The authors have investigated H-ANN for yield prediction and demonstrated the benefits of this algorithm over the others. The authors of [8] present an ML approach to optimize water usage for irrigation purposes in agriculture. The authors [9] have proposed the use of IoT devices for LoRa communication which fared better compared to the others in terms of device cost, power consumption, and range offered combined. In [10] authors have proposed a master–slave mesh network with the LoRa server acting as the master and the LoRa gateway transmitting data to the MQTT which then uploads data to a remote cloud server. The authors of [11] have proposed an ML model that processes sensory data to recommend the irrigational needs of agricultural land. This literature review provides insights into how modern technology and advanced agricultural methodologies enable precision agriculture to succeed in terms of cost, efficiency, and accuracy. This review focuses on the aspects mentioned below:

- Use and applicability of IoT sensors in precision agriculture and comparison of various IoT devices in terms of communication capabilities and possible data gathering method/s.
- The applications of ML algorithms in concepts regarding yield prediction, weed detection, optimized irrigation, and others are discussed.
- Further, an overview of modifications done in order to integrate IoT and ML has been explored.
- Lastly, the possible challenges and future trends have been mentioned in precision agriculture.

Section 1 provides us with an idea about techniques and technologies that go into precision agriculture. Section 2 covers the proposed system with it further divided into subsections. Sections 1, 2, and 3 briefly explain Tier-1, Tier-2, and Tier-3 of the three-tier architecture. Section 3 explores the global predictive module and local predictive model being implemented along with the WSN node and its modules being elucidated. Section 4 goes over the expected results for the proposed system. Section 5 states the concluding summarized remarks and expected future trends. Section 6 states the referenced materials used to put together this review article.

2 Literature Survey

There exists exhaustive literature on precision agriculture. This survey studies 3 essential approaches currently considered in precision agriculture trends. This includes AI/ML-based approaches for crop yield prediction and recommending decisions to optimize farming. The second approach includes IoT/WSN devices to amass environmental data and monitor agriculture; lastly, the overall integration of WSN and AI/ML-based solutions was used.

Reddy et al. [8] has described various models utilized in crop yield prediction with existing models using algorithms such as CNN, KNN, ANN, MCNN, H-ANN, Random Forest, MLR, Neural Networks, LSTM, and DNN commonly. The crops were analyzed against different algorithms which considered features such as soil nutrients, humidity, temperature, rainfall, precipitation, radiation, wind speed, and air pressure. The SVM classifiers were used to grade tomatoes and an accuracy of 97.7% was achieved. CNN and MCNN used for crop yield prediction gave an RMSE value of 1396.4 and a relative error of 8.465, though it reduced prediction error it resulted in reduced efficiency and unwanted relative error being introduced. H-ANN used for crop yield production prediction reflected the RMSE value of 4.72. Another approach for yield prediction was a combination of ANN and MLR, it gave RMSE values of 0.051, and the use of DNN for the same gave us a training RMSE value of 10.55 and validation RMSE value of 12.79. It inferred that data availability and feature selection heavily influenced the model's accuracy.

Gopal et al. [9] have explained the use of Hybrid ANN (H-ANN), a combination of MLR with ANN to improve the accuracy by incorporating the MLR coefficients and associated bias. The data for the study was collected from the state of Tamil Nadu in India. The dataset in consideration was cleaned using the techniques of missing value treatment, and outlier treatment. Various algorithms were used to obtain a set of features which then were used as input to the MLR model to attain the best feature subset. Ultimately features, such as area, temperature, number of open wells and available tanks, length of the canal, were considered. The overall performance was superior for the H-ANN in contrast to other algorithms such as SVR, KNN, RF, conventional MLR, and conventional ANN. To put it into context MLR and conventional ANN give RMSE value of 0.098 which is 0.037 more than the H-ANN, with error rates of 0.064 and 0.041 respectively, and correlation of 0.92

and 0.99, respectively. The selection of appropriate initial weights using MLR was instrumental in obtaining the near-global optimal minima.

In this paper, the authors Singh et al. [10] present applications of ML to optimize irrigation using soil moisture prediction considering factors like temperature, humidity, and weather forecast data from the Internet to make better real-time decisions for watering crops. Multiple linear regression (MLR), Elastic net regression (ENR), Gradient boosting regression tree (GBRT), and Random Forest regression (RFR) are discussed. Two strategies have been employed, the first one is in order to predict soil moisture, first, and the soil temperature is predicted using ENR and MLR models. Comparatively, ENR performed better with a 2.25 mean squared error in contrast with 3.4 from MLR. The second strategy does not use soil temperature. The soil moisture was predicted using both strategies and it was found that using predicted temperature as a parameter the GBRT model performed the best with MSR, and R-squared values of 4.04, 0.94 respectively in comparison to 19.26, 0.78 respectively in the case of RFR, and 40.51, 0.55 as for ENR. In the second strategy without using temperature as a parameter and using the GBRT model, we get an MSR of 5.78 and an R-squared value of 9.93 respectively. In conclusion, the GBRT model with predicted soil temperature performed the best in predicting soil moisture.

Suji Prasad et al. [11] have proposed a system that uses LoRa communication between IOT devices that are deployed in smart agriculture systems. Also, authors in [11] have compared LoRa, WiFi, Zigbee, and Bluetooth communication technologies and have concluded that LoRa consumes low power for communication, is cheap, and has a better range comparatively. The deployed sensor nodes measured the temperature, soil moisture, and humidity of the field. The data observed from the sensor nodes showed that the temperature is negatively correlated with humidity. This system helps farmers monitor the farmlands and assists to increase the yield. The current system can be improved further by using an unsupervised clustering algorithm for sensor data analysis, thus automating field control. The efficiency of the WSN nodes could also be improved by using mesh networks.

Chaopisit and Kaewpraek [12] and Queralta et al. [13] have proposed a system that uses a mesh networking system for LoRa-based WSN for agricultural fields. The proposed solution uses a master-slave mesh network where all the nodes send data to the LoRa server (Master), the server then transmits data to the LoRa gateway which uploads the data to a remote cloud server using MQTT. The nodes capture the temperature, soil moisture, and humidity of the field. Each node was deployed at a distance of 400 m² on the farmland. The nodes have 3 states: running, transmission, and deep sleep. In the running state, the node lasted for 55 h 24 Min, in transmission mode the node lasted for 39 h 8 Min, and in deep-sleep mode, the node lasted for 193 h 54 Min. The proposed system was able to communicate at a distance of 850 m with a signal strength of 101 dBm. The proposed system can be further improved by designing a better antenna design that consumes less power and provides a longer range. The system also lacks a web dashboard or mobile application for monitoring. The nodes also need a solar energy-based rechargeable battery.

Bhoi et al. [14] proposed and implemented a three-level architecture for irrigation recommendations using ML. The three levels include the Crop field level, Cloud

level, and User level. At the crop field level, different sensors were deployed on the field to monitor soil moisture, soil temperature, air temperature, and humidity. Each sensor collects the data twice a day and sends it to the Aurdino. Aurdino forwards the data to the storage at the cloud level. The stored data is then passed to the ML models. The ML model has three parts. First is a regression model which helps to predict the parameters in advance. Further, the values are passed to a clustering model that uses a binary classification algorithm to categorize data into clusters of two in order to reduce errors. The first part is irrigation required and the second is irrigation not required. At the user level, a mobile application is developed to display all the information about the crop field to the user in real time. Thus the three-part ML model can be used to effectively and efficiently help recommend the irrigational requirements of the farm.

Kanuru et al. [15], in this paper discuss the use of IoT, ML, and a GPS module to determine the type and amount of fertilizer to use. The soil type, region, and crop type are considered to determine the fertilizer. The ML models such as KNN (unsupervised ML algorithm), the Naïve Bayesian model (classification algorithm), and lastly, the linear SVC (supervised ML algorithm) have been explored. The implementation consisted of 4 parts: obtaining relevant datasets, processing data, finding a correlation between soil type and chemical used, and using ML on data. After implementing all the 3 ML algorithms stated above, it was observed that KNN had the highest accuracy of 0.8145 followed by linear SVC with 0.777, and finally Naïve Bayes with 0.759. Hence the fertilizer selection, based on region and crop type reduces the wastage of fertilizer and saves cost.

From the existing literature survey, we conclude that there are various techniques involved in the field of precision agriculture but there is a need for a standard architecture that provides all the necessary technological components which work in synergy.

3 Proposed System

With the ever-expanding population, the food demand keeps increasing hence there is a need to develop a system to meet the demands. The reduced agricultural land area has resulted in farming needs being fulfilled by farmlands at remote locations where technology such as GPS/GSM/GPRS have not yet been developed and large lands are present for farming. Thus considering the above concerns there is a need for a precision agriculture system that can be deployed across farms at remote locations that are easily scalable. The proposed solution consists of a 3-tier architecture as shown in Fig. 1.

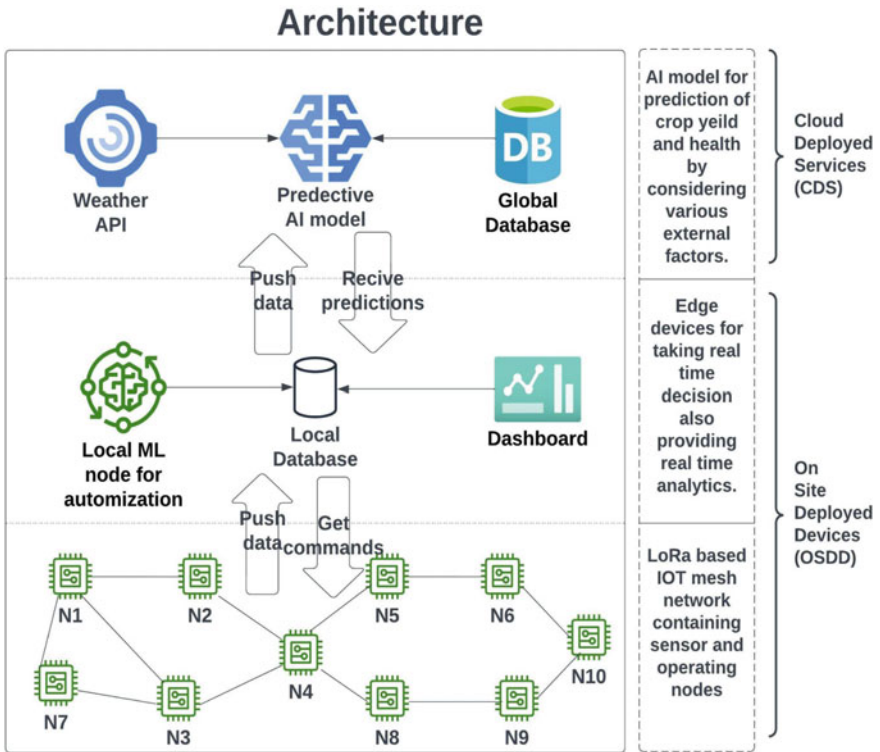


Fig. 1 System architecture

Tier 3.1 consists of:

- 3.1.1 Weather API server.
- 3.1.2 Global database.
- 3.1.3 Predictive AI model.

Tier 3.2 Consists of:

- 3.2.1 Real-time local database.
- 3.2.2 Local ML models.
- 3.2.3 Analytical dashboard.

Tier 3.3 Consists of:

- 3.3.1 Actuator nodes.
- 3.3.2 Sensing nodes.
- 3.3.3 Sink nodes.

Tier 3.1:

3.1.1 Weather API server: This system consists of a server that sends timely requests to the weather server for real-time data of the weather conditions of the land location. The data gathered includes humidity, temperature, wind speed, chances of rain/drought or a natural calamity, etc.

3.1.2. Global database: this system consists of a MongoDB server that stores the data from the weather API and the local database periodically after 12 h.

3.1.3. Predictive ML model: The ML model receives data from the global server. The input data consists of two contexts: the global context in which the data is gathered from the weather stations via API and the on-field data such as soil temperature, humidity, and wind speed. The model works on these parameters to provide suggestions that will help the farmer prepare for the predicated conditions.

Tier 3.2:

3.2.1. Real-time local database: this database stores the real-time data gathered by the WSN nodes which consist of soil temperature, humidity, soil NPK values, and wind speed.

3.2.2. Local ML models: There are 3 ML models that generate appropriate results to assist decision-making. In the yield prediction model, multiple features are taken into account to predict the agricultural yield from the farmland. The features in question range from soil health, environmental elements, agricultural practices, and the size of farmland. The soil moisture prediction model paired with weather forecasting data will make better decisions for water irrigation to the soil and improve the efficiency of water usage. The soil fertility model will be used for making decisions on fertilizer needs. The soil type and crop type are the main features that the model will consider for determining the type of fertilizer to use.

3.2.3. Analytical dashboard: A dashboard for the farmer to get the farm statistics in real-time via a global database which is connected to the WSN on-field sensor nodes. Each node extracts the soil data such as NPK values and moisture every 30 min and which in turn is used by the ML model to predict the crop yield which the farmer can refer to amongst other data displayed. The dashboard also queries the locally stored database to get updated farm results faster. The dashboard will be provided via a mobile application which will be implemented by Flutter framework and can run on the website as well.

Tier 3.3:

3.3.1 Actuator nodes: This is a type of WSN node that senses the environment upon receiving the commands from the sink node, it takes the decision accordingly. Here the actuator node consists of an electromagnetic switch through which large appliances, such as Motors, Pumps, can be operated.

3.3.2. Sensing nodes: The node consists of a LoRa transceiver which is responsible for the communication of data between the sensing node and the mesh network. The

node consists of a separate microcontroller which is connected to the probes that measure soil humidity, and temperature, as well as a DHT11 sensor, and a wind speed sensor. The node also consists of a real-time clock that tracks observed values and timings.

3.3.3. Sink nodes: The sink nodes in the mesh network forward data from the nodes to the Internet. Each node except the sink node will try to find the best path to send the data to the sink node. The path may involve direct sink-node-to-sense-node communication or hop-to-hop communication until it reaches the sink node. There must be at least 4 sink nodes installed.

4 Implementation

The ML module is divided into 2 main parts.

1. Global predictive module.
2. Local ML model for automation.

1. Global Predictive Module:

This module uses a global database to make crop yield predictions based on various parameters. The global database houses the data received from sensory devices. For the yield prediction having a good dataset is important and with data coming in from the sensors as long as the sensors aren't faulty it's expected that the ML model predicts accurately. The data undergoes pre-processing to remove null values, missing values, values out of range, incorrect data types, etc. Next, is the feature extraction of data; features essential for yield prediction are extracted for informative and non-redundant data that expedites and facilitates subsequent steps of learning and generalization. The features include soil nutrient and moisture content, humidity, temperature, solar intensity, wind speed, atmospheric pressure, land size, and others. Once feature extraction is completed the dataset created, is split into learning and testing datasets (this step is done only for model creation). An ML algorithm is applied to the data in order to analyze data as per set constraints and provide us with relevant numbers for an accurate prediction for a given dataset. The obtained results include accuracy, RMSE value, as well as yield predicted.

2. Local ML Model:

The irrigational model automates the water irrigation systems and fertilizer systems. The ML model predicts the soil moisture based on the data received from sensors which are then stored in a local database, the data includes current soil moisture, soil temperature, humidity, and weather forecast data received from weather API. ML models like Gradient boosting regression tree (GBRT) can be used for such tasks.

Figure 2 shows the relationship between features like temperature, and water level at various depths. From the correlation heatmap, we can infer that temperature and

moisture level have a negative correlation thus if the temperature rises then correspondingly water level is expected to decrease, and also when the depth observation increases the above inference becomes stronger. The fertilizer model determines the type of fertilizer that has to be used based on the land type, plant type, and NPK values received from the sensor. ML models like KNN can be used to determine the type of fertilizer and the appropriate amount can be determined with Eq. 1 like those provided by the NSW department of primary industries.

$$F = \frac{N}{N_f} \times 100 \tag{1}$$

where:

- F* Amount of fertilizer in kg/ha
- N* Nutrient in kg/ha
- N_f* % of nutrient in fertilizer.

Figure 3 depicts the layout of a sample user interface for farmers to check the real-time statistics of the field along with NPK values. The second screen shows the extended navigations provided such as crop yield prediction, detailed statistics of farms, fertilizer content along with monthly analytics.

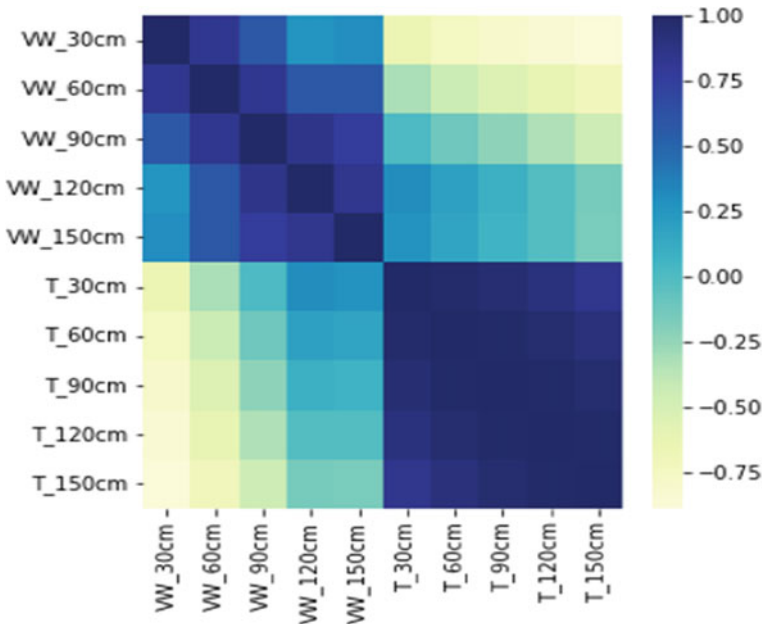


Fig. 2 Correlation heatmap of temperature and humidity at relative depths

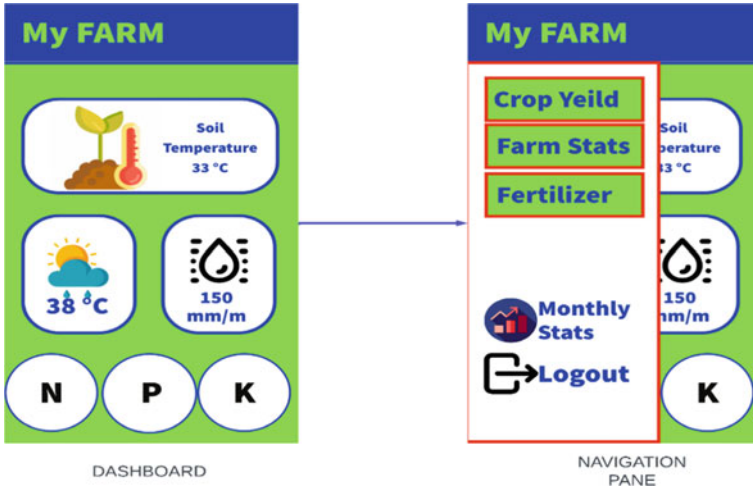


Fig. 3 Sample UI interface

In Fig. 4 the WSN nodes consist of a central Microcontroller interconnected with sensors, Communication modules, and Control devices. The whole system is powered by a Power Supply Unit.

The WSN nodes consist of the following modules:

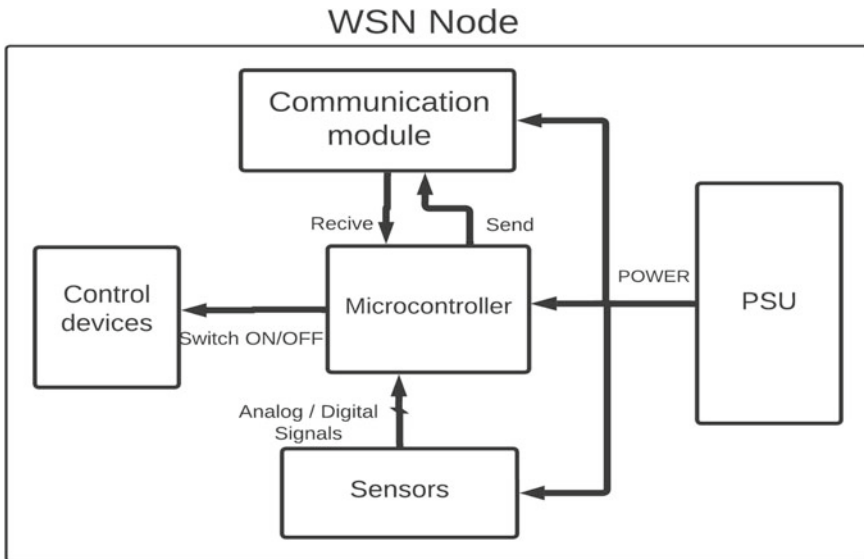


Fig. 4 Block diagram of WSN node

1. **Microcontroller:** The microcontroller is the heart of the WSN node, it is connected to the sensors that can either send digital or analog signals. It helps convert sensor analog signals into digital signals through its Analog to Digital Converters (ADCs). We have used an ESP32 microcontroller which uses a 32-bit ADC thus it has a higher sampling rate. IT also contains General Purpose Input Output (GPIO); these are physical pins on the microcontroller that act as I/O. The ESP 32 has a total of 25 GPIO pins out of which 15 are ADC, 21 are Pulse Width Modulation (PWM) pins, and digital output pins. The digital output pins can be used to control relays and other switching appliances while PWM pins can be used to control variable controls such as stepper motors or as a motor speed controller.
2. **Communication module:** This module handles all the communication requests done by the microcontroller and the other nodes. The communication technology can change as per the requirements of the mesh network but this module provides reliable data forwarding and data sending in a mesh topology.
3. **Sensors:** The sensors used in the WSN node consist of a DHT11 humidity and temperature sensor, a capacitive soil moisture sensor, and an airspeed measurement sensor.
4. **Control devices:** The control device may contain a relay, contactor, or an actuator that on receiving a command from the microcontroller performs a switching action to the connected appliance. The switching appliance may be a pump or a motor valve, etc.
5. **Power Supply Unit:** The PSU consists of a solar panel that charges the Li-ION batteries. The charging is monitored by a BMS circuit and the supply voltage of the solar panel is filtered with a buck converter. The batteries are supplied with a constant voltage of 5 V to charge. The PSU provides 5 V to the node.

Figure 5 is the 3D model that shows the WSN nodes deployed around the crops. The deployed nodes consist of a solar panel and the node itself. A sink node is situated at the center of the field which connects sensor nodes to the Internet. The node density can vary as per the communication medium used between nodes.

5 Results

The proposed system helps monitor the agrarian environment by measuring temperature, humidity, soil moisture, wind speed, and NPK values using the WSN mesh network. The data gathered is then effectively stored in the database which allows the use of an analytical program to display the real-time status of the farm with the help of a Dashboard or a Mobile app. The soil irrigation and fertilizer ML models provide timely recommendations for the appropriate amount of fertilizing resources to be used and also when to be used. This data is then forwarded to the bigger global data storage periodically. The predictive ML model uses the weather API to get the predicted weather data and uses the data from the global database to make a decision

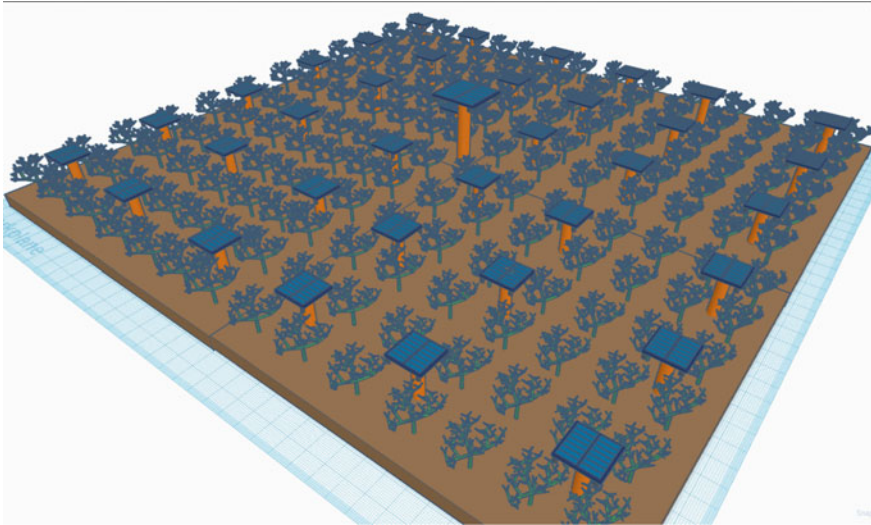


Fig. 5 3D model of the agriculture field

that will help farmers to be prepared for the coming days for example: If the ML model gets data from the weather API that high temperature is expected tomorrow then the model will check current soil moisture levels and available water quantity and prepare for the next day by filling water in the tank to an opposite amount.

6 Conclusion and Future Scope

The proposed architecture establishes a standard framework for building a WSN-based precision farming. The architecture helps improve the utilization of natural resources such as water, and fertilizer. It also helps improve the yield of agricultural land. The architecture also incorporates the weather forecasting ability which helps the farmer to prepare for future conditions in advance. Lastly, the dashboard helps to provide an accurate representation of the farm conditions and helps manage and improve the system easily if needed. The proposed system can be used in large as well as small farmlands and can be scaled as the need increases. Although the current solution works for most cases, there are some improvements that can enhance the functionality of the system. A standard WSN mesh protocol can be used to improve the node's battery life and increase the reliability of the communication in the mesh network. Different dashboard views can also be incorporated to provide perspective on the different aspects of the farm. The accuracy of the model can be further improved by using an unsupervised ML algorithm that can be trained on the gathered agricultural land data.

References

1. World Population 1950–2023
2. Countries in the world by population (2023)
3. When Will The World Run Out of Food
4. How modern food can regain its nutrients
5. Effect of COVID-19 on agricultural production and food security: A scientometric analysis
6. Impact of COVID-19 on people's livelihoods, their health and our food systems
7. Adoption of Precision Agriculture Technologies
8. Reddy DJ, Rudra Kumar M (2021) Crop yield prediction using machine learning algorithm. In: 2021 5th international conference on intelligent computing and control systems (ICICCS). IEEE, pp 1466–1470
9. Gopal PSM, Bhargavi R (2019) A novel approach for efficient crop yield prediction. *Comput Electron Agric* 165:104968
10. Singh G, Sharma D, Goap A, Sehgal S, Shukla AK, Kumar S (2019) Machine learning based soil moisture prediction for Internet of Things based smart irrigation system. In: 2019 5th international conference on signal processing, computing and control (ISPCC). IEEE, pp 175–180
11. Suji Prasad SJ, Thangatamilan M, Suresh M, Panchal H, Rajan CA, Sagana C, Gunapriya B, Sharma A, Panchal T, Sadasivuni KK (2022) An efficient LoRa-based smart agriculture management and monitoring system using wireless sensor networks. *Int J Ambient Energy* 43(1):5447–5450
12. Chaopisit I, Kaewpraek N (2020) Low-cost remote-sensing-node based on LoRa mesh network for smart agriculture. *Int J Sci Innov Technol* 3(2):35–42
13. Queralta JP, Gia TN, Zou Z, Tenhunen H, Westerlund T (2019) Comparative study of LPWAN technologies on unlicensed bands for M2M communication in the IoT: beyond LoRa and LoRaWAN. *Proc Comput Sci* 155:343–350
14. Bhoi A, Nayak RP, Bhoi SK, Sethi S, Panda SK, Sahoo KS, Nayyar A (2021) IoT-IIRS: Internet of Things based intelligent-irrigation recommendation system using machine learning approach for efficient water usage. *Peer J Comput Sci* 7:e578
15. Kanuru L, Tyagi AK, Aswathy SU, Fernandez TF, Sreenath N, Mishra S (2021) Prediction of pesticides and fertilizers using machine learning and Internet of Things. In: 2021 international conference on computer communication and informatics (ICCCI). IEEE, pp 1–6

Low Power High Speed Modified GDI Full Adder Implemented by NCFET Technologies



Priyanka Kattepogu and Atul S. M. Tripathi

Abstract This study introduces a novel full adder architecture that combines excellent performance with low power consumption. The proposed design utilizes the Gate Diffusion Input (GDI) and Modified Gate Diffusion Input (MGDI) approaches, which are known for their high speed and low power consumption properties. Additionally, Negative Capacitance Field Effect Transistor (NCFET) technology is used to implement the proposed full adders, which has been shown to improve digital circuit efficiency. The simulation results demonstrate that the proposed full adders significantly reduce power consumption and latency while maintaining a comparable area to existing designs. Overall comparison is made in terms of time delay, power consumption, and number of transistors used. The power consumption is approximately reduced to 90% and delay is approximately reduced to 40%, making an effective and efficient full adder logic design.

Keywords GDI · MGDI · NCFET

1 Introduction

The full adder circuit is a critical component of digital systems used for arithmetic operations such as addition and subtraction. It takes two binary digits and a carry-in bit as inputs and generates a sum bit and a carry-out bit using logic gates like XOR, AND, and OR gates. By connecting multiple full adders, wider adders can be created to handle larger numbers. The circuit can handle up to eight inputs, allowing it to process a byte of data. Full adder circuits are designed for high speed, energy efficiency, low power consumption, and fewer transistors. By optimizing the design and implementing advanced techniques, full adder circuits can achieve a balance

P. Kattepogu · A. S. M. Tripathi (✉)
School of Electronics Engineering, VIT-AP University, Amaravati, India
e-mail: atul.tripathi@vitap.ac.in; atul.tripathi.ece10@itbhu.ac.in

P. Kattepogu
e-mail: priyanka.21mvd7007@vitapstudent.ac.in; priyankakattepogu37946@gmail.com

between these factors, leading to improved performance, reduced power consumption, and enhanced energy efficiency. These advancements contribute to the progress and sustainability of digital technology, enabling the realization of innovative and power-efficient applications across various domains.

2 Literature Review

There are various designs for full adders with different transistor counts, power consumption levels, and performance metrics [1]. Pass-transistor logic-based designs with fewer transistors are popular for reducing power consumption, but they may result in degraded output signals and are not suitable for low voltage operations. Designs such as Carry Lookahead, Ripple Carry, Manchester Carry Chain, and Hybrid Adder address specific factors such as power consumption, latency, and space efficiency. The 28-transistor full adder provides reliable operation but may consume more power and have a larger footprint. The 20 T, 16 T, and 14 T designs use transmission gate logic, pass-transistor logic, and a combination of pass-transistor and transmission gate logic, respectively, to reduce the transistor count while maintaining functionality. The 12 T design uses complementary pass-transistor logic to achieve a further reduction in transistor count while minimizing area and power consumption [1].

3 Proposed Approaches

3.1 Design a GDI-Based Full Adder with 10 Transistors

In 2001, A. Morgenshtein, A. Fish, and I. A. Wagner developed the GDI approach. It entails utilizing the straightforward cell in Fig. 1 and is a design methodology that is known for its outstanding power efficiency. By using GDI, designers can improve logic swing and minimize static power dissipation, enabling the implementation of various logic functions with fewer transistors. This optimizes resource utilization.

The GDI method is a digital circuit design technique that uses a simple cell structure to reduce transistor count, power consumption, chip area, and delay. It can implement multiple logic gates using a single cell and is useful for designing circuits that require low power consumption, fast operation, small chip area, and reduced transistor count. However, it may introduce challenges such as increased sensitivity to process variations and reduced noise margins. The GDI cell has three inputs and can perform various logic functions, including a complete adder that adds two binary numbers and a carry input to produce sum and carry outputs [2, 3]. The GDI technique can design a full adder using only 10 transistors, leading to improved performance (Table 1).

Fig. 1 GDI cell

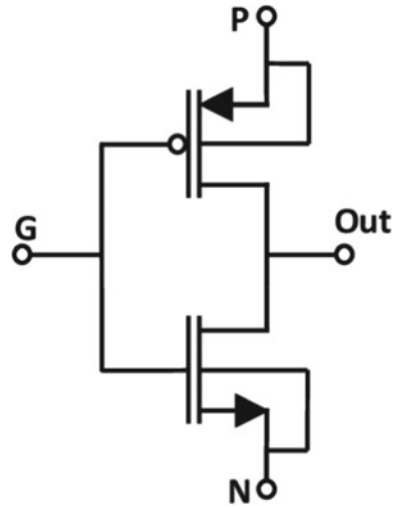


Table 1 Various functions of GDI cell

G	P	N	D	Function
A	B	0	$A'B$	F1
A	1	B	$A' + B$	F2
A	B	1	$A + B$	OR
A	0	B	AB	AND
A	B	C	$A'B + AC$	MUX
A	1	0	A'	NOT

A 10-transistor complete adder’s schematic based on the figure displays the GDI method [4]. The design is built upon the four-transistor XOR logic concept, with two such gates used to carry out the addition and obtain the sum. The carry output is generated using pass transistors, as depicted in the figure (Fig. 2).

A new 10-transistor full adder based on GDI was designed using Cadence Virtuoso and GPDK 90. The design features fewer transistors, which reduces power consumption and chip area, leading to cost savings and simplified design. The switching activity of the new full adder is also lower compared to previous designs. The full adder was tested with various pulse voltages applied to input signals, and the outputs for sum and carry were observed and compared to earlier full adder designs. Latency and power usage calculations were made, and the GDI-based full adder achieved lower power usage and transistor count (Fig. 3).

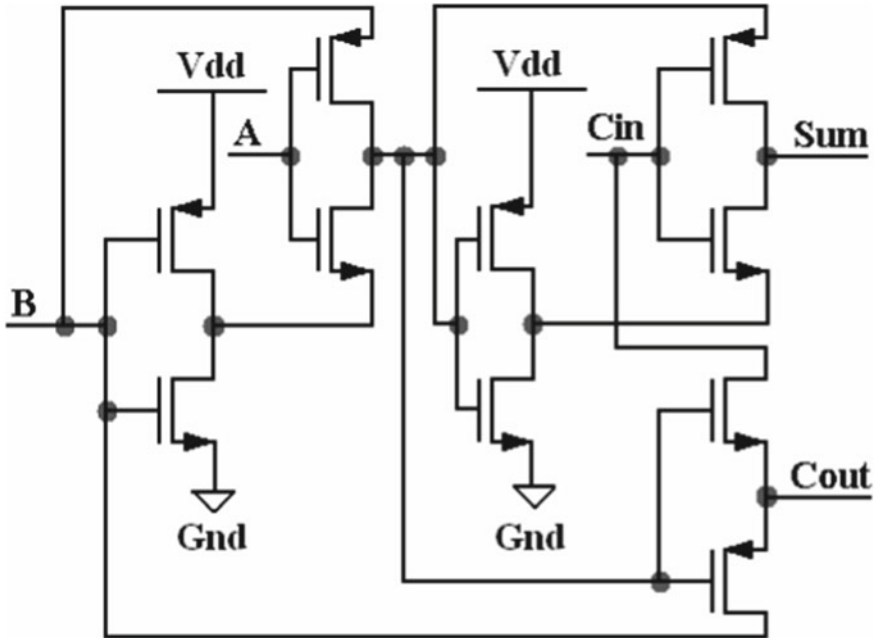


Fig. 2 GDI-based 10 T full adder

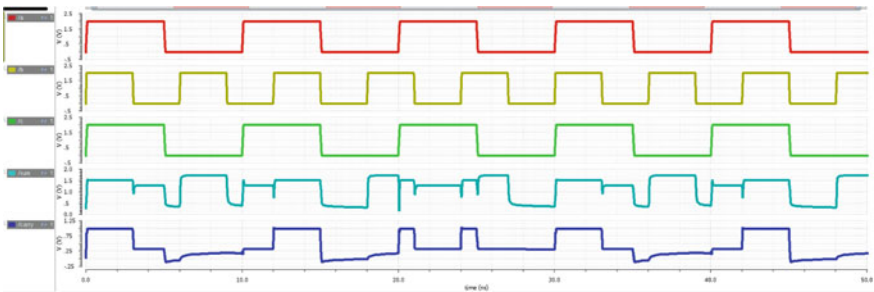


Fig. 3 Transient response of GDI-based full adder with 10 transistors

3.2 Drawbacks of GDI Technique

- The primary drawback of utilizing a GDI cell is [5] that it may result in compromised output quality. When G is set to 0, the PMOS transistor is activated, allowing the P input to pass to the output. Similarly, when G is set to 1, the NMOS transistor is activated, passing the N input to the output. However, in such cases, a weak logic 0 or a weak logic 1 may be produced by the PMOS and NMOS transistors, respectively. It is therefore necessary must alter the GDI cell to produce a strong logic 0 that may be used as an acceptable input for the following stage.

- The output stage of GDI does not provide a strong logic high (1) or logic low (0) signal.
- This is due to inadequate biasing of the bulk end, resulting in a drop in the threshold voltage of the circuit.

3.3 Design a Modified GDI Technique Full Adder with 8 Transistors

The conventional GDI technology has been improved with the MGDI method used in digital circuit design. MGDI aims to address the issue of degraded output signal quality in GDI cells, specifically the weakened logic levels produced by the NMOS and PMOS transistors. In MGDI, additional transistors and circuit modifications are introduced to improve the output characteristics of the GDI cell. The modifications are designed to overcome the limitations of the standard GDI technique and enable the production of stronger logic levels at the output (Fig. 4).

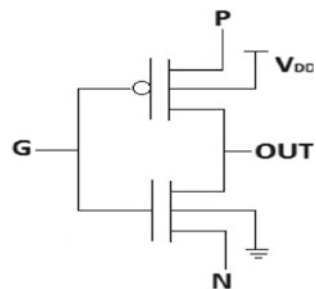
In a modified GDI cell [2, 6] the supply voltage, also called as V_{dd} , is linked to the bulk node of the PMOS transistor. The NMOS transistor’s bulk node is linked to GND, a low constant voltage, in the redesigned GDI cell.

3.3.1 Implementation of Full Adder Using Modified GDI Technique

The XOR gate design is the cornerstone upon which the entire adder’s design is built. This suggested architecture makes use of three transistor XOR gates [7] (Fig. 5).

To improve and minimize static power consumption, a new implementation XOR logic has been created with the use of MGDI cell. Only 8 transistors (8 T) are required for this MGDI complete adder design, resulting in reduced complexity and physical size. Two XOR logic gates and one Multiplexer (MUX) logic unit make up the 1-bit MGDI complete adder [3]. Three transistors are normally used to build XOR gates when utilizing the conventional method. However, utilizing two p-type Metal–Oxide–Semiconductor (pMOS) units and one n-type Metal–Oxide–Semiconductor

Fig. 4 Modified GDI cell



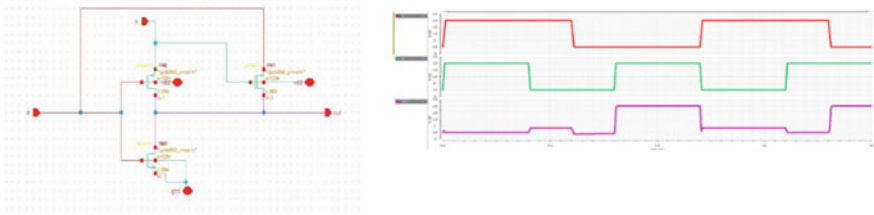


Fig. 5 XOR logic is implemented by using MGDI technique of 3 transistors and its transient response

(nMOS) unit, the modified GDI approach may produce a three transistor XOR gate. Output of modified GDI-based full adder is as follows (Figs. 6 and 7).

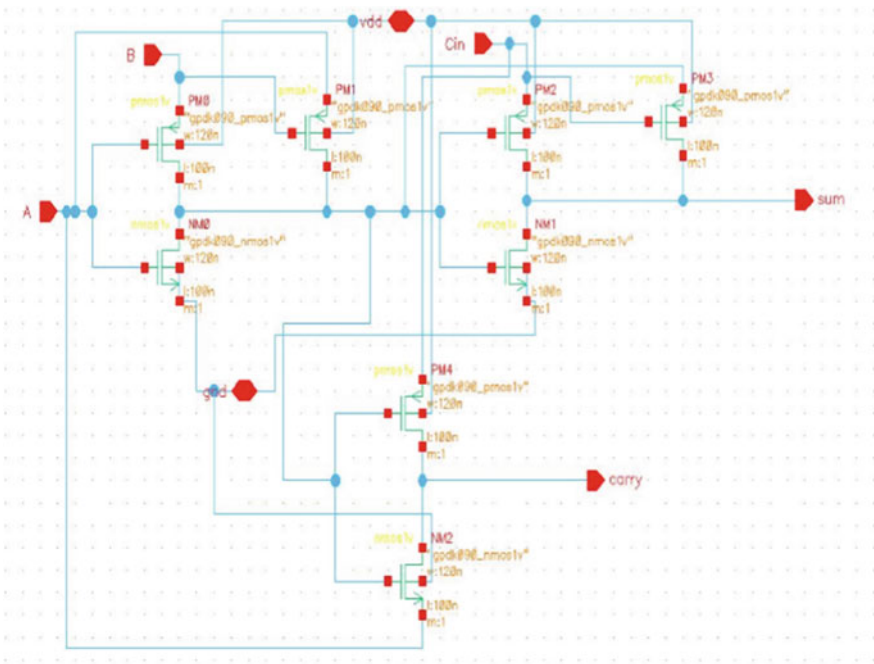


Fig. 6 Schematic of modified GDI-based full adder with 8 transistors

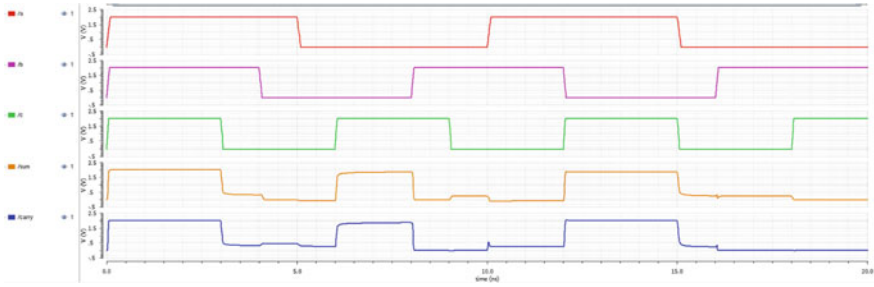


Fig. 7 Transient response of modified GDI (MGDI)-based full adder with 8 transistors

3.4 Utilizing NCFET Technology for Designing Full Adders

3.4.1 Introduction of MOS Devices

The heat generated by billions of transistors on a single IC chip is a significant challenge for the semiconductor industry. To address this issue, reducing the supply voltage is a promising technique for reducing power consumption. However, in conventional MOSFETs, scaling the supply voltage has reached its fundamental limits, with a subthreshold swing (SS) of 60 mV/decade being the threshold. To achieve lower supply voltages, FETs with SS values below 60 mV/decade, such as tunnel FETs and NCFETs, are being explored. These FETs have potential for low-power operation, leading to reduced power dissipation and lower off-current. FETs are classified into two groups based on their subthreshold slope: MOSFETs, FINFETs, FDSOI, and similar devices with SS greater than 60 mV/decade, and tunnel FETs and NCFETs with SS less than 60 mV/decade [8].

3.4.2 Tunnel FET Device

TFETs have the potential to offer advantages over conventional transistors, such as lower subthreshold swing and leakage current, leading to lower power consumption and improved energy efficiency. However, TFETs face challenges such as achieving high on-current levels and integrating them into existing fabrication processes. Researchers are working on developing new tunnel junction designs and material systems to improve TFET performance and optimizing TFET design and fabrication processes to achieve practical and reliable devices for various applications. Despite these challenges, TFETs are a promising technology that could significantly improve the performance of electronic devices.

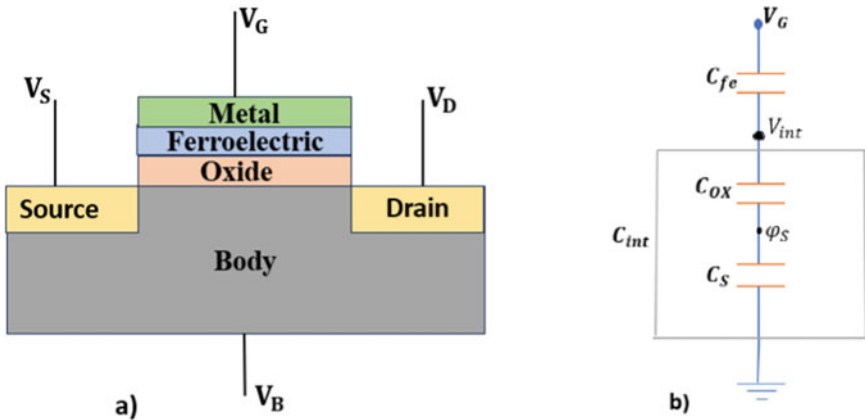


Fig. 8 a Structure of NCFET. b Capacitance model of NCFET

3.4.3 Concept of NCFET

NCFETs leverage the unique properties of certain ferroelectric materials to achieve enhanced performance. Ferroelectric materials exhibit a characteristic called the ferroelectric phase transition, where the material undergoes a reversible polarization shift with an electric field. By integrating a ferroelectric substance into the transistor’s gate oxide, NCFETs can effectively reduce the effective capacitance of the transistor, making it appear as if it has a negative capacitance. This phenomenon can improve transistor performance by reducing power consumption and increasing switching speeds. The structure and capacitance model of NCFETs [8] are designed to take advantage of this unique behavior of ferroelectric materials (Fig. 8).

A. The GDI-based full adder is implemented by using NCFET technology with 10 transistors

The GDI-based 10-transistor complete adder is a digital circuit that performs binary arithmetic addition operations. It uses the GDI technique and negative capacitance concept to reduce transistor count and power consumption while improving performance. The circuit is divided into a sum stage, a carry stage, and a final stage. Negative capacitance compensates for parasitic capacitance, leading to faster switching times, improved noise margins, and reduced power consumption. Additionally, XOR gates can be designed using only four transistors, further reducing transistor count and power consumption. The circuit is implemented using NCFET technology, which has a low subthreshold slope and negative capacitance, reducing the number of transistors needed for full adders. The full adder circuit takes two inputs (A and B) and a carry-in signal (C_{in}) and produces two outputs (sum and carry). The expressions to obtain these outputs are given as follows (Figs. 9 and 10):

Sum = $\mathbf{a \ XOR \ b \ XOR \ C_{in}}$ and Carry = $\mathbf{a \cdot b + b \cdot C_{in} + C_{in} \cdot a}$, it can be written as
 Carry = $\mathbf{(a \ XOR \ b) \cdot C_{in} + a \cdot b}$

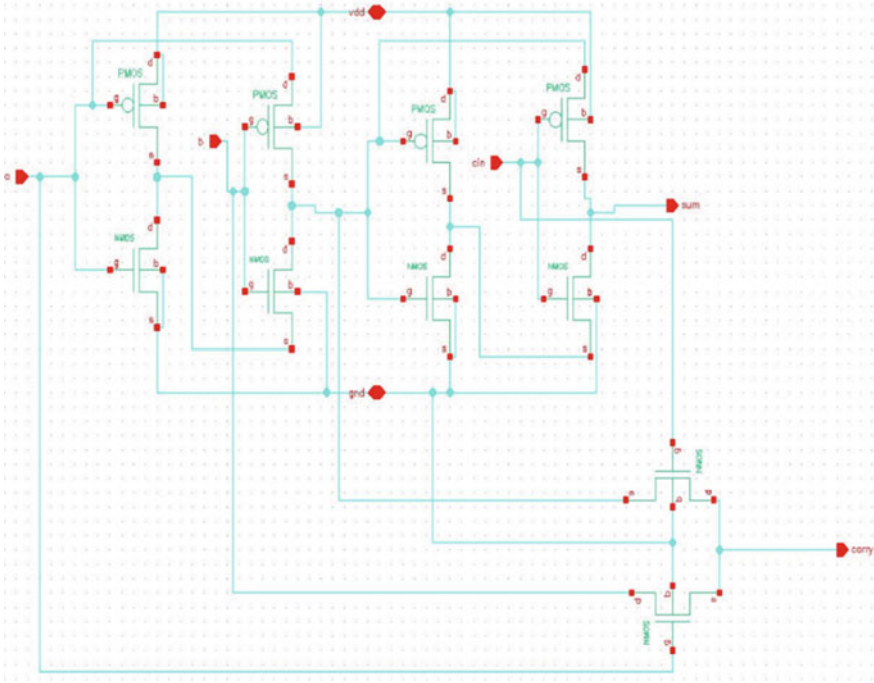


Fig. 9 Schematic of 10 transistor full adder with GDI technique in NCFET technology

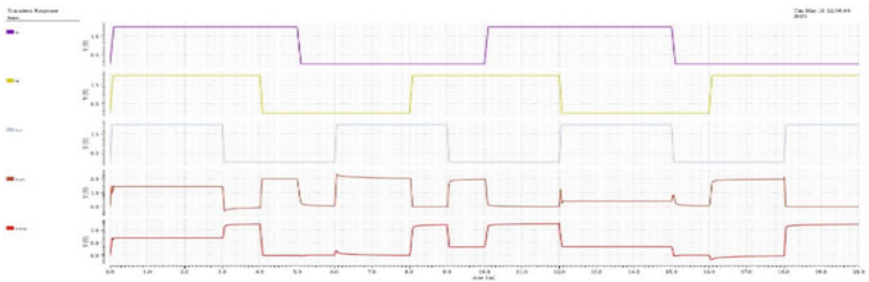


Fig. 10 Output response of GDI full adder in NCFET technology

The figure displays the result of a full adder that is GDI-based that has been implemented using NCFET technology. Upon observation, the output waveform is degraded. For example, when the input changes from 100 to 000, the expected behavior is for the sum output to transition from 1 to 0. However, in the figure shown, when the sum is 0 for inputs A, B, and C_{in} being 0, 0, 0 respectively, the output exhibits a small voltage, as illustrated in the figure. This small for the next stage, voltage can be mistakenly perceived as logic 1 leading to an incorrect output.

B. Modified GDI-based full adder of 8 transistors is implemented by using NCFET technology

Designing a full adder circuit requires consideration of factors such as high speed, energy efficiency, low transistor count, and low power consumption. A full adder circuit enables the execution of advanced arithmetic functions in digital circuits. The GDI approach was used to design a full adder with 10 transistors, but output voltage degradation and high threshold drop were observed. Modifications were made to the gate diffusion input technique by biasing the bodies of both PMOS and NMOS transistors to achieve strong logic 1 and 0 outputs. Currently, an 8-transistor MGDI approach is being developed in negative capacitance technology. This approach is expected to have a lower transistor count, reduced power consumption, and a smaller chip size. By improving transistor count and power consumption, the full adder circuit can achieve improved performance, cost-effectiveness, and environmental sustainability (Fig. 11).

The above schematic is designed with 8 transistors with the help of modified GDI technique concept in NCFET technology to avoid the weak logic 0 and weak logic

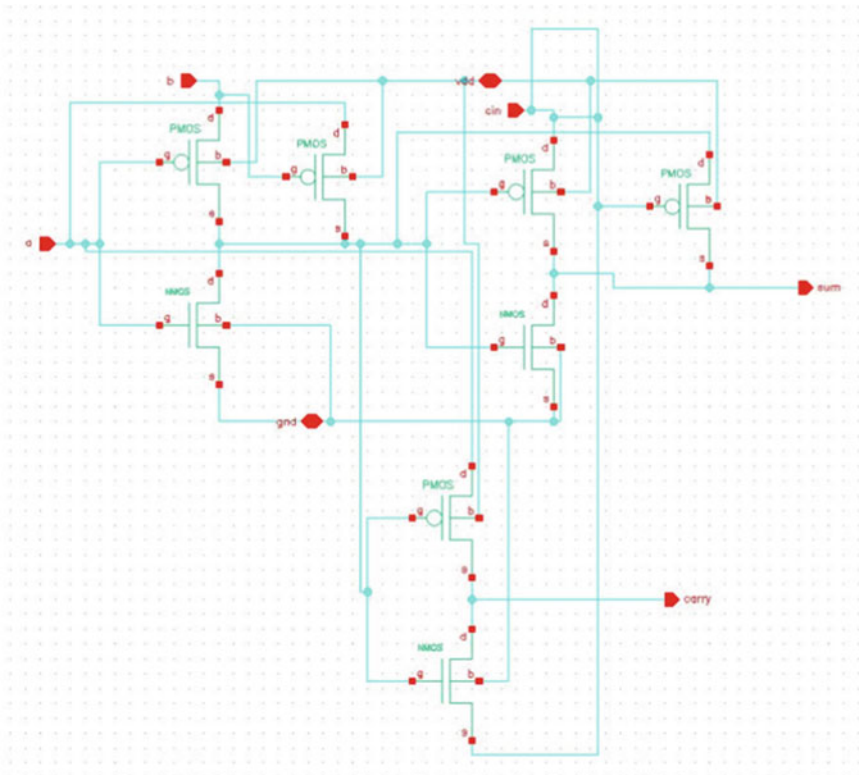


Fig. 11 Schematic of modified GDI-based full adder in NCFET technology

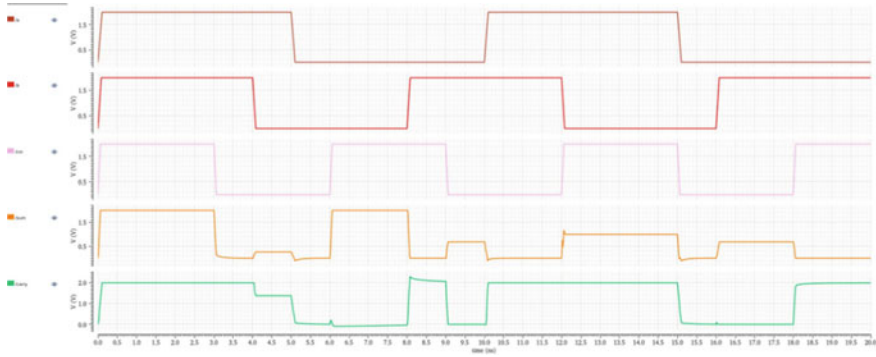


Fig. 12 Transient response of modified GDI-based full adder in NCFET technology

land also avoid the output voltage swing. The symbol test for the GDI technique of 8 transistors full adder in NCFET technology is shown in Fig. 12.

3.5 Advantages of NCFET Technology

NCFETs offer several advantages in full adder circuits. The negative capacitance effect compensates for the capacitive loading of the circuit, improving propagation time and reducing delay. NCFETs can also boost effective transconductance, reducing power dissipation. They offer improved noise immunity and robustness against fluctuations in input voltage. NCFETs enable compact circuit designs, reducing the number of transistors and silicon area, contributing to lower manufacturing costs and increased integration density. NCFETs can be implemented using conventional CMOS processes, allowing for easy integration into existing electronic systems without significant changes to the manufacturing process.

4 Results and Discussion

The effectiveness and efficiency of a full adder logic design can be evaluated based on its worst-case delay, power consumption, and power delay product. In this regard, the proposed design, which incorporates the Gate Diffusion Input (GDI) and Modified Gate Diffusion Input (MGDI) methods, has demonstrated promising results. Specifically, it has achieved a reduction of approximately 90% in power consumption and 40% in delay compared to a traditional full adder design that uses CMOS technology or NCFET technology. Table 2 provides a comparative analysis of the performance parameters of these different designs.

Table 2 Delay versus Avg power consumption between CMOS and NCFET technologies

Parameter	CMOS technology		NCFET technology	
	GDI	Modified GDI	GDI	Modified GDI
Transistor count	10	8	10	8
Average power (μ W)	59.09	18.54	6.768	0.127
Delay sum (ns)	5.03	4.06	3.001	3.006
Delay carry (ns)	5.026	4.05	4.017	2.988
Power delay product (fj)	297.2	75.3	20.31	0.38

5 Conclusion

The paper's consuls have presented a novel design of full adders that achieves excellent performance and minimal power usage. The proposed design utilizes both the MGDI and the GDI techniques, which are known for their high-speed capability and little power consumption. In addition, the proposed full adders are implemented using NCFET technology, which has been shown to enhance the performance of digital circuits. The results of the simulation show that the suggested full adders significantly reduce power consumption and delay. While maintaining comparable area. The proposed full adder is compared with existing Hybrid CMOS 1-bit full adders. Based on the simulation results overall improvements of 40%, 90%, and 92% are observed in reduction of delay, power consumption, and power delay product, respectively. The proposed full adders could be used in a variety of digital systems that call for high speed and low power consumption such as mobile computing, Internet of Things, and wearable devices. The proposed design can be further optimized by exploring additional circuit techniques and optimization strategies.

References

1. Joshi M, Chauhan D, Kaushik BK (2015) A new high-speed full adder cell. *World Appl Sci J* 33:599–603
2. Dhar K (2014) Design of a low power, high speed, energy efficient full adder using modified GDI and MVT scheme in 45nm technology. In: 2014 international conference on control, instrumentation, communication and computational technologies, pp 1–5
3. Kai W, Ahmad N, Jabbar M (2019) Design of low power 8-bit gate-diffusion input (GDI) full adder using variable body bias (VBB) technique in 90nm technology. *Indonesian J Electr Eng Comput Sci* 14(2):912–920. <https://doi.org/10.11591/ijeecs.v14.i2.pp912-920>
4. Pandey SP, Gupta R, Akashe S, Vidyarthi A (2013) Analysis and optimization of active power and delay of 10T full adder using power gating technique at 45 nm technology. *J VLSI Sig Proc* 2:51–57
5. Hiremath S, Koppad D (2012) Low power full adder circuit using Gate Diffusion Input (GDI) MUX. In: Proceedings of the 2012 conference, pp 53–56. <https://doi.org/10.1049/cp.2012.2493>
6. Uma R, Dhavachelvan P (2012) Modified gate diffusion input technique: a new technique for enhancing performance in full adder circuits. *Procedia Technol* 6:74–81

7. Roy Chowdhury S, Banerjee A, Roy A, Saha H (2008) A high-speed 8-transistor Full Adder design using novel 3-transistor XOR gates. *Int J Electron Circ Syst* 3(1):1–6
8. Rahi SB, Tayal S, Upadhyay A (2021) A review on emerging negative capacitance field effect transistor for low power electronics. *Microelectron J* 116(105242)

Design and Analysis of Non-vulnerable PRBS Generation with Internal Shift and XOR of LFSR



Jagadeshwaran Ramasamy  and Dhandapani Samiappan 

Abstract Linear feedback shift register (LFSR) generates the pseudorandom binary sequence (PRBS) used in the 3G wireless standards code division multiple access (CDMA). LFSR has inherent properties and widely applied in digital random number generator, encoding techniques, built-in self-test (BIST) and cryptography. PRBS sequence length becomes vital and depends on the hardware complexity, namely registers. In this, PRBS generation with internal shift performed with XOR on the conventional LFSR increases the PRBS sequence length with reduced hardware complexity. Results are compared with conventional techniques to reveal the properties of the proposed method on par with the conventional techniques with double the sequence length while retaining same hardware complexity and reduced switching activity. With this, future wireless communication increases the number of codes, and hence, the number of users accommodation increases drastically with low intentional interference.

Keywords LFSR · CDMA · PRBS · BIST

1 Introduction

Linear feedback shift register (LFSR) has numerous applications in the field of wireless communication, fault coverage, device under test, pattern generation and data compression. Emergence is essential in every field as the world's loop corner is digitizing very fast. In the Third Generation Partnership Project, CDMA plays vital role in handling subscribers with unique code at end user communication with increased data rates. CDMA employs special sequences as codes and has added features and robustness against interference. Increasing demand for reducing interference needs

J. Ramasamy (✉) · D. Samiappan
Saveetha Engineering College, Saveetha Nagar, Chennai, India
e-mail: Jagadeshwaran231@gmail.com

D. Samiappan
e-mail: dhandapani@saveetha.ac.in

orthogonal sequences for spreading the signal over the same channel and to increase the subscribers with in the channel. Various improvements are carried our recently for the generation of PRBS using LFSR.

Interference is the main issue in wireless communication. Interference was effectively handled by CDMA method. Every subscribers's information is transmitted by encoded with specially generated orthogonal codes and decoded with same code at the other end. To enhance the security, the multibit LFSR technique was proposed [1]. In this, instead of single bit shift at each clock cycle, multiple limited bits are shifted in a clock time. Multibit LFSR performs well in 32-bit LFSR generation, with complexity in how much bits shifted in a clock cycle. Multibit LFSR effectively used in lightweight applications of message hidden in image for steganography applications was discussed [2].

LFSR reseeding scheme was developed to reduce the power dissipation [3], by effectively handling the switching transition. Reducing transition in conjunction with generation of specified number of bits in the sequence length provides compatibility of compression techniques. LFSR was used in reducing the bit position errors with reduced decoding latency utilizing the autocorrelation among the different PRBS codes [4]. This method revamped the failures of code synchronoization with improved the success rate of detection. Ordering of orthogonal array employed using LFSR was discussed [5]. LFSR-based pattern generation for path delay fault was proposed to detect long-path undetectable delay [6]. It matches the similarity between the detected and target paths to compute the seed value for PRBS.

Padding of LFSR seed based on decompression logic [7] was proposed to increase the effectiveness of detecting the target faults with compact test data volume. EMI suppression based on PRBS and data inclusive with XOR to reduce radiation was proposed [8]. In order to satisfy the emergence of edge computing devices, LFSR-based pruning sparsity technique was introduced [9] with reducing area of the devise and saving energy and power.

Observation based on the modified output sequence based on swapping the MSB with 2×1 multiplexer was proposed to reduce the peak and average power of the conventional LFSR [10]. It reduces the bit transition slightly and avails 25% smooth shift in the run-length in comparison with conventional LFSR. To generate the PRBS sequence of flexible length, vertical stacked LFSR [11] was proposed by stacking the highest LFSR over the lower LFSR. Combining internal and external LFSR surge the PRBS sequence length as integrating two sets of LFSR [12] was introduced with little hardware complexity. Modified PRBS sequence that satisfies properties of the main stream conventional LFSR in order to be compatible with the LFSR application was discussed [13]. The characteristics provide similarity among the conventional and proposed codes, and it plays vital role in wireless communication in reducing interference.

The paper is presented as follows. Section 2 deals the background of the proposed method. Section 3 introduces the block diagram and algorithm of the proposed method. Section 4 discusses about the result attained. Finally, section concludes.

Table 1 Primitive polynomial

Bits (n)	Feedback polynomial (F_n)	Period T_n
2	$a^2 + a + 1$	$2^n - 1$
3	$a^3 + a^2 + 1$	$2^n - 1$
4	$a^4 + a^3 + 1$	$2^n - 1$
5	$a^5 + a^3 + 1$	$2^n - 1$
6	$a^6 + a^5 + 1$	$2^n - 1$
7	$a^7 + a^6 + 1$	$2^n - 1$
	\vdots	

2 Background

Conventional LFSR is the simplest PRBS generation with minimal hardware complexity. Technique to improve the PRBS sequence length was discussed. VS-LFSR technique is employed to increase the PRBS length based on the need for number of codes/subscribers. Two techniques are discussed in this section with its block diagram.

2.1 LFSR

Linear feedback shift register (LFSR) is a linear combination of digital logic connected by the sequence of registers with the seed value and regenerative feedback to the registers in order of n -bit primitive polynomial for the sequence of n . Feedback pattern is essential in the logic combination to generate the PRBS of longer length with the seed value as input. The primitive polynomial is the minimal polynomial seed value to LFSR registers to generate the PRBS of specific length. PRBS generated has longer length patterns, namely maximum length sequence (MLS). An n -bit LFSR generates MLS in the range of $2^n - 1$.

Table 1 shows primitive polynomial for PRBS generation of maximum length sequence of $2^n - 1$ and with the same time period. Figure 1 picturizes the 4-bit conventional LFSR with the series of registers and linear combination of the primitive polynomial as input (i_s). The block diagram extends for n -bit LFSR with the corresponding primitive polynomial as feedback.

2.2 VS-LFSR

Vertical stacked LFSR is developed to arbitrarily change the PRBS sequence length in multiples of maximum length period of each n -bit of order smaller than the overall bit length. The method increases the PRBS sequence length while trade-off with number

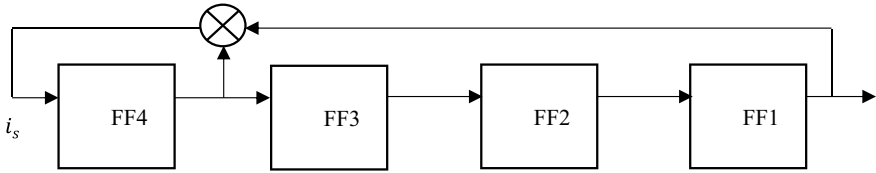


Fig. 1. 4-bit conventional LFSR

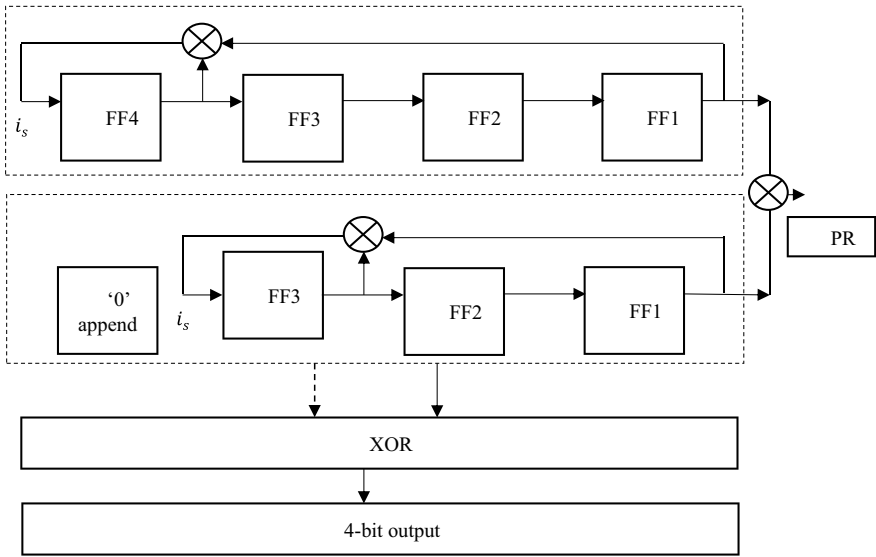


Fig. 2 Vertical stacked LFSR

of registers. VS-LFSR adheres to the properties of conventional LFSR, except the method generates 2^n states. Figure 2 idealized the concept of VS-LFSR.

3 PRBS Generation with Shift and XOR of LFSR

The proposed method is shown in Fig. 3. The n -bit conventional LFSR is the base for the proposed method. The state output at each stage is left shifted by value one and padded with zero to equalize the length of the conventional LFSR. The equivalence operation, bitwise XOR is carried out on conventional LFSR and left shifted LFSR sequence to generate the output bit sequence. In order to double the sequence length, the sequence output is taken at two levels of clock. At leading edge, the conventional LFSR outputs as PRBS sequence. At the trailing edge, the shifted version output

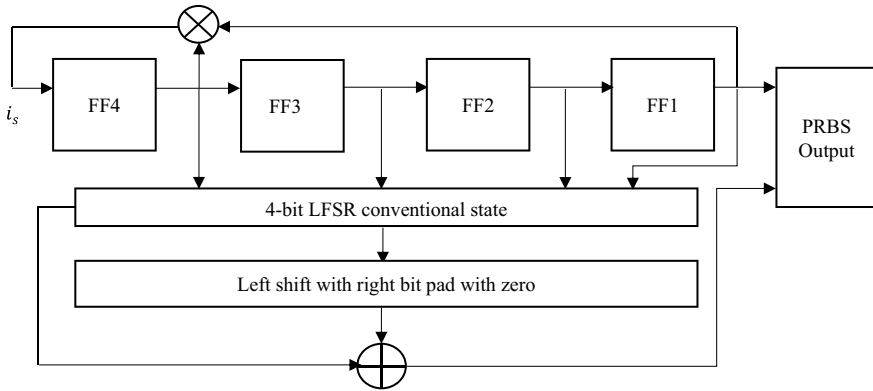


Fig. 3 PRBS generation with left shift—XOR of proposed method

Table 2 Algorithm

PRBS generation with left shift—XOR
Initialize the seed value of LFSR
Define parameters: clock, No. of registers, bit length = i
Evaluate the edge of the clock
While edge is leading
$CS = NS;$
$Input = XOR(CS(i), CS(i + 1))$
$NS = input \& CS(1:i - 1);$
$PRBS = NS(i)$
While edge is trailing
$CS_N = NS(2:i) \& '0'$
$NS_N = XOR(NS, CS_N)$
$PRBS = NS_N(i)$

is considered. Altogether the bit sequence was doubled with the same hardware complexity. Table 2 shows the algorithm of the proposed method.

4 Result and Discussion

The proposed method accommodates double the subscriber in the CDMA with the generation of double the PRBS than the conventional techniques. Table 3 shows the comparison of the proposed and existing methods of PRBS generation. The proposed method has maintained the orthogonal properties of the sequences and reduced the switching activity; hence, transaction power dissipation is reduced. The simulation was carried in the Xilinx 14.7 on the target device xc6slx4 with the sequence generated with the clock frequency of 1 GHz. Conventional LFSR generates

the signal for the time period of $(2^n - 1)ns$. The proposed method generates the signal for the time period of $2(2^n - 1)ns$.

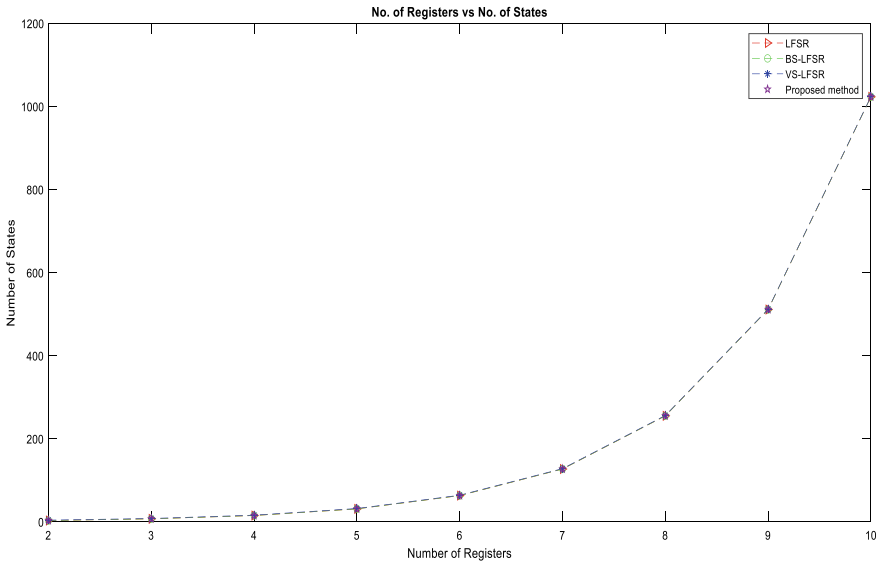
Figure 4a shows the comparison between the number of states and number of registers employed in the conventional and proposed techniques. Comparatively, conventional LFSR, BS-LFSR and proposed method produce same number of output states, whereas VS-LFSR has all zero states included in the domain. On comparison, VS-LFSR has generated one additional state than the conventional and proposed method. Hence, VS-LFSR covers all states for n-bit, no struck-at faults may occur in the state diagram. Figure 4b shows the comparison between the number of input bits and PRBS length. The proposed method outperforms in the generation of the sequence length by double the times, as output carried at each edge of the clock signal.

Figure 5 shows overall transition or run-length of the PRBS of the proposed method for 4-bit LFSR in the conventional method. Figure shows run-length, considering each place value as input for 4-bit length. As per LFSR property, run-length is ceil function of half of the total number of states maintained in each case. Hence, the property of LFSR is preserved. Figure 5a shows the number of transitions occurred as PRBS generated for MSB. Figure 5b shows the number of transitions occurred as PRBS generated for next lower of MSB. Figure 5c shows the number of transitions occurred as PRBS generated for next higher of LSB. Figure 5d shows the number of transitions occurred as PRBS generated for LSB.

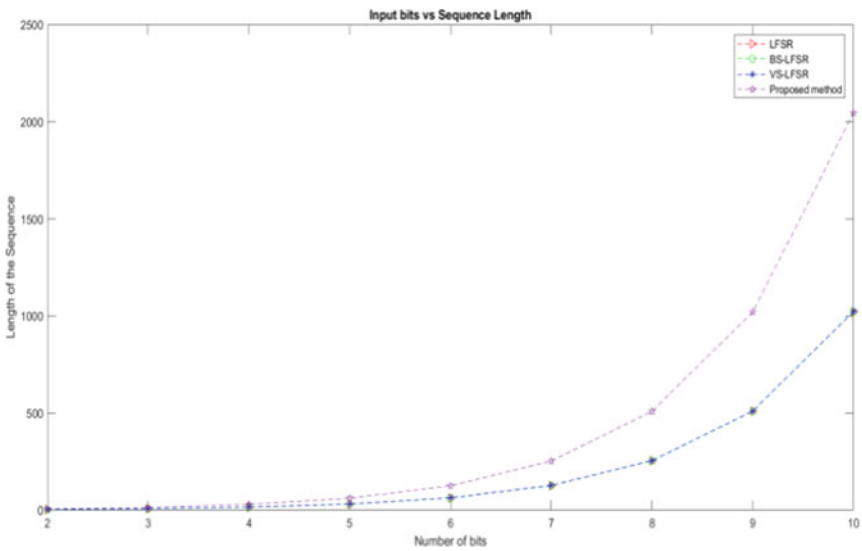
Figure 6a shows PRBS sequence run-length for various positions of the states for $n = 4$ in the proposed method, when LSB value is output. On comparing the conventional and proposed methods for 4-bit LFSR with PRBS from LSB of the registers, the system performance is the same in terms of switching activity with the ratio of number of PRBS length and number of transitions. Figure 6b shows PRBS sequence run-length for various positions of the states for $n = 4$ in proposed method, when LSB value is output. On comparing the conventional and proposed methods for 4-bit LFSR with PRBS from LSB of the registers, the system performance is outperforming in terms of switching activity and power dissipation that is reduced.

Table 3 Comparison of LFSR, BS-LFSR, VS-LFSR and proposed method

Method	Number of bits	Number of states	Number of flip-flops	Hardware complexity	Period	Run-length	PRBS sequence length
LFSR	n	$(2^n - 1)$	n	F_n	$(2^n - 1)$	$\frac{(2^n - 1)}{2}$	$(2^n - 1)$
BS-LFSR	n	$(2^n - 1)$	n	$F_n + \frac{M_n}{2}$	$(2^n - 1)$	$\frac{(2^n - 1)}{2}$	$(2^n - 1)$
VS-LFSR	n	2^n	$\frac{n^2 + n - 2}{2}$	$\sum_{k=0}^{n-2} F_{n-k}$	$\prod_{m=0}^{n-2} (2^{n-m} - 1)$	$\frac{\prod_{m=0}^{n-2} 2^{n-m} - 1}{2}$	$\prod_{m=0}^{n-2} (2^{n-m} - 1)$
Proposed method	n	$(2^n - 1)$	n	F_n	$2(2^n - 1)$	$2^n - 1$	$2(2^n - 1)$



(a)



(b)

Fig. 4 a Number of states versus registers, b number of bits versus sequence length

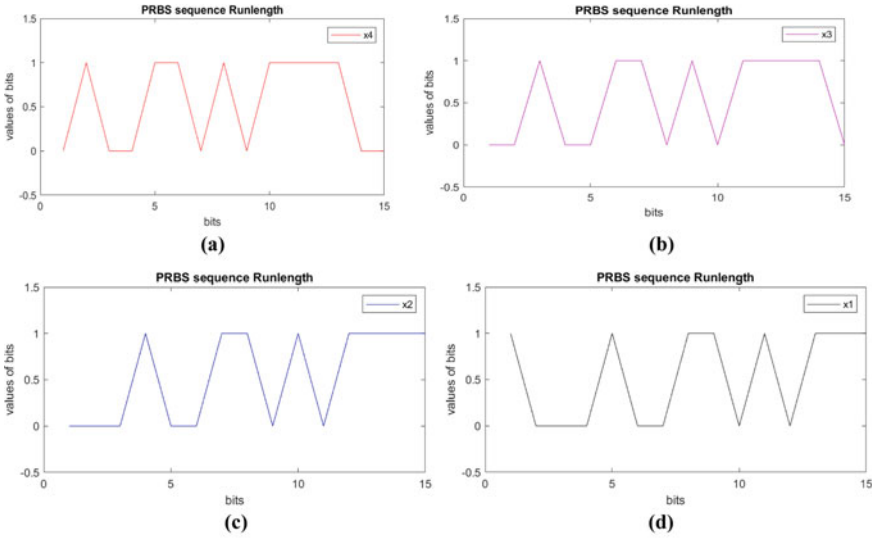


Fig. 5 PRBS sequence run-length for various positions of the states for $n = 4$ in conventional method **a** MSB, **b** MSB-1, **c** LSB + 1, **d** LSB

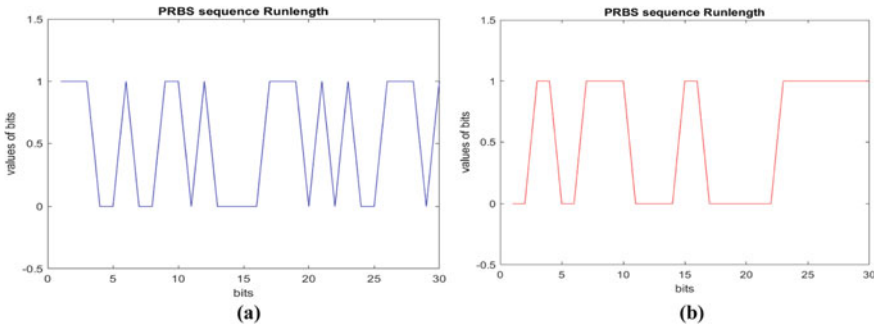


Fig. 6 PRBS sequence run-length for various positions of the states for $n = 4$ in proposed method, **a** LSB value is output, **b** MSB value is output

Table 3 shows the comparison between conventional and proposed methods with the parameter of n -bit input. Hardware complexity of the proposed method is maintained on par with the conventional LFSR with increase in the PRBS length. Autocorrelation between the generated sequences is carried out and adheres to the property of LFSR.

5 Conclusion

In this paper, the proposed method of shifted XOR of LFSR was designed and studied. It was outperformed with existing technologies in the generation of PRBS sequence length. In comparison with BS-LFSR, switching activity was reduced which leads to reduction in power dissipation without the multiplexer circuit. Increasing the number of codes concludes, increase in the number of users in CDMA is twice than the conventional LFSR method with same complexity. The technique increases the complexity to decode the cipher code as code doubles. In the future, the method applied to the spread spectrum analyzes the error performance and its compatibility.

References

1. Datta D, Datta B, Dutta HS (2017) Design and implementation of multibit LFSR on FPGA to generate pseudorandom sequence number. In: 2017 devices for integrated circuit (DevIC), Kalyani, India, pp 346–349. <https://doi.org/10.1109/DEVIC.2017.8073966>
2. Shunnar M, Othman A, Awad A (2022) A study to improve image steganography using linear feedback shift register. In: 2022 international conference on computer engineering, network, and intelligent multimedia, Surabaya, Indonesia, pp 344–348. <https://doi.org/10.1109/CENIM56801.2022.10037395s>
3. Lee J, Toubna NA (2007) LFSR-reseeding scheme achieving low-power dissipation during test. *IEEE Trans Comput Aided Des Integr Circuits Syst* 26(2):396–401. <https://doi.org/10.1109/TCAD.2006.882509>
4. Moon S, Park S, Lee J-H, Lee Y (2019) Rapid Balise telegram decoder with modified LFSR architecture for train protection systems. *IEEE Trans Circuits Syst II Express Briefs* 66(2):272–276. <https://doi.org/10.1109/TCSII.2018.2844731>
5. Castoldi G, Moura L, Panario D, Stevens B (2017) Ordered orthogonal array construction using LFSR sequences. *IEEE Trans Inf Theory* 63(2):1336–1347. <https://doi.org/10.1109/TIT.2016.2634010>
6. Pomeranz (2019) LFSR-based test generation for path delay faults. *IEEE Trans Comput Aided Des Integr Circuits Syst* 38(2):345–353. <https://doi.org/10.1109/TCAD.2018.2812120>
7. Pomeranz (2021) Padding of LFSR seeds for reduced input test data volume. *IEEE Trans Comput Aided Design Integr Circ Syst* 40(5):1004–1008. <https://doi.org/10.1109/TCAD.2020.3011665>
8. Park et al (2019) Polynomial model-based eye diagram estimation methods for LFSR-based bit streams in PRBS test and scrambling. *IEEE Trans Electromagn Compat* 61(6):1867–1875. <https://doi.org/10.1109/TEMC.2019.2900055>
9. Karimzadeh F, Cao N, Crafton B, Romberg J, Raychowdhury A (2021) A hardware-friendly approach towards sparse neural networks based on LFSR-generated pseudo-random sequences. *IEEE Trans Circuits Syst I Regul Pap* 68(2):751–764. <https://doi.org/10.1109/TCSI.2020.3037028>
10. Abu-Issa S, Quigley SF (2009) Bit-swapping LFSR and scan-chain ordering: a novel technique for peak- and average-power reduction in scan-based BIST. *IEEE Trans Comput Aided Des Integr Circuits Syst* 28(5):755–759. <https://doi.org/10.1109/TCAD.2009.2015736>
11. Ramasamy J, Samiappan D (2022) A modified PRBS: vertical stacked LFSR primitive polynomial for secure data communication. *Procedia Comput Sci* 215:947–954. <https://doi.org/10.1016/j.procs.2022.12.097>
12. Vikranth CS, Rakesh K, Jagadeesh B, Mohammad D, Somanathan GR, Bhakthavatchalu R (2021) Design and analysis of test pattern generator by combining internal and external

- LFSR. In: 2021 5th international conference on trends in electronics and informatics (ICOEI), Tirunelveli, India, pp 240–244. <https://doi.org/10.1109/ICOEI51242.2021.9452847>
13. Velazquez-Gutierrez M, Vargas-Rosales C (2017) Sequence sets in wireless communication systems: a survey. *IEEE Commun Surv Tutor* 19(2):1225–1248. <https://doi.org/10.1109/COMST.2016.2639739>

ML- and IoT-Based Crop Prediction System



Kushagra Sharma and Deepak Kumar

Abstract The use of IoT and machine learning technologies to develop intelligent agricultural system is discussed here. This integrated system addresses the three critical aspects of farming such as crop prediction, fertilizer recommendation, and automatic irrigation. The crop prediction system leverages advanced ML algorithms to analyse crucial factors. This analysis enables the system to suggest the most suitable crop for specific environmental conditions. Similarly, the fertilizer recommendation system tailors its recommendations based on the crop type and existing soil nutrient levels, optimizing crop yield and overall efficiency. Additionally, the automatic irrigation system utilizes real-time soil moisture levels and accurate weather forecasts to schedule irrigation effectively, minimizing water waste while providing the ideal amount of water for the crops. The study successfully implements and rigorously evaluates the performance of these advanced systems, presenting their achievements and potential benefits through meticulous simulations.

Keywords Artificial intelligence · Crop recommendation · Irrigation system · Fertilizer recommendation

1 Introduction

An innovative agricultural system is under development, leveraging the power of machine learning to incorporate crop selection, autonomous irrigation, and fertilizer recommendation. The primary aim of this intelligent system is to alleviate physical labour, conserve energy, and ultimately boost overall productivity for farmers [1].

K. Sharma

Department of Electronics and Communication, Bharat Institute of Technology Meerut, Meerut, Uttar Pradesh, India

D. Kumar (✉)

Bharat Institute of Technology, Meerut, Uttar Pradesh, India

e-mail: Kr.deepak05@gmail.com

The background study forms the foundational basis of this research, encompassing several essential aspects. The main project objective centres around the development of an intelligent agriculture system, aimed at optimizing field yield, providing valuable insights into crop rotation, continuous area assessment, and enhancing irrigation through selective irrigation techniques. In this context, an affordable smart farming module based on IoT and ML is introduced, utilizing cutting-edge methods to improve the accuracy of results [2, 3].

The research comprehensively explores the scope of machine learning in precision agriculture, with a specific focus on agriculture and food production. It involves a thorough comparison of various machine learning models' performance in agricultural contexts, identifying the most suitable models for agricultural applications. Additionally, the study delves into the practical implementation of multiple MLAs, for sensor data analytics in the agricultural ecosystem. A case study of an IoT-based smart farm prototype is also presented, providing valuable insights into its real-world application and benefits.

Moreover, introduces a deep reinforcement-based machine learning for smart agriculture IoT system, incorporating layers for edge computing, data transmission, data collection, and cloud computing. This system seamlessly integrates agricultural production with advanced information technologies, such as AI and cloud computing, to significantly enhance food output. The algorithms section of this study focuses on crop recommendation by leveraging key parameters such as N, P, and K values through the implementation of machine learning algorithms. A decision tree algorithm is proposed for smart irrigation, utilizing temperature, humidity, and moisture as the input parameters. The study further conducts a comprehensive analysis of various machine learning concepts for an IoT-based smart agriculture system and suggests the adoption of support vector machine (SVM) and decision tree algorithms to optimize farming processes and enhance crop growth based on soil data [4].

For crop quality and yield assessment, the study explores the integration of ML with computer vision techniques for the classification of different crop images. The proposed systems and solutions include a smart farming system that relies on artificial intelligence methods for prediction. This encompasses data collection through sensors, data storage and cleaning, and predictive calculation using AI techniques facilitated through a WSN.

Furthermore, the study presents a solution architecture for designing key performance indicators (KPIs) in precision agriculture and recommends an IoT-based smart agriculture structure with a predictive approach called WPART. This approach utilizes machine learning techniques to forecast crop productivity and drought, facilitating informed decision-making. Additionally, the study introduces an intelligent irrigation system that employs the Internet of Things (IoT) and a long short-term memory network (LSTM) to achieve precise irrigation scheduling.

Based on the cited research works, an advanced system has been developed to provide crop recommendations by considering various factors such as weather and soil conditions, crop requirements, and prevailing climatic circumstances. Moreover,

this system also facilitates automatic irrigation, weather prediction, and open weather API derived from crop and field characteristics.

First section of this paper describes about the agricultural challenges raised by environmental factors, offering possible solutions and introducing the proposed system. This section widely explains the need for innovative approaches to address these challenges and serves as a foundation for subsequent discussions on the functionalities and effectiveness of the integrated system. The second section describes about the proposed methodology. It summarizes the conceptual framework and technologies used in developing the integrated system; it also explains about the systematic approach taken to design a robust and efficient solution. The third section proposes an overview of the three main subsystems constituting the integrated agricultural system: the crop prediction system, the autonomous watering block, and the fertilizer directing system. Each subsystem is explained upon, illumination off how machine learning algorithms, weather forecasts, and crop characteristics synergistically contribute to their functionalities. The consequent section of the paper incorporates the results and discussions derived from simulations of the integrated systems. It presents a detailed analysis of the performance and outcomes of the crop recommendation, automatic watering, and fertilizer recommendation systems. Additionally, this section also describes effectiveness and potential suggestions of adopting these systems in real-world agricultural practices [5].

Finally, the paper is concluded with a complete summary of the work accomplished. It highlights the key achievements, emphasizes the significance of the integrated system, and summarizes possible future directions for further improvement. Overall, this research pays valuable knowledge to the field of agricultural technology, showcasing the huge potential of IoT and machine learning in transforming crop management practices for enhanced efficiency and sustainability.

2 Method Proposed

To optimize crop yield, the development of a multifunctional and integrated system is crucial. This sophisticated system consists of three primary subsystems: the crop recommendation system, the automatic watering system, and the fertilizer recommendation system. The crop recommendation system evaluates diverse soil and meteorological conditions at specific locations to recommend suitable crops for achieving optimal growth. The automatic watering system assesses soil moisture levels and weather data, determining the need for watering based on the plant's dampness requirement. Lastly, the fertilizer recommendation system offers guidance on the appropriate amount and timing of fertilizer application to ensure healthy crop development. The amalgamation of these three subsystems into a single main system holds the potential to significantly improve crop growth and yield.

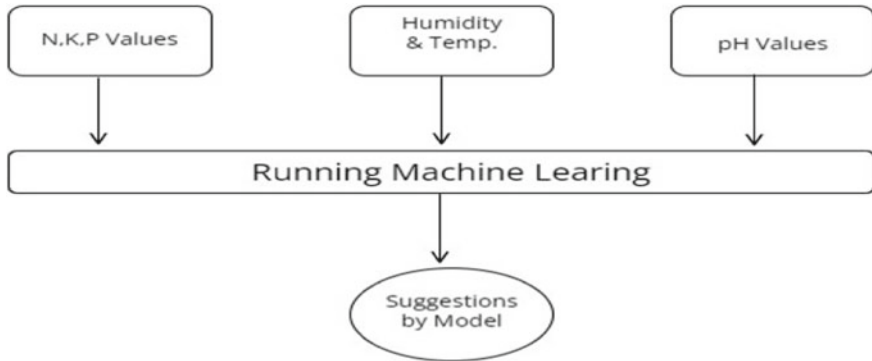


Fig. 1 Block diagram of process

2.1 Recommendation System for Crops

This system functions by initially collecting temp., rainfall data, and damp data, for the chosen location and utilizing specialized sensors to assess the Nitrogen, Phosphorus, Potassium (NPK) levels and pH of land. The weather data, comprising temperature, dampness, and rainfall, is acquired through an open weather API on a cloud platform. The collected data is then fed into a machine learning model, to determine the optimal crops for the specific environmental conditions. The initial discrete data values which are clustered using the accuracy of five distinct K-Means clustering machine learning models—Random Forest Classifier (RFC), Gradient Boosting Classifier (GBC), Decision Tree (DT), Extreme Gradient Boosting (XGB), and K-Nearest Neighbours (KNN)—are subsequently compared. Finally, the data collected from both the sensors and the cloud platform is transmitted through the most appropriate means. By conducting this comprehensive analysis, the system suggests the most suitable crop that is likely to flourish in the given location and environmental circumstances (Fig. 1).

2.2 Autonomous Watering System

This innovative irrigation system incorporates a range of sensors and advanced technologies to optimize crop irrigation and foster robust plant growth. It employs a moisture sensor to closely monitor the soil's damp content and utilizes a cloud platform which is an open weather API to observe and analyse weather conditions. Flow sensors are integrated to accurately calculate the quantity of water supplied to the soil. Additionally, temperature values are periodically monitored at fixed intervals (P hours).

The system operates by intelligently assessing the soil moisture levels and weather forecasts. When low soil moisture and no imminent rain are detected, it evaluates the

temperature conditions. If the temperature falls within the normal range, the system transverse the pump, watering the plants for some minutes to fulfil the crop minimum mist requirements. In cases of high temperatures, the pump extends its operation to $X + Y$ minutes, providing ample water to meet the crop's maximum moisture demands. However, when rain is predicted, the system intelligently postpones watering until the rain has begun, thereby avoiding unnecessary irrigation.

This intelligent irrigation system relies on microcontrollers to execute precise watering strategies based on factors such as flow sensor data, the field size, and specific water needs of the harvested crop. As a result of this advanced technique, plants receive the ideal amount of moisture, leading to enhanced growth and maximizing crop yields.

In conclusion, this innovative irrigation system, powered by state-of-the-art sensor technologies and microcontrollers, ensures that crops receive the optimal water supply, adapting effectively to varying weather conditions. By diligently monitoring soil moisture and weather forecasts, the system optimizes irrigation timings, culminating in healthier and more productive crops. This integrated approach to irrigation marks a significant advancement in agricultural practices, fostering sustainability, and efficiency in crop cultivation.

2.3 Fertilizer Suggestion System

The fertility of the soil is crucial for the successful growth of crops, and it largely depends on the levels of essential proponentes like nitro (N), potash (K), and phosphorus (P) present in the soil. To correctly measure these values, a dedicated NPK sensor is used. Each crop has a specific set of NPK values that are used when evaluating the soil quality. The data received from the NPK sensor is then carefully examined with the type of crop being grown. This allows the system to check whether the soil fertilization is at an optimal level. If the soil fertility is found to be ideal, the system recommends using regular fertilizers to maintain the soil's fertility. However, if the soil fertility falls below the best level, the system provides valuable suggestions for applying additional fertilizers alongside the standard ones to meet the crop's specific nutrient requirements. Additionally, the system offers precise fertilization dosing recommendations, ensuring that the soil fertility is preserved and optimized until the harvest period. By providing tailored fertilization advice, this intelligent system plays a vital role in promoting and sustaining the health and productivity of crops throughout their entire growth cycle.

3 Discussion and Results

The proposed models have undergone simulation, and this section presents the results and findings.

3.1 Crop Recommendation System

Table 1 presents a comprehensive comparison of the accuracy achieved by five distinct models. Based on the research findings, it is evident that the K-Nearest Neighbours (KNN) machine learning algorithm outperforms the other models when applied to the data set, making it the ideal choice for implementation in the crop recommendation system.

Figure 2 displays essential input values such as the pH level and the N, K, and P nutrient levels of the land. Additionally, the system acquires crucial meteorological data, including humidity, rainfall, and temperature, from the cloud platform, considering the provided city input. By integrating all this relevant information, the KNN machine learning model efficiently identifies the most suitable crop for cultivation in the specific area, leading to an accurate recommendation, such as suggesting the cultivation of oranges.

Furthermore, Table 2 illustrates the crop prediction system’s diverse suggestions for various discrete type of in data set, underscoring the system’s adaptability to different atmospheric conditions. This exemplifies the system’s capacity to offer personalized crop recommendations, thereby enhancing agricultural practices and optimizing crop yields for increased productivity and success.

Table 1 Comparison of accuracy achieved by five distinct models

	Model	Accuracy	Thain acc
1	DT	0.993182	1
2	GBC	0.995455	1
3	KNN	0.977273	0.988636
4	RFC	0.997727	1
5	XGB	0.993182	1

```

Enter the Value of N : 37
Enter the Value of P : 34
Enter the Value of K : 30
Enter the Value of pH : 6.79
Enter the City : Delhi
Crop Suggested : Banana
PS C:\Users\ehskusa\OneDrive - Ericsson\Documents\VScode\ML\Numpy>
    
```

Fig. 2 Simulation for crop recommendation system

Table 2 Crop recommendation and their different atmospheric parameters

N	P	K	Temperature	Humidity	pH	Rainfall	Label
90	42	43	20.87974371	82.00274423	6.502985292	202.9355362	Rice
71	54	16	22.61359953	63.69070564	5.749914421	87.75953857	Maize
13	60	25	17.13692774	20.59541693	5.68597166	128.256862	Kidney beans
3	49	18	27.91095209	64.70930606	3.692863601	32.67891866	Moth beans
19	55	20	27.43329405	87.80507732	7.18530147	54.73367631	Mung bean
56	79	15	29.48439992	63.19915325	7.454532137	71.89090748	Black gram
2	24	38	24.55981624	91.63536236	5.922935513	111.9684622	Pomegranate
91	94	46	29.36792366	76.24900101	6.149934034	92.82840911	Banana
119	25	51	26.47330219	80.92254421	6.283818329	53.65742581	Watermelon
115	17	55	27.57826922	94.11878202	6.776533055	28.08253201	Muskmelon
24	128	196	22.75088787	90.69489172	5.521466996	110.4317855	Apple
22	30	12	15.78144173	92.51077745	6.354006744	119.035002	Orange

3.2 Autonomous Watering System

In Table 3, the water requirements for each crop are presented, indicating the recommended range of moisture in millimetres (mm) essential for promoting healthy growth. The system guarantees that the soil mist content is maintained for this optimal range to support the crop’s well-being effectively.

To achieve this, the system collects input data concerning the specific cultivated crop and the corresponding city of cultivation. Utilizing the cloud platform, the system gathers data on temperature and rainfall to make well-informed irrigation decisions.

Table 3 Water requirements for different crop

Crop	Minimum (in mm)	Maximum (in mm)
Apple	800	1230
Banana	1100	1200
Coffee	200	300
Cotton	500	920
Grape	800	300
Maize	350	1500
Orange	400	380
Rice	1200	900

3.3 *Fertilizer Recommendation System*

The successful growth of crops relies heavily on the NKP data, which are crucial for getting selective outcomes. To ensure healthy crop development, the system thoroughly analyses these values and recommends suitable fertilizers based on the obtained results.

The analysis involves input data, including the N, K, and P reading retrieved from the land through the NKP sensing device, also as information about the specific crop being cultivated. This data enables the system to propose the most appropriate fertilizers and their corresponding dosages, facilitating the crop's healthy growth. By closely detecting changes in soil NKP data, the system effectively ensures that the crop receives the essential nutrients it needs, resulting in improved growth and enhanced overall yield.

4 Conclusion

A novel system has been developed, capable of providing crop recommendations that suit the specific local climate conditions and soil characteristics, as well as automating the watering process for plants as needed and suggesting the most appropriate fertilizers on behalf of the type of crop. The system is designed to offer personalized suggestions for a wide range of crops, including rice, grape, maize, cotton, apple, banana, coffee, and orange, ensuring optimal outcomes for farmers. By adopting this technology, farmers can make informed decisions, reduce manual labour, save energy, and ultimately boost their agricultural productivity. As part of future advancements, the system aims to incorporate a disease prediction feature that utilizes image classification to forecast crop diseases, thereby further enhancing its functionality and contributing to the sustainable development of agriculture.

References

1. Varghese R, Sharma S (2018) Affordable smart farming using IoT and machine learning. In: 2018 second international conference on intelligent computing and control systems (ICICCS), pp 645–650
2. Mekonnen Y, Namuduri S, Burton L, Sarwat A, Bhansali S (2019) Machine learning techniques in wireless sensor network based precision agriculture. *J Electrochem Soc* 167(3):037522
3. Bu F, Wang X (2019) A smart agriculture IoT system based on deep reinforcement learning. *Future Gener Comput Syst* 99:500–507
4. Singh RK, Berkvens R, Weyn M (2021) AgriFusion: an architecture for IoT and emerging technologies based on a precision agriculture survey. *IEEE Access* 9:136253–136283
5. Togneri R et al (2019) Advancing IoT-based smart irrigation. In: *IEEE Internet Things Mag* 2(4):20–25

Multifocus, Infrared and Visible Light Image Fusion Using Non-subsampled Shearlet Transform and SUSAN Operator



Dileep Kumar Sharma  and Abhilasha Sharma

Abstract Image fusion is the multipurpose and multidisciplinary field studied by most of the researchers these days. This research applies non-subsampled shearlet transform (NSST) and a human visual system operator to fuse multifocus and infrared visible light image pairs. Initially the images are decomposed in different subbands using NSST and then local energy and SUSAN operator based fusion rules are applied on these subbands. The fusion results are compared visually and using different performance measures with three existing fusion schemes. The fusion comparison of the schemes show that the proposed scheme is capable of fusing the multifocus and infrared visible image pairs. The average run time of the proposed scheme is quite low as compared to other schemes.

Keywords Multifocus image fusion · Infrared–visible light image fusion · Smallest univalue segment assimilating nucleus · Non-subsampled shearlet transform

1 Introduction

Imaging sensors are widely used to capture different regions of interest in a scene or organ. Due to the limitation of depth of focus of digital cameras, it is not possible to acquire all the regions of an image with equal focus. In one image some of the objects or regions appear to be in focus whereas objects at other distances from image sensor may not be in focus and appear blurred. Image fusion techniques are employed to combine such images to get a complementary image rich in focused information [1]. Similarly a visible image acquired in proper illumination condition appears clear however, it doesn't show targets clearly if not properly illuminated [2]. Thermal radiation objects such as humans and animals are clearly visible in infrared

D. K. Sharma (✉) · A. Sharma
Chitkara University School of Engineering & Technology, Chitkara University, Himachal Pradesh, India
e-mail: dileep.kumar@chitkarauniversity.edu.in

(IR) images [3]. The process to obtain the combined image in such cases carrying the enhanced information is called as image fusion [4, 5]. Fusion process can be carried in various ways such as combining pixels, regions or features. Pixel fusion technique combines pixels directly, region based method divides the image into regions and features of regions are employed in fusion process. Pixel fusion method is most widely studied by the researchers because of its simplicity and chances of introducing distortion are less. After acquiring images preprocessing is the first step in fusion process which involves registration and conversion from one color domain to other domain etc. In next step spatial domain methods or multiscale transforms are used to split the images into various scales and bands. Multiscale transforms (MST) are mostly preferred because their response matches with human visual system. After this step the pixels with outstanding activity are searched which take part in the construction of fusion image. In an alternative approach grouping of pixels and consistency verification may also be done before reconstruction of the final fused image to improve the fusion results.

Many researchers in state of art literature have done fusion of multifocus, IR-visible image pairs. Yang et al. [1] have used nonsubsampled contourlet (NSCT) transform to fuse multifocus images. The authors have applied the sum modified Laplacian and energy of log Gabor filters as fusion rules for low and high frequency subbands, respectively. Li et al. [6] have compared multiresolution transforms such as different types of wavelets, curvelets and contourlets to fuse multifocus, medical and IR-visible image pairs. Liu et al. [7] have utilized neighbour distance and morphological operations to fuse multifocus images. Liu et al. [4] combined MST and sparse representation in the fusion of multifocus, IR-visible images to overcome the individual limitation of these methods. Kim and Hwang [8] and Tang et al. [9] applied deep learning network to fuse IR-visible images. Tian et al. [10] fused multifocus images using two stage model. The first stage contains DenseNet network—polarization self-attention for training and fusion stage contains edge feature map. In our recent research in [11], we have applied smallest uni-value segment assimilating nucleus (SUSAN) operator in the fusion of CT-MRI images, however the suitability of this operator in the fusion of multifocus and IR-visible images is not tested. This research is an attempt to explore the suitability of this operator in the fusion of multifocus and IR-visible image pairs.

2 Methodology

Decompose the source images into low frequency (LF) and high frequency (HF) subbands using NSST. Low frequency subbands are fused using weighted sum of pixel intensities within a 3×3 window where the weight is varied in accordance with a Gaussian curve. This is expressed as below:

$$A_{\text{Low}}(i, j) = \sum_{x=-1}^1 \sum_{y=-1}^1 w \times C_1(x + i, y + j) \quad (1)$$

where, $A_{\text{Low}}(i, j)$ represents the activity level of a LF subband pixel at position i, j of a source image, w is Gaussian weight, and C_1 represents the intensity of LF subband pixel at position i, j . The corresponding pixels of two LF sub-band images are compared and the pixel with high activity as expressed by Eq. (1) is kept to form the LF sub-band of the fused image.

The SUSAN operator is applied to fuse HF subband pixels. This operator gives an edge map to find the pixels with high edge response that are involved in pixel fusion [12]. It is given as:

$$\text{map}(\vec{a}, \vec{a}_0) = \exp\left(-\left(\frac{g(\vec{a}) - g(\vec{a}_0)}{t_h}\right)^6\right) \quad (2)$$

where, \vec{a} represent position of nucleus, \vec{a}_0 indicate position of pixel under the mask, $g(\vec{a})$ represent intensity at position \vec{a} and t_h is intensity gradient threshold which is set to 20. The edge information of corresponding pixels of the two HF sub-bands of source images are compared and the pixel which has high edge information is kept in fusion image. Final fused image is reconstructed by taking inverse NSST of the fused LF and HF subbands of NSST decomposition.

3 Dataset

A secondary dataset containing 05 multi-focus image pairs and 05 infrared visible image pairs of sizes 256×256 and 320×240 is downloaded from [13].

4 Performance Parameters

The proposed method is compared using entropy, mutual information, normalized mutual information and mean parameters which are defined as follows:

Entropy represents the average texture in the image [14, 15].

Mutual information measures the statistical similarity between the sources images and fusion image.

Normalized mutual information (NMI) is calculated by finding MI between the source and fused images and dividing it by the square root of the product of their entropies [11].

Mean quantifies average grey level in the fusion image.

A higher value of all these parameters indicates a better fused image.

5 Results and Discussion

Experiment 1: Visual analysis of the proposed method is done with three state of the art methods named as stationary wavelet (SWT) [16], adaptive sparse representation (ASR) [17] and NSST based scheme [18]. Visual analysis and comparison of the developed method with other methods is done for multi-focus and visible infrared image pairs. Figure 1 shows the visual fusion results of various methods for a pair of multi-focus images. From this figure, it is seen that resolution of far focus objects is slightly compromised for the SWT scheme as compared to other schemes. The resolution of near focus objects is compromised for the NSST based scheme and the proposed scheme as compared to the ASR method. Figure 2 shows the fusion results of different methods for a visible infrared image pair. The fusion differences between different methods are visible in the arrow marked region of this figure. There is a reduction in the intensity of the rectangular region as marked by arrow for ASR scheme. The SWT, NSST and the proposed scheme fuse this region in their output however the contrast of the rectangular region is better in the proposed scheme.

Experiment 2: In this experiment quantitative comparison w.r.t. performance parameters is done. It is seen from Table 1 that average entropy and average mean value of the proposed method is highest among all the other schemes as shown in bold. This shows that the fusion images obtained by the presented scheme have higher amount of texture and brightness. Average MI and NMI parameters of the proposed scheme are better than SWT and NSST based schemes for the multi-focus images

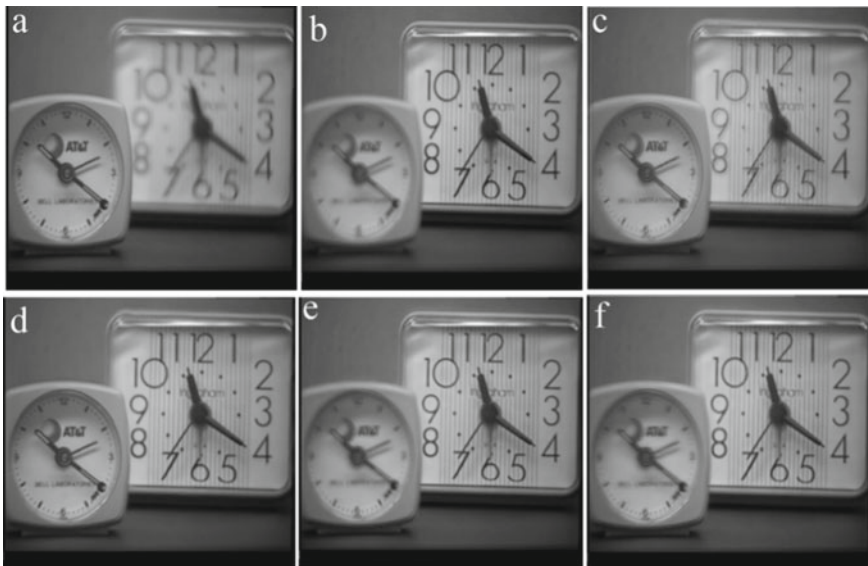


Fig. 1 Comparative fusion results of a multifocus image pair: **a** near focus image, **b** far focus image, **c** SWT method, **d** ASR scheme, **e** NSST, **f** proposed scheme

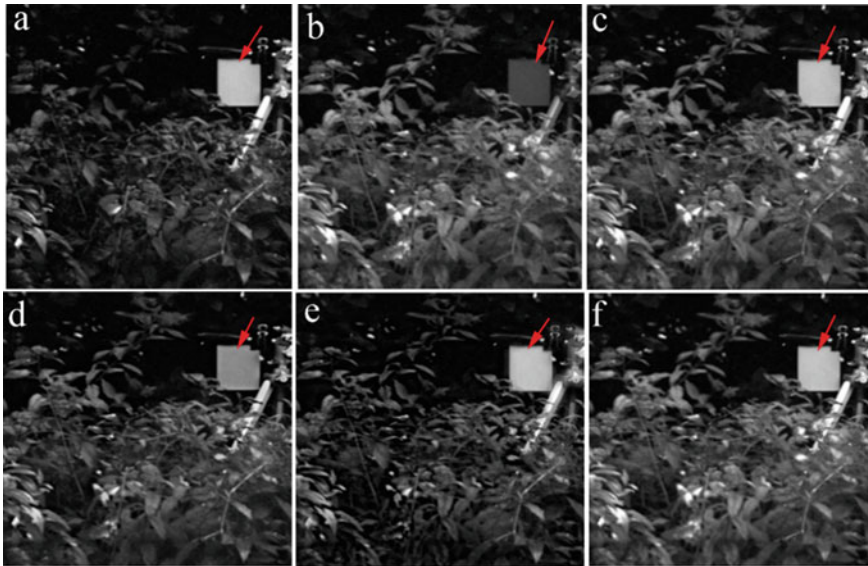


Fig. 2 Comparative fusion results of a visible IR image pair: **a** visible image, **b** infrared image, **c** SWT method, **d** ASR scheme, **e** NSST, **f** Proposed scheme

dataset. However, these parameters are smaller than ASR scheme. The highest value of these parameters are shown in bold and signifies higher amount of source information available in the fused image. Table 2 shows that the average MI and NMI parameters of the proposed scheme are better than ASR and NSST based methods for visible infrared images dataset. However, these are smaller than SWT scheme. The average entropy and average mean value of the proposed method is highest among all the schemes.

Experiment 3: The time taken to fuse the different image pairs is recorded and displayed in Table 3. From this table it clear that the run time taken by the proposed SUSAN based method is quite low as compared to other schemes.

Table 1 Average parametric performance of different fusion approaches for 05 pairs of multifocus images

Approach	MI	NMI	Entropy	Mean
Near focus image	–	–	–	106.7332
Far focus image	–	–	–	105.6590
SWT [16]	8.0016	1.1059	7.2343	111.4271
ASR [17]	8.4194	1.1639	7.1380	109.6484
NSST [18]	8.0649	1.1139	7.1436	109.5336
Proposed	8.1294	1.1239	7.2609	112.3857

Table 2 Average parametric performance of different fusion approaches for 05 pairs of infrared visible light images

Approach	MI	NMI	Entropy	Mean
Visible image	–	–	–	56.5538
Infrared	–	–	–	92.9196
SWT [16]	5.1928	0.7231	7.2054	102.1175
ASR [17]	3.5683	0.5127	6.9748	74.8838
NSST [18]	3.7532	0.5717	7.0085	67.6141
Proposed	4.3510	0.6841	7.2109	103.0180

Table 3 Average computation time (in second) comparison

Dataset	SWT [16]	ASR [17]	NSST [18]	Proposed
Multifocus	143.2951	360.0857	58.6937	7.1743
Visible-infrared	142.3598	219.4421	60.5845	7.9241

6 Conclusion

A SUSAN and weighted sum of pixel intensity based multifocus and IR-visible light image fusion scheme using NSST has been proposed in this research. Fusion quality comparison has been done with three state of art schemes. From the results it is evident that the NSST-SUSAN based method can be employed to fuse multifocus and IR-visible images with comparable results and minimum computation time as compared to other schemes.

References

1. Yang Y, Tong S, Huang S, Lin P (2014) Multifocus image fusion based on NSCT and focused area detection. *IEEE Sens J* 15(5):2824–2838
2. Lewis JJ, O’Callaghan RJ, Nikolov SG, Bull DR, Canagarajah N (2007) Pixel- and region-based image fusion with complex wavelets. *Information fusion* 8(2):119–130
3. Kaur R, Tiwari RK, Maini R, Singh S (2023) A framework for crop yield estimation and change detection using image fusion of microwave and optical satellite dataset. *Quaternary* 6(2):28
4. Liu Y, Liu S, Wang Z (2015) A general framework for image fusion based on multi-scale transform and sparse representation. *Information Fusion* 24:147–164
5. Singh S, Tiwari R, Sood V (2022) Estimation of landcover types over himalayan region with the classification of optical and microwave-based image fusion dataset. *Int Arch Photogramm Remote Sens Spat Inf Sci* 43:523–528
6. Li S, Yang B, Hu J (2011) Performance comparison of different multi-resolution transforms for image fusion. *Information Fusion* 12(2):74–84
7. Liu C, Long Y, Mao J (2016) Energy-efficient multi-focus image fusion based on neighbor distance and morphology. *Optik* 127(23):11354–11363
8. Kim J-H, Hwang Y (2023) Infrared and visible image fusion using a guiding network to leverage perceptual similarity. *Comput Vis Image Underst* 227:103598

9. Tang W, He F, Liu Y (2022) YDTR: infrared and visible image fusion via Y-shape dynamic transformer. *IEEE Transactions on Multimedia*
10. Tian B, Yang L, Dang J (2023) Fine-grained multi-focus image fusion based on edge features. *Sci Rep* 13(1):2478
11. Ramlal SD, Sachdeva J, Ahuja CK, Khandelwal N (2022) Multimodal medical image fusion using nonsubsampling shearlet transform and smallest uni-value segment assimilating nucleus. *Int J Pattern Recognit Artif Intell* 36(04):2257001
12. Smith SM, Brady JM (1997) SUSAN—a new approach to low level image processing. *Int J Comput Vision* 23(1):45–78
13. <http://home.ustc.edu.cn/~liuyu1>. Last accessed 09 Aug 2017
14. Ranta S, Gupta S, Sharma DK (2022) Image enlargement technique based on the combination of SVD and Hermite interpolation. In: 2022 2nd international conference on emerging frontiers in electrical and electronic technologies (ICEFEET). IEEE, pp 1–5
15. Gupta S, Sharma DK, Ranta S (2022) A new hybrid image enlargement method using singular value decomposition and cubic spline interpolation. *Multimedia Tools and Applications* 81(3):4241–4254
16. Ramlal SD, Sachdeva J, Ahuja CK, Khandelwal N (2018) Brain CT and MR image fusion framework based on stationary wavelet transform. In: *Advances in computer and computational sciences*. Springer, pp 445–453
17. Liu Y, Wang Z (2015) Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Proc* 9(5):347–357
18. Kong W, Zhang L, Lei Y (2014) Novel fusion method for visible light and infrared images based on NSST–SF–PCNN. *Infrared Phys Technol* 65:103–112

Leveled Approach of Context Setting in Semantic Understanding of Remote Sensing Images



Stuti Ahuja  and Sonali Patil 

Abstract Geographic data is carried via high spatial resolution (HSR) remote sensing (RS) images. Understanding HSR remote sensing images requires a thorough exploration of geo-objects and their spatial relationships. The task of semantic understanding is divided into four subtasks preprocessing, object detection, identifying geospatial relations, and semantic description generation. These semantic descriptions will be of great significance in the field of urban planning, RS image retrieval, military intelligence, etc. A leveled approach for context setting is proposed in this paper for the semantic understanding of RS images. This approach comprises various steps such as scene classification, multi-label classification, object detection and labeling, description generation using geospatial relationships, and finally generating detailed descriptions for input RS images. Results show that the proposed method performs well in generating a description that considers the overall context of the RS scene.

Keywords Scene classification · Object detection · Multi-label classification · Semantic understanding · Remote sensing · Image processing

1 Introduction

Semantic understanding of RS images generates a geospatial relation description, instead of simply predicting individual keywords related to the RS scene. To implement accurate geospatial relation-based description generation, geo-objects should

S. Ahuja (✉)

K.J. Somaiya College of Engineering, University of Mumbai, Mumbai, India

e-mail: stuti.ahuja@somaiya.edu

NMIMS School of Technology Management and Engineering, Navi Mumbai, India

S. Patil

K.J. Somaiya College of Engineering, Somaiya Vidyavihar University, Mumbai, India

e-mail: sonalipatil@somaiya.edu

be initially recognized, and their geospatial relation to be expressed as basic components of image semantic understanding. This paper proposes an intelligent system that will effectively provide geospatial semantics for high spatial resolution remote sensing image understanding.

Over the years, the issue of information retrieval and mining from remote sensing images has received the majority of attention in research on this paradigm [1]. These methods mainly make use of the notion of visual attention, scene-specific contexts, knowledge graphs, and game theoretic optimization [2–7].

After identifying the need for semantic descriptions in the form of sentences, a few RS image description methods are proposed in the literature [8–11]. Few approaches for the semantic understanding of RS images have been proposed in the literature [12–15]. Most of them are generating sentences based on the sentences the deep learning model trained on. Even though the generated description describes the contents present in the image, the overall context of the entire image is missing.

The main contribution of this work can be listed as: (i) A methodology is devised to generate multi-sentence descriptions for remote sensing images. (ii) The descriptions are generated at multiple stages of semantic understanding for context setting.

The rest of the paper is organized as follows. The proposed methodology for generating detailed descriptions for RS images using leveled approach is discussed in Sect. 2. Section 3 comprises results and discussions followed by the conclusion in Sect. 4.

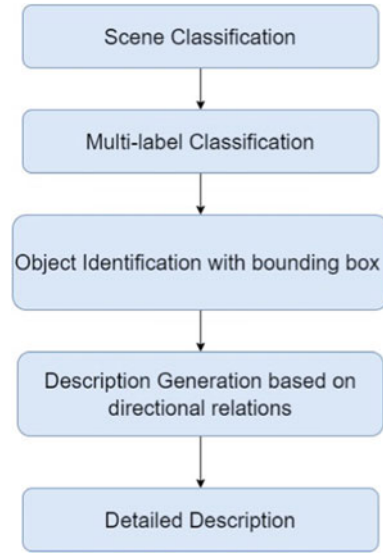
2 Methodology

The leveled approach for context setting is proposed here for semantic understanding of RS images to give context to the overall image, then gradually leads toward detailed description. This approach comprises various steps such as scene classification, multi-label classification, object detection and labeling, description generation using geospatial relationships and finally generates detailed descriptions for input RS images as shown in Fig. 1.

2.1 Scene Classification

Semantic-level scene classification of remote sensing image seeks to assign each remote sensing image with a distinct class label. This area has recently emerged as a new research stream by virtue of rapid development in machine learning techniques. Scene classification gives the idea about what the image is by generating a one-word caption for the image. A variety of CNN-based deep learning approaches have been proposed in the literature for remote sensing scene classification [16–21]. Scene classification can be binary or multi-class, based on the number of classes used during the training phase of deep learning. Each image in a binary classification is given

Fig. 1 Leveled approach for context setting in RS scene understanding



one and only one label from two mutually exclusive classes, while in a multi-class classification, each data sample is given one and only one label from more than two classes.

CNNs typically include three layers: convolutional, pooling, and dense layers and are used for classification purposes. The dropout layer is also essential for avoiding overfitting. The proposed methodology seeks to increase the CNN model’s image scene classification accuracy. Considering the model’s strong performance, ResNet, a residual network, is chosen. The abbreviation ResNet stands for Residual Network, a specialized network introduced by Microsoft. ResNet connects inputs of network layers into farther parts of that network to allow shortcuts.

2.1.1 Dataset Details

The datasets which are available for RS image scene classification are summarized in Table 1. The NWPU-RESISC45 dataset [16] is used here for scene classification.

2.2 Multi-label Classification

High spatial resolution RS image comprises multiple geospatial objects. Detecting and labeling these objects will be of great significance in various applications such as image retrieval, image understanding. Single-class classifications will not suffice the

Table 1 Datasets available for RS image scene classification [16]

Datasets	Images per class	Scene classes	Total images	Spatial resolution (m)	Image sizes	Year
UC Merced land-use	100	21	2100	0.3	256 × 256	2010
WHU-RS19	~ 50	19	1005	Up to 0.5	600 × 600	2012
SIRI-WHU	200	12	2400	2	200 × 200	2016
RSSCN7	400	7	2800	–	400 × 400	2015
RSC11	~ 100	11	1232	0.2	512 × 512	2016
Brazilian coffee scene	1438	2	2876	–	64 × 64	2015
NWPU-RESISC45	700	45	31,500	~ 30 to 0.2	256 × 256	2016

problem of object identification and annotation; therefore, multi-class land cover classification is required to identify various object categories from RS images. Multi-class classification [22, 23] is the method of identifying multiple object classes present in the RS image and annotating the image with those object classes. Multi-class classification is implemented using Fastai deep learning API.

2.2.1 Dataset Details

UC Merced Land Use Dataset [24]: Details of the dataset are given in Table 2. This dataset has 2100 high-resolution RS images in total. There is a vector, $C = [c_1, c_2, c_3, \dots, c_{17}]$, for each image, representing the presence or absence of the objects of the classes from c_1 to c_{17} . Different images will have different numbers of object classes associated with them in multi-label classification; therefore, the number of labels for every image will vary. These vectors are required for providing training to the deep learning model.

Table 2 Table captions should be placed above the tables

Classes	17 classes: airplane, bare soil, buildings, cars, chaparral, court, dock, field, grass, mobile-home, pavement, sand, sea, ship, tanks, trees, water
Total number of images	2100
Size of image	256 × 256 pixels
Resolution	

2.3 Object Identification with Bounding Boxes

Various object detection techniques are available which can be categorized broadly as machine learning-based, deep learning-based, and object-based image analysis (OBIA) [25]. YOLOv5 object detection is used here for getting objects in the form of a bounding box which also gives the location of each object. YOLOv5 is popular because of its speed and accuracy [26–29]. This location information of objects is further used in the identification of relations between objects.

2.4 Description Generation Based on Relationships

The directional and proximity relationships are determined and relationship triplets are formed after object identification and labeling are complete. For the purpose of creating descriptions for RS images, several methods exist [30]. The basic template-based method is employed to generate semantic descriptions because the proposed methodology places greater emphasis on relationship triplet generation. The template used is “The ⟨Object i ⟩ is on the ⟨direction⟩ side of the ⟨Object j ⟩” because directional relationships are taken into account in this situation. Each relationship triplet, “Object 1, Direction, Object 2,” results in a single sentence. The detailed methodology for relationship triplet generation is given in [31].

2.5 Detailed Description

The detailed semantic description is generated by combining the results of scene classification, multi-label classification, and directional and proximity relation identification at the end of all processing steps.

3 Results and Discussion

3.1 Scene Classification Results

The RS image scene classification is implemented using two deep learning techniques: first Sequential Convolution Neural Network (CNN) model with the Keras framework, and second, the ResNet18 model with the Fastai framework, and the results of both models are compared with each other. Scene classification results of sequential CNN and ResNet18 are shown in Figs. 2 and 3, respectively. In Fig. 3, two labels are there on every image. The first represents the ground truth, i.e., actual class to which the image belongs and the second label is the predicted class.



Fig. 2 Scene classification results using basic Keras sequential CNN model



Fig. 3 Scene classification results using ResNet18 (Fastai framework)

Table 3 Performance evaluation of scene classification

Model	Validation accuracy
Keras sequential CNN	0.5437
ResNet18 (Fastai framework)	0.9111

3.1.1 Performance Analysis

The performance analysis of scene classification is shown in Table 3 and the confusion matrix for ResNet18 is shown in Fig. 4.

The results show that ResNet18 not only outperforms the sequential CNN model but also other machine learning and deep learning techniques [17]. Scene classification when implemented on the same dataset using handcrafted feature-based techniques such as color histogram, LBP, and GIST gives maximum accuracy around 27%. Unsupervised machine learning techniques such as BoVW give maximum

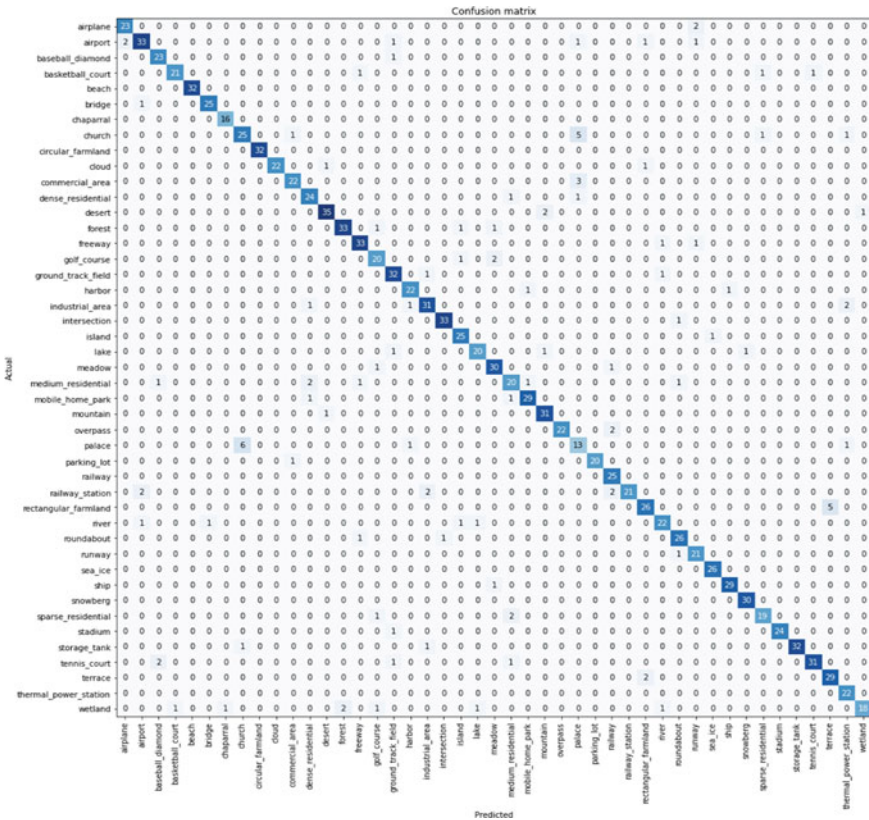


Fig. 4 Confusion matrix for scene classification using ResNet18 model

accuracy of 45% and when deep learning models are used (AlexNet, VGGNet, and GoogleNet) give maximum accuracy of around 80%.

3.2 Multi-label Classification Results

As a first step in the implementation of multi-class classification using deep learning, a few transforms have been applied to RS images to increase the range of our data collection and promote effective model learning. To improve the dataset, random flipping and normalization of certain images are carried out and data are created for training the deep learning model using the data bunch API of the Fastai library. After performing preprocessing, two deep learning models have been implemented to test the performance of multi-class classification, ResNet-50 and AlexNet. These deep learning architectures have been tested on the UC Merced land use dataset. The results show that the deep learning approach recognizes geospatial objects and assigns multiple labels to RS images based on the detected objects with 96% accuracy using the ResNet-50 model and 94% accuracy using the AlexNet model. The results of multi-label classification using ResNet-50 are shown in Fig. 5. The results of the multi-label classification show that AlexNet and ResNet-50 outperform other deep learning models when implemented on the same dataset.

3.3 Object Detection Results

Object detection and labeling are carried out using the DIOR dataset [25]. The sample results of YOLOv5 object detection on DIOR dataset images are shown in Fig. 6. The performance evaluation of object detection is carried out using the mean average precision parameter as given in Eqs. (1) and (2). Table 4 shows the performance analysis of the model.

Mean average precision (mAP),

$$AP = \sum_{k=0}^{n-1} [\text{Recalls}(k) - \text{Recalls}(k + 1)] * \text{Precisions}(k), \quad (1)$$

$$mAP = \frac{1}{n} \sum_{k=1}^n AP_k. \quad (2)$$

The performance analysis shows that the overall mAP using YOLOv5 object detection with the Darknet framework is 62.1%. Various other deep learning frameworks were implemented in literature [25] on DIOR object detection datasets such as ResNet-101, RetinaNet, VGG16, Faster R-CNN, Mask R-CNN, and YPLOv3. Out of all, RetinaNet, when combined with ResNet-101, and PANet with ResNet-101

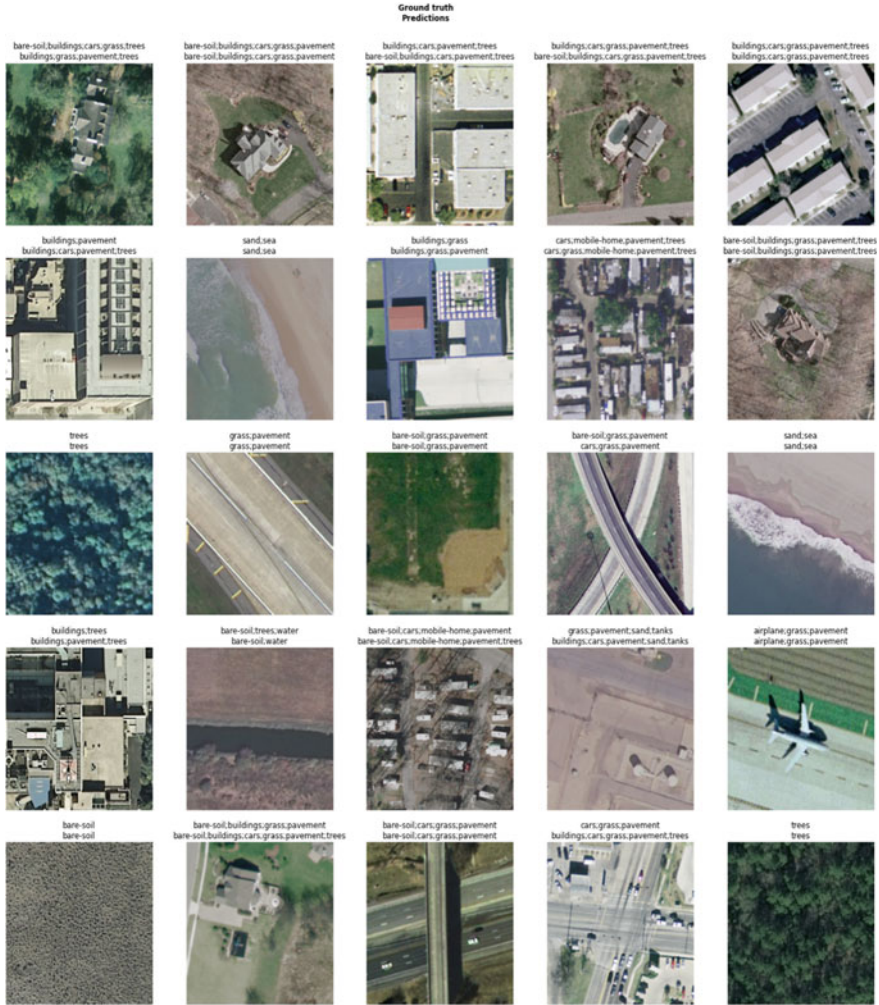


Fig. 5 Multi-label classification results using ResNet-50 (Fastai framework)

give a maximum mAP of 66.1% which is comparable with our YOLOv5 model. Also, YOLOv5 is fast and simpler as compared to other models.

3.4 Description Generation Based on Relationships

One sentence is generated for each relationship triplet (Object 1, Direction, Object 2). The description generated for one sample image is shown in Fig. 7.

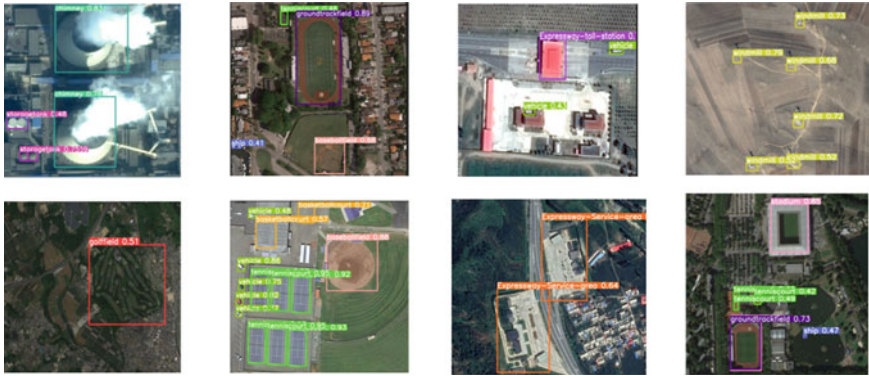


Fig. 6 Object detection results using YOLOv5 on the DIOR dataset

Table 4 Performance evaluation of object detection

Class	Images	Instances	P	R	mAP@.5
All	2501	25,954	0.739	0.582	0.621
Airplane	2501	99	0.438	0.307	0.329
Airport	2501	521	0.056	0.207	0.279
Baseball field	2501	6018	0.702	0.328	0.391
Basketball court	2501	1461	0.88	0.807	0.864
Bridge	2501	445	0.88	0.755	0.8_
Chimney	2501	745	0.864	0.796	0.833
Dam	2501	360	0.713	0.425	0.46
Expressway service area	2501	405	0.75	0.689	0.758
Expressway toll station	2501	671	0.654	0.359	0.435
Golf course	2501	217	0.831	0.498	0.618
Ground track field	2501	143	0.675	0.566	0.588
Harbor	2501	7672	0.766	0.761	0.79
Overpass	2501	146	0.834	0.445	0.5_
Ship	2501	644	0.756	0.658	0.686
Stadium	2501	137	0.586	0.547	0.505
Storage tank	2501	1612	0.811	0.88	0.899
Tennis court	2501	213	0.842	0.7	0.705
Train station	2501	4148	0.786	0.672	0.707
Vehicle	2501	154	0.794	0.708	0.79
Wind mill	2501	143	0.557	0.524	0.493

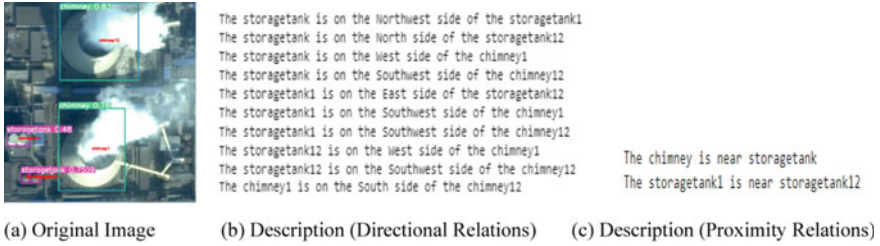


Fig. 7 Description generation results using directional and proximity relationship

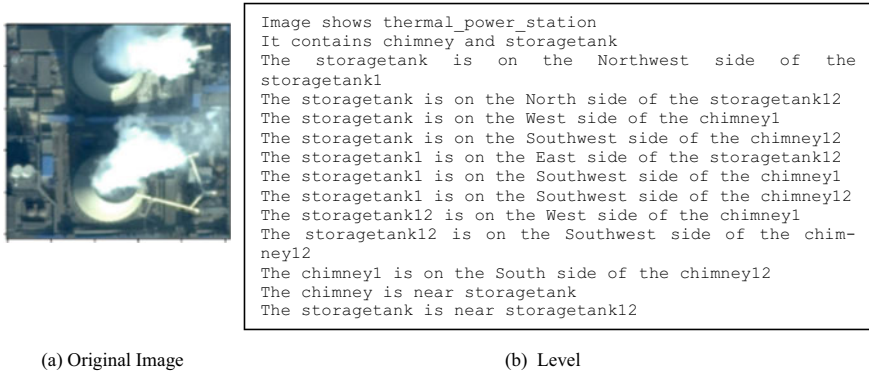


Fig. 8 Detailed description generation using leveled approach

3.5 Detailed Description

The detailed semantic description is generated by combining the results of scene classification, multi-label classification, and directional and proximity relation identification at the end of all processing steps. One example is shown in Fig. 8.

4 Conclusion

Leveled approach for context setting for semantic understanding of RS images is proposed in this report. This approach is divided into various steps such as scene classification, multi-label classification, object detection, and labeling, description generation using geospatial relationships, and finally generates detailed descriptions for input RS image.

Scene classification is carried out using ResNet18 with the Fastai deep learning architecture on the NWPU-RESISC45 dataset with 45 classes. When applied to the same dataset, this technique outperforms other machine learning and deep learning

models with 91% accuracy. AlexNet and ResNet-50 models are used for multi-label classification along with the Fastai deep learning framework. ResNet-50 model gives around 96% accuracy on UC Merced 17 class dataset, which is better than other models. Object detection and labeling are carried out using YOLOv5 with 20 class DIOR dataset giving around 62% accuracy (mean average precision). Due to the speed and simplicity, YOLOv5 model is chosen for object detection in the proposed framework. Then, directional and proximity relations have been identified and represented in the form of sentences. At the end, the detailed description is generated by combing outputs of all stages and context setting is achieved using this leveled approach. Even though the proposed method effectively generates detail description for the input RS image, the work can be extended by creating a common dataset for all of the steps that would train the model as a whole.

References

1. Gu Y, Wang Y, Li Y (2019) A survey on deep learning-driven remote sensing image scene understanding: scene classification, scene retrieval and scene-guided object detection. *Appl Sci* 9(10)
2. Sreela SR, Idicula SM (2019) Dense model for automatic image description generation with game theoretic optimization. *Information* 10(11)
3. Xu N, Liu AA, Liu J, Nie W, Su Y (2019) Scene graph captioner: image captioning based on structural visual representation. *J Vis Commun Image Represent*
4. Anderson P, Fernando B, Johnson M, Gould S (2016) SPICE: semantic propositional image caption evaluation
5. Yang Z, Qin Z, Yu J, Hu Y (2018) Scene graph reasoning with prior visual relationship for visual question answering, pp 1–12
6. Yao T, Pan Y, Li Y, Mei T (2018) Exploring visual relationship for image captioning, vol 1, pp 1–16
7. Gao W, Zhu Y, Zhang W, Zhang K, Gao H (2019) A hierarchical recurrent approach to predict scene graphs from a visual-attention-oriented perspective. *Comput Intell* 35(3):496–516
8. Zhang CLX, Li X, An J, Gao L, Hou B (2017) Natural language description of remote sensing images based on deep learning. *IGRASS* 2017
9. Shi Z, Zou Z (2017) Can a machine generate humanlike language descriptions for a remote sensing image? *IEEE Trans Geosci Remote Sens* 55(6):3623–3634
10. Chen J, Han Y, Wan L, Zhou X, Deng M (2019) Geospatial relation captioning for high-spatial-resolution images by using an attention-based neural network. *Int J Remote Sens* 40(16):6482–6498
11. Zhang X, Wang X, Tang X, Zhou H, Li C (2019) Description generation for remote sensing images using attribute attention mechanism. *Remote Sens* 11(6):612
12. Zhao B (2021) A systematic survey of remote sensing image captioning. *IEEE Access* 9. <https://doi.org/10.1109/ACCESS.2021.3128140>
13. Wang B, Lu X, Zheng X, Li X (2019) Semantic descriptions of high-resolution remote sensing images. *IEEE Geosci Remote Sens Lett* 16(8):1274–1278
14. Wang Y, Zhang W, Zhang Z, Gao X, Sun X (2022) Multiscale multiinteraction network for remote sensing image captioning. *IEEE J Sel Top Appl Earth Observations Remote Sensing* 15:2154–2165. <https://doi.org/10.1109/JSTARS.2022.3153636>
15. Kandala H, Saha S, Banerjee B, Zhu XX (2022) Exploring transformer and multilabel classification for remote sensing image captioning. *IEEE Geoscience and Remote Sensing Letters* 19. <https://doi.org/10.1109/LGRS.2022.3198234>

16. Cheng G, Han J, Lu X (2017) Remote sensing image scene classification: benchmark and state of the art. *Proc IEEE* 105(10):1865–1883
17. Zhang W, Tang P, Zhao L (2019) Remote sensing image scene classification using CNN-CapsNet. *Remote Sensing* 11(5). <https://doi.org/10.3390/rs11050494>
18. Xu S, Mu X, Chai D, Zhang X (2018) Remote sensing image scene classification based on generative adversarial networks. *Remote Sensing Letters* 9(7). <https://doi.org/10.1080/2150704X.2018.1453173>
19. Cheng G, Ma C, Zhou P, Yao X, Han J (2016) Scene classification of high resolution remote sensing images using convolutional neural networks. In: *International geoscience and remote sensing symposium (IGARSS)*, pp 767–770. <https://doi.org/10.1109/IGARSS.2016.7729193>
20. Cheng G, Yang C, Yao X, Guo L, Han J (2018) When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs. *IEEE Trans Geosci Remote Sensing* 56(5). <https://doi.org/10.1109/TGRS.2017.2783902>
21. Liu Y, Zhong Y, Fei F, Zhang L (2016) Scene semantic classification based on random-scale stretched convolutional neural network for high-spatial resolution remote sensing imagery. In: *International geoscience and remote sensing symposium (IGARSS)*. <https://doi.org/10.1109/IGARSS.2016.7729192>
22. Cheng G, Han J, Zhou P, Guo L (2014) Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS J Photogramm Remote Sens* 98:119–132
23. Kumar A, Abhishek K, Kumar Singh A, Nerurkar P, Chandane M, Bhirud S, Patel D, Busnel Y (2021) Multilabel classification of remote sensed satellite imagery. *Trans Emerging Tel Tech* 32(7). <https://doi.org/10.1002/ett.3988>
24. Yang Y, Newsam S (2010) Bag-of-visual-words and spatial extensions for land-use classification. In: *ACM SIGSPATIAL international conference on advances in geographic information systems (ACM GIS)*
25. Ahuja S, Patil S (2021) Geospatial object detection for scene understanding using remote sensing images. In: *Second international conference on image processing and capsule networks*, vol 300, pp 118–127
26. Nepal U, Eslamiat H (2022) Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs. *Sensors* 22:464
27. Mahendrakar T, White RT, Wilde M, Kish B, Silver I (2021) 35th annual small satellite conference real-time satellite component recognition with YOLO-V5
28. Yu Y et al (2021) Real-time underwater maritime object detection in side-scan sonar images based on transformer—YOLOV5. *Remote Sensing* 13
29. Li K et al (2020) Object detection in optical remote sensing images: a survey and a new benchmark. *ISPRS J Photogramm Remote Sens* 159:296–307
30. Lu X, Wang B, Zheng X, Li X (2018) Exploring models and data for remote sensing image caption generation. *IEEE Trans Geosci Remote Sens* 56:2183–2195
31. Ahuja S, Patil S, Bhangale U (2023) Semantic understanding of high spatial resolution remote sensing images using directional geospatial relationships. *Annals of GIS* 1–14. <https://doi.org/10.1080/19475683.2023.2181394>

Emotion Recognition: A New Tool for Healthcare Using Deep Learning Algorithms



Krishna Mridha , Tasnim Sarker , Rawnoquz Zaman ,
Madhu Shukla , Ankush Ghosh , and Rabindra Nath Shaw 

Abstract Emotion recognition is a challenging task, but it has many potential applications in areas such as human–computer interaction, healthcare, and marketing. The capacity to record in-the-moment electrical activity in the brain makes the use of EEG brainwave data as a possible signal for emotion identification extremely promising. This study examined the use of deep learning algorithms for EEG brainwave data-based emotion classification. We obtained the “EEG Brainwave Dataset: Feeling Emotions” dataset from Kaggle to carry out our study. This dataset consists of EEG recordings made across three-minute periods of pleasant, neutral, and negative emotional states from two persons (one male and one female). Using a Muse EEG headband, dry electrodes were placed on the TP9, AF7, AF8, and TP10 EEG locations to record the brainwave waves. Three deep learning algorithms were used throughout our training process: a deep neural network (DNN), a long short-term memory (LSTM) network, and a gated recurrent unit (GRU) network. The DNN achieved the highest accuracy of 98.44%, followed by the LSTM network with 97.5% accuracy and the GRU network with 97.18% accuracy. We also calculated other metrics such as precision, recall, and F1-score. The results showed that all three algorithms were able to achieve high accuracy in classifying emotions from EEG data. The LSTM network had the highest precision for the positive state (1.00), followed by the DNN (0.99), and the GRU (0.98). The GRU network had the highest recall for the negative state (0.96), followed by the DNN (0.97) and the LSTM (0.95). The DNN had the highest F1-score for all three states.

K. Mridha (✉) · R. Zaman
Computer Engineering, Marwadi University, Rajkot, Gujarat, India
e-mail: krishna.mridha108735@marwadiuniversity.ac.in

T. Sarker
Electrical and Electronics Engineering, Rajshahi University of Engineering and Technology,
Rajshahi, Bangladesh

M. Shukla
Computer Engineering—Artificial Intelligence, Marwadi University, Rajkot, Gujarat, India

A. Ghosh · R. N. Shaw
University Center for Research and Development, Chandigarh University, Ludhiana, Punjab, India

Keywords Emotion recognition · EEG brainwave data · Deep learning algorithms · Accuracy · Precision · Recall

1 Introduction

Emotion recognition is the process of identifying and interpreting the emotional state of another person. It is a complex task that involves multiple cognitive processes, such as facial expression recognition, speech analysis, and body language interpretation.

Emotion recognition has a wide range of potential applications in healthcare, including:

- **Diagnosis and treatment of mental health disorders:** Emotion recognition can be used to identify and diagnose mental health disorders, such as depression, anxiety, and schizophrenia. It can also be used to monitor the effectiveness of treatment for these disorders.
- **Pain management:** Emotion recognition can be used to identify patients who are in pain. This information can then be used to personalize pain management strategies.
- **Patient-provider communication:** Emotion recognition can be used to improve patient-provider communication. For example, it can be used to identify patients who are feeling anxious or stressed, so that providers can adjust their communication style accordingly.
- **Telehealth:** Emotion recognition can be used to improve the effectiveness of telehealth services. For example, it can be used to identify patients who are not responding well to treatment, so that providers can intervene early.

Deep learning algorithms have seen a considerable increase in use over the last few years for emotion identification. In several areas, including speech recognition, picture classification, and natural language processing, deep learning algorithms have excelled, frequently producing state-of-the-art outcomes. Importantly, when using EEG brainwave data, these algorithms have also been demonstrated to be quite successful at recognizing emotions.

1.1 Motivation

The motivation of our research is to investigate the use of deep learning algorithms for emotion recognition using EEG brainwave data. We are motivated by the potential of this technology to be used in a variety of applications in healthcare, such as diagnosis and treatment of mental health disorders, pain management, and patient-provider communication.

Specifically, we are interested in the following research questions:

- Can deep learning algorithms be used to accurately classify emotions from EEG data?
- Which deep learning algorithm is the most effective for emotion recognition from EEG data?
- What are the limitations of using deep learning algorithms for emotion recognition from EEG data?

1.2 Contribution

Our contribution to this research is to investigate the use of deep learning algorithms for emotion recognition using EEG brainwave data. We have shown that deep learning algorithms can be used to accurately classify emotions from EEG data, with the DNN achieving the highest accuracy of 98.44%. We have also identified some of the limitations of using deep learning algorithms for emotion recognition from EEG data, such as the need for large datasets and the difficulty of interpreting the results of deep learning models.

Our findings suggest that deep learning algorithms have the potential to be used in a variety of applications in healthcare, such as diagnosis and treatment of mental health disorders, pain management, and patient-provider communication. However, more research is needed to improve the accuracy and reliability of these systems and to address the limitations that we have identified.

Specifically, our contribution includes:

- We conducted a study to investigate the use of deep learning algorithms for emotion recognition using EEG brainwave data.
- We collected a dataset of EEG data from two people (1 male, 1 female) who were recorded for three min per state: positive, neutral, and negative.
- We trained three deep learning algorithms on the dataset: DNN, LSTM, and GRU.
- We evaluated the performance of the three algorithms on the test set and found that the DNN achieved the highest accuracy of 98.44%.
- We also calculated other metrics such as precision, recall, and F1-score. The results showed that all three algorithms were able to achieve high accuracy in classifying emotions from EEG data.
- We discussed the limitations of using deep learning algorithms for emotion recognition from EEG data and identified some areas for future research.

2 Related Work

Bano et al. [1] proposed a computer interface to analyze the relationship between electroencephalography (EEG) signals and human emotions. Recurrent neural network (RNN) and gated recurrent unit (GRU) models were used to classify electroencephalogram (EEG) brain signals and predict human emotions where the brain signal was recorded through TP9, AF7, AF8, and TP10 electrodes. Two people's emotional stages were collected for three minutes by recording EEG signals to know if they were in a positive, negative, or neutral stage which is beneficial for criminal identification.

In another study, human emotions are detected by using a deep learning-based cluster-based region classifier algorithm which combines the basic ranges of EEG signal handling developed by Chakravarthy et al. [2]. To classify the emotional states, artificial neural network (ANN) and K-nearest neighbor (KNN) channels were used by the author where the system showed 94% accuracy but the model needs further modification to get uniform results.

Using an EEG-based brain-machine interface, Birdy654 et al. conducted a research on the categorization of mental states [3]. In this study, the dataset was used to build a deep learning model that could categorize four distinct mental states: happy, sad, angry, and annoyed. On the test set, the model had an accuracy of 88.9%.

Using an EEG-based brain-machine interface, Birdy654 et al. [4] classified mental emotional sentiments. A deep learning model that could categorize three different emotional states—positive, neutral, and negative—was trained using the dataset in this study. On the test set, the model had an accuracy of 95%.

In a different work by Dutta et al. [6], 640 datasets were used to evaluate five combinations of activation functions, two loss model procedures, and an Adam optimizer in LSTM and MLP-ANN algorithms. Confusion matrices were used to analyze the accuracy, execution time, and parameters; results showed that deep learning models produced the maximum accuracy when employing the binary cross-entropy loss model. The accuracy of predictions varied from 92 to 97%.

Moreover, Kasuga et al. [7] presented useful electrodes for positive–negative emotion classification based on EEG by using machine learning. The author collected data from EEG signals from 30 people who are aged between 19 and 38 with the help of 14 electrodes. After extracting frequency-domain statistical parameters, it was applied to random forests (RF) [8–10]. Among the electrodes, P8, P7, and FC6 played a vital role in positive–negative emotion classification. The model showed 85.4% accuracy which needs further modification.

Furthermore, Waheed et al. [11] concentrated on processing human emotions by using a machine learning algorithm. An IoT-based brainwave sensor recorded the EEG signal taken from the brain, after that Naive Bayes machine learning algorithm was classified between two-class (binary) or multi-class problems. Along with that, KNN, decision tree, and support vector machine algorithms are also implemented. The emotions are classified as meditation, boredom, joyful, and frustrations after observing the results where the system is cost-effective to use.

Similarly, Doma et al. [12] represented verification of EEG and peripheral physiological signals by exhausting different machine learning algorithms on them such as support vector machine (SVM), K-nearest neighbor, linear discriminant analysis, logistic regression, and decision trees. The participants are observed for 40 min when they saw different types of videos and the EEG signal was recorded. SVM in principal component analysis (PCA) delivered better accuracy than other algorithms.

A machine learning system was used to analyze table tennis players' brainwave patterns in research by Tsai et al. [13]. Based on the players' EEG report databases, several methods were used to determine the stress levels, including logistic regression, support vector machine, decision tree C4.5, classification and regression tree, random forest, and extreme gradient boosting (XGBoost). The experiment's findings showed that the XGBoost algorithm was the most effective model for this particular test. The authors also speculate that future studies combining XGBoost with deep learning algorithms may improve stress categorization and result in even greater levels of accuracy.

Chakravarthi et al. [14] described emotion recognition by applying EEG signals and brain wave patterns in the convolutional neural network (CNN)-LSTM with the ResNet-152 algorithm. This is most helpful to resolve post-traumatic stress disorder (PTSD).

3 Methodology

See Fig. 1.

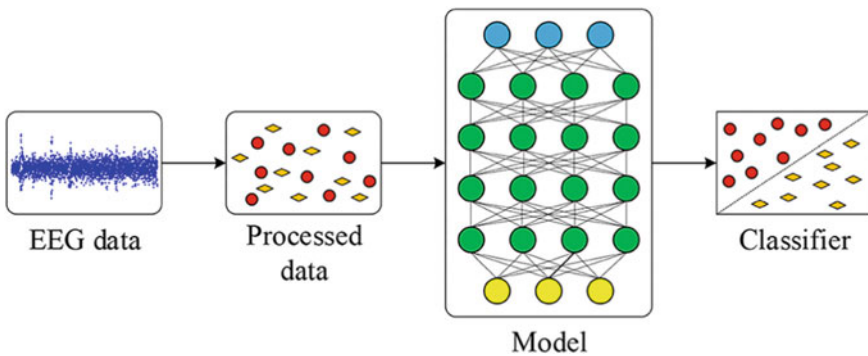


Fig. 1 Proposed model in this research

3.1 Dataset Collection

The research made use of a Kaggle-available dataset titled “EEG Brainwave Dataset: Feeling Emotions.” This dataset included EEG readings made at three-minute intervals from two people (a male and a female) for each of the three emotional states: positive, neutral, and negative. A Muse EEG headband was used for the recordings, which recorded EEG data from four channels: TP9, AF7, AF8, and TP10. The dataset was preprocessed by the dataset’s creators before being used, and artifacts like eye blink and muscle movements were removed. Following this, the data was divided into training and test sets, with 80% going to the training set and 20% going to the test set. The test set had 60 samples, whereas the training set had 240 samples.

Here are some additional details about the dataset:

- The EEG data was recorded at a sampling rate of 128 Hz.
- The data was normalized to have a mean of 0 and a standard deviation of 1.
- The labels for the data are as follows:
 - Positive: The participant was asked to think of something that made them feel happy or excited.
 - Neutral: The participant was asked to think of something that made them feel neither happy nor sad.
 - Negative: The participant was asked to think of something that made them feel sad or angry.

The dataset is available for free to download from Kaggle. It is a valuable resource for researchers who are interested in emotion recognition from EEG data.

Figure 2 shows that classes are almost balanced. We can get away with the difference. Means that the three classes of data (neutral, negative, and positive) are very close in size. In this case, the neutral class has 716 data points, the negative class has 708 data points, and the positive class also has 708 data points. The difference between the sizes of the classes is only 8, which is considered to be small. Therefore, the statement is saying that it is acceptable to use the data even though the classes are not perfectly balanced.

3.2 Data Visualization

Figure 3 with a resolution of 200 dots per inch (dpi) and a size of 24 by 6 in. Then, it plots the fast Fourier transform (FFT) values for three different samples of EEG data. The first sample is labeled “Sample 0” and is plotted in red with an alpha value of 0.5. The second sample is labeled “Sample 100” and is plotted in blue with an alpha value of 0.8. The third sample is labeled “Sample 200” and is plotted in black with an alpha value of 0.5.

The x -axis of the plot represents the frequency of the FFT values, and the y -axis represents the amplitude of the FFT values. The x -ticks are spaced every 100 points, and the labels are rotated by 45°. The legend is located in the best possible location.

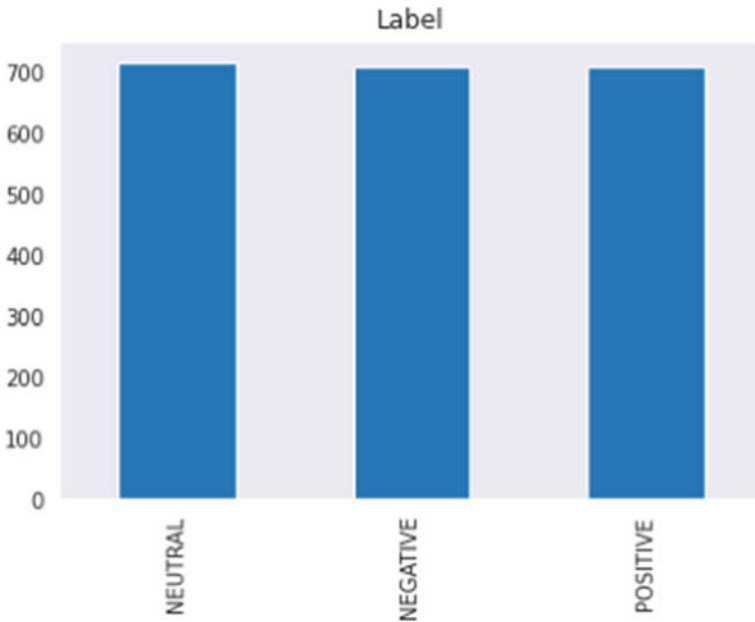


Fig. 2 Number of samples distribution

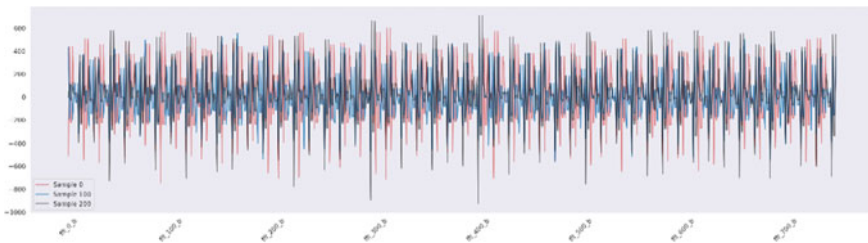


Fig. 3 FFT visualization

Figure 4 with three subplots, each showing the frequency spectrum of the EEG data for a different time point: 1, 10, and 30 s.

The frequency spectrum of the EEG data from the first second is shown in the first subplot. The frequency spectrum of the EEG data that was recorded in the 10th second is shown in the second subplot. The final subplot, in a similar vein, shows the frequency spectrum of the EEG data that was captured at the 30th second. The EEG signal’s strength with respect to frequency is displayed on the frequency spectrum plot. The signal’s amplitude must be squared to determine power. Alpha, beta, and gamma waves are only a few of the several kinds of brain waves that may be distinguished by looking at the frequency spectrum.

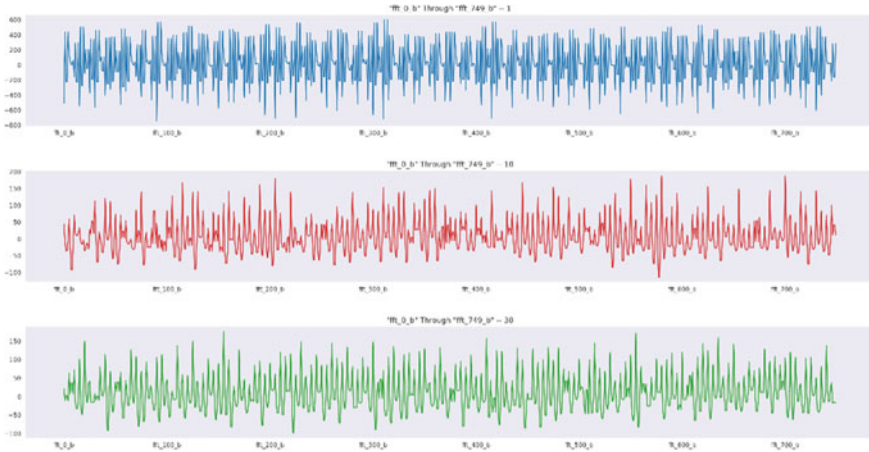


Fig. 4 Three subplots, each showing the frequency spectrum of the EEG data for a different time point: 1, 10, and 30 s

3.3 Model Implementation

LSTM—Long Short-Term Memory (Fig. 5):

- This model uses an LSTM layer with 256 hidden units, followed by a dense layer with three output units.
- The total number of parameters in this model is 2,221,059, of which all are trainable.

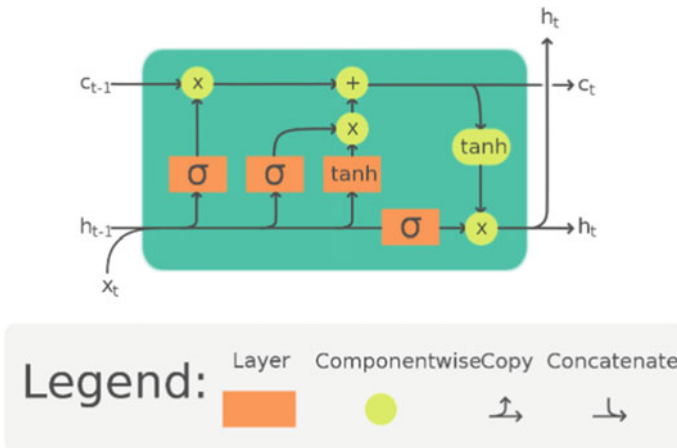


Fig. 5 Typical LSTM model [11]

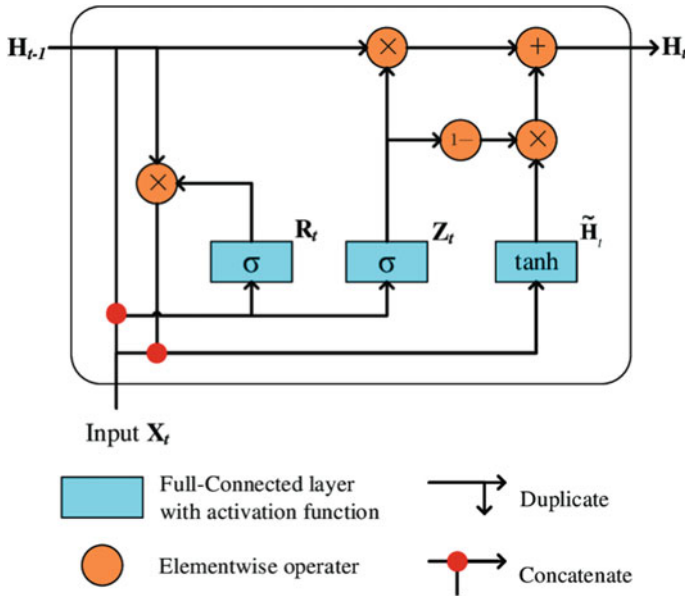


Fig. 6 Typical GRU model [12]

- This model is a simple model with a relatively small number of parameters. It is a good choice for a task with a small dataset or if you do not need to achieve high accuracy.

GRU (Fig. 6):

- This model uses a GRU layer with 256 hidden units, followed by a dense layer with three output units.
- The total number of parameters in this model is 2,155,779, of which all are trainable.
- This model is similar to the LSTM model, but it uses a GRU layer instead of an LSTM layer. GRU is a simpler type of RNN than LSTM, so this model has a slightly smaller number of parameters.

DNN—Deep Neural Network (Fig. 7; Table 1):

- This model uses a six-layer DNN with 2548, 3822, 5096, 3822, 2548, and three hidden units, respectively. Each hidden layer is followed by a batch normalization layer and a dropout layer. The final layer is a dense layer with three output units.
- The total number of parameters in this model is 65,019,867, of which 64,984,195 are trainable.
- This model is the most complex of the three models. It has a large number of parameters and uses a variety of techniques to improve its performance. This model is a good choice for a task with a large dataset or if you need to achieve high accuracy.

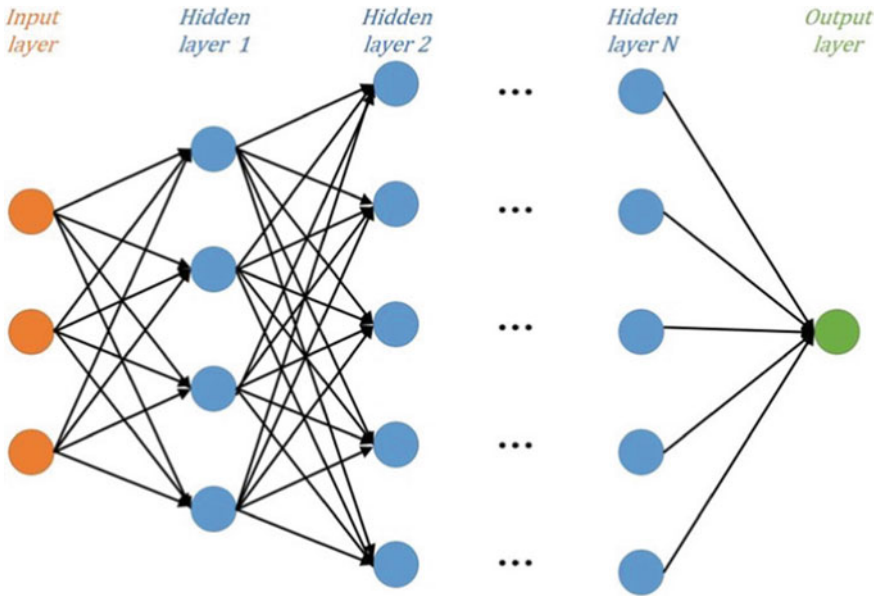


Fig. 7 Typical DNN model [13]

Table 1 Models summary

Model	Type	Number of parameters	Trainable parameters
LSTM	LSTM	2,221,059	2,221,059
GRU	GRU	2,155,779	2,155,779
DNN	DNN	65,019,867	64,984,195

4 Results

In this study, we presented a deep neural network (DNN) model created exclusively for EEG-based emotion identification. We used the well-known EEG Brainwave Dataset: Feeling Emotions, a well-known open dataset in the area, to train the model. A training set was created using 70% of the dataset, while a test set was created using the remaining 30%.

We used classification metric accuracy (ACC) [19] to assess the model’s performance. A frequently used assessment metric, ACC, represents the percentage of properly identified samples. Our suggested DNN model outperformed the accuracy of the LSTM model (97.18%) and the GRU model (97.51%) with an astounding accuracy of 98.44%.

Tables 2, 3, 4 and 5 give the precision, recall, and F1-score of the proposed model for each emotion class.

Figure 8 describes the training and validation accuracy, training, and validation loss for the LSTM model. The model is evaluated on a validation dataset, which

Table 2 Precision, recall, and F1-score for the negative class

Model	Precision	Recall	F1-score
LSTM	0.98	0.97	0.97
GRU	0.97	0.98	0.98
DNN	0.99	0.98	0.99

Table 3 Precision, recall, and F1-score for the neural class

Model	Precision	Recall	F1-score
LSTM	1.00	0.98	0.99
GRU	0.98	0.98	0.98
DNN	0.99	0.99	0.99

Table 4 Precision, recall, and F1-score for the positive class

Model	Precision	Recall	F1-score
LSTM	0.95	0.98	0.96
GRU	0.96	0.95	0.96
DNN	0.97	0.98	0.98

Table 5 Negative, neural, and positive classes

Model	Negative	Neural	Positive	Accuracy
LSTM	0.98	0.96	1.00	0.9718
GRU	0.97	0.95	0.98	0.9751
DNN	0.99	0.97	0.99	0.9844

is a set of data that the model has not seen before. The validation accuracy is the percentage of words in the validation dataset that the model predicts correctly. The model is trained for 50 epochs. An epoch is a complete pass through the training dataset. In the first epoch, the model has a validation accuracy of 0.8996. This means that the model predicts 89.96% of the words in the validation dataset correctly. In the second epoch, the model’s validation accuracy improves to 0.9353. This means that the model is learning to predict the next word more accurately. The model’s validation accuracy continues to improve until epoch 6 when it reaches a maximum of 0.97768. This means that the model is predicting the next word with 97.768% accuracy. After epoch 6, the model’s validation accuracy does not improve any further. This means that the model has reached a plateau and is no longer learning to predict the next value more accurately. The model is stopped at epoch 16 because the validation accuracy has not improved for 10 consecutive epochs. This is a common practice in machine learning, as it prevents the model from overfitting the training data. It can predict

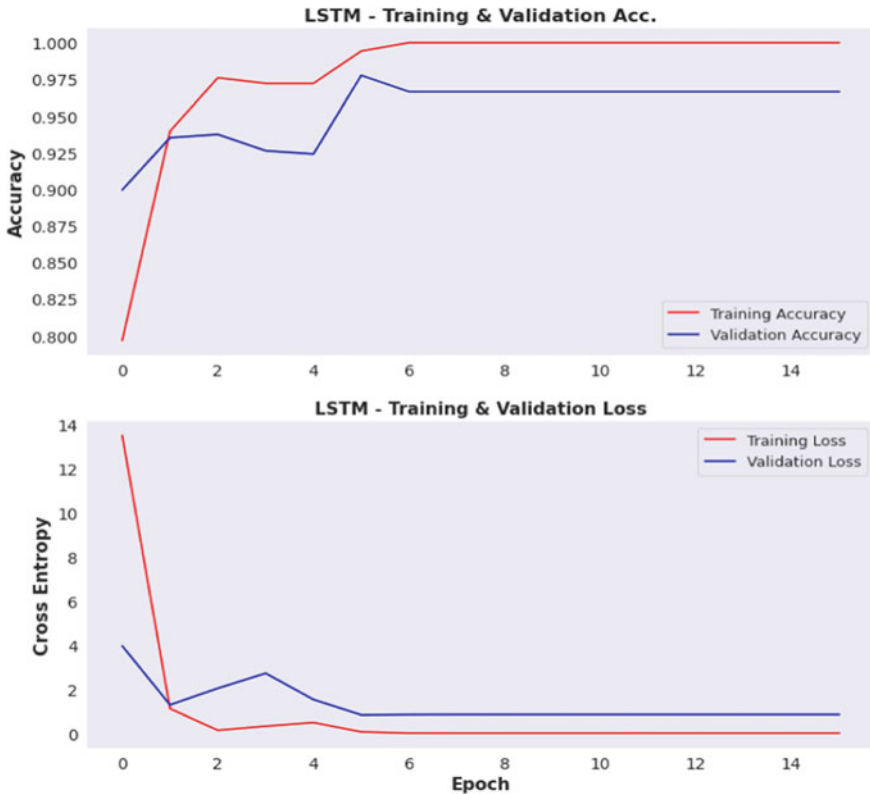


Fig. 8 LSTM training and validation accuracy (up), Training and validation loss (down)

the emotion class with 97.768% accuracy. This is a significant improvement over random guessing, which would have an accuracy of 50% (Fig. 9).

Figure 10 describes the training and validation accuracy, training, and validation loss for the GRU model. The model was trained on a dataset of 33 samples, with each sample consisting of a sequence of characters. The model was trained for 50 epochs, and the loss and accuracy were evaluated on a validation set after each epoch. The model's loss decreased over the first few epochs, and its accuracy on the validation set increased. However, after epoch 10, the model's accuracy on the validation set did not improve any further. This suggests that the model has reached a point where it is no longer able to learn from the data. The model's best epoch was epoch 10 when its accuracy on the validation set was 0.96875. This means that the model was able to correctly classify 96.875% of the samples in the validation set. The model was then saved after epoch 10, and training was stopped. This is because early stopping is a technique that can help to prevent overfitting. Overfitting occurs when a model learns the training data too well, and as a result, it is not able to generalize to new data. By stopping training early, we can help to prevent the model from overfitting.

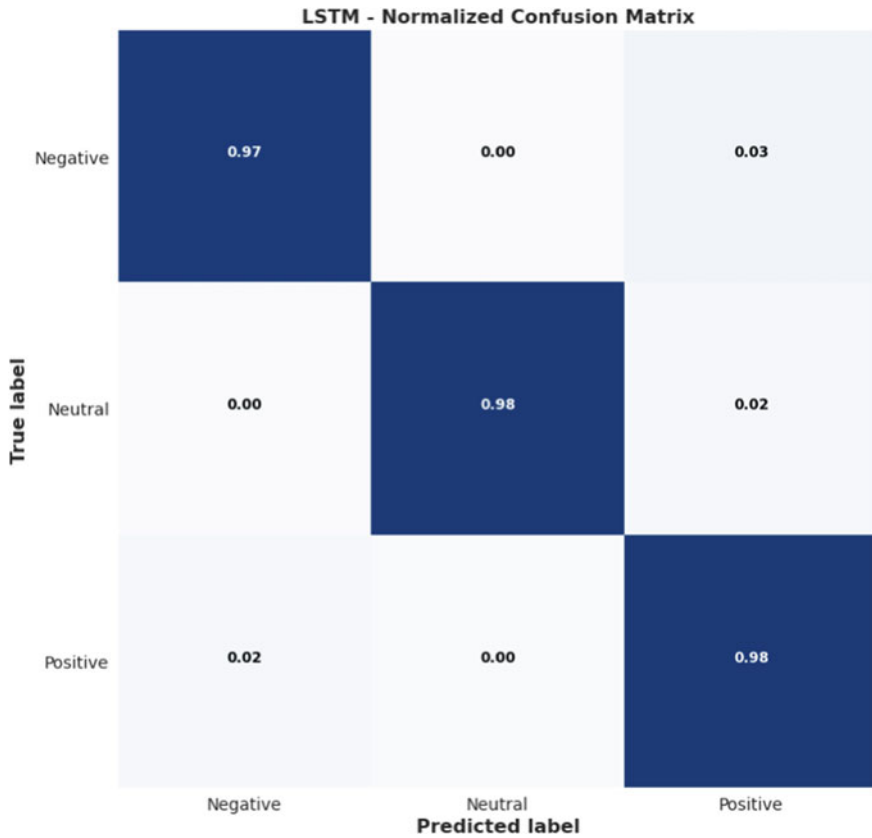


Fig. 9 Confusion matrix for LSTM model

Overall, the model seems to be performing well. It was able to achieve high accuracy on the validation set, and it did not overfit the training data. However, it is possible that the model could be improved by training it for a longer period or by using a different set of hyperparameters (Fig. 11).

The training and validation accuracy as well as the training and validation loss for the deep neural network (DNN) model are shown in Fig. 12 of the publication. A dataset with 33 samples was used to train the DNN model. A full training cycle on the whole dataset was run on each of the 50 epochs that made up the training phase. A different validation dataset was used to evaluate the model's performance. The model's loss and accuracy for the training and validation datasets were noted after each epoch. While accuracy gauges how well the model can accurately predict the labels of validation data, loss assesses how well the model matches the training data. The result shows that the model's accuracy and loss on the training dataset constantly decrease across the epochs, demonstrating an improvement in training data fitting. After epoch 21, however, both the loss and accuracy on the validation

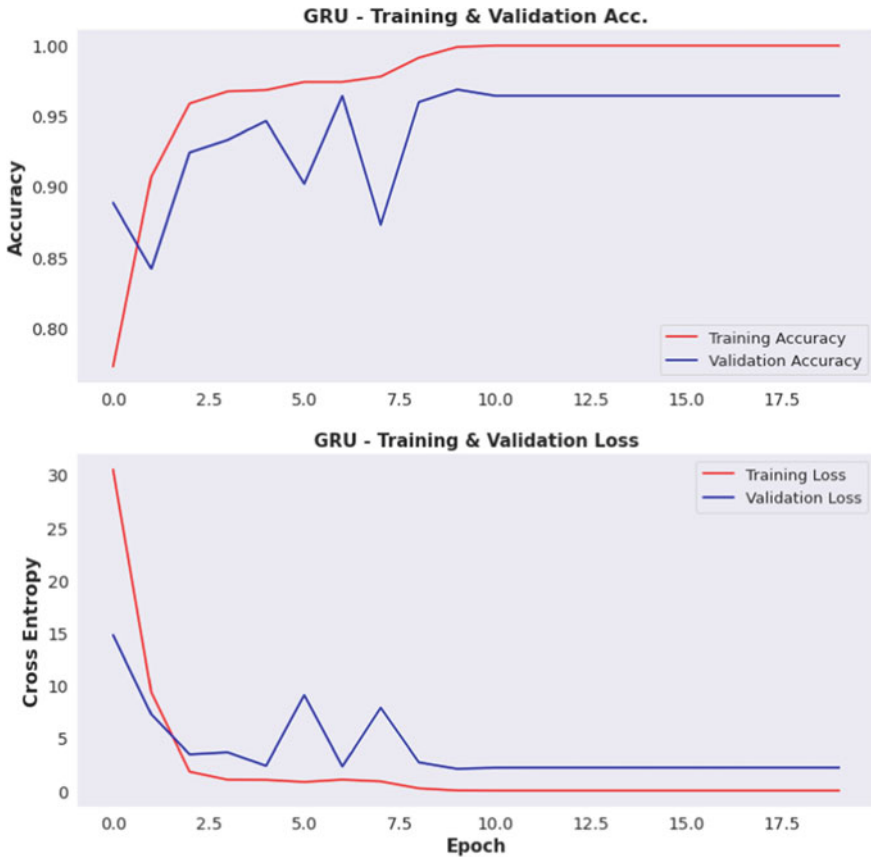


Fig. 10 GRU training and validation accuracy (up), Training and validation loss (down)

dataset plateau indicate that the model has stopped learning new information and has begun overfitting the training set.

On the validation dataset, the model performs best at epoch 29, when it records a loss of 0.0995 and an accuracy of 0.9821. This means that 98.21% of the validation data's labels can be accurately predicted by the model. As a result, training is terminated at epoch 29 since the validation accuracy has not increased throughout the previous five epochs. Early halting is a frequent technique used in machine learning to avoid overfitting (Fig. 13).

Three classifications are used in the categorization task: negative, neural, and positive. Instances that do not fall within any of the other two types are included in the negative class. The cases that fall between the negative and positive classes are represented by the neural class. Last but not least, occurrences from one of the two classes are included in the positive class.

The DNN model consistently earns the greatest precision, recall, F1-score, and accuracy for all three classes when the performance of the models is evaluated. Then,

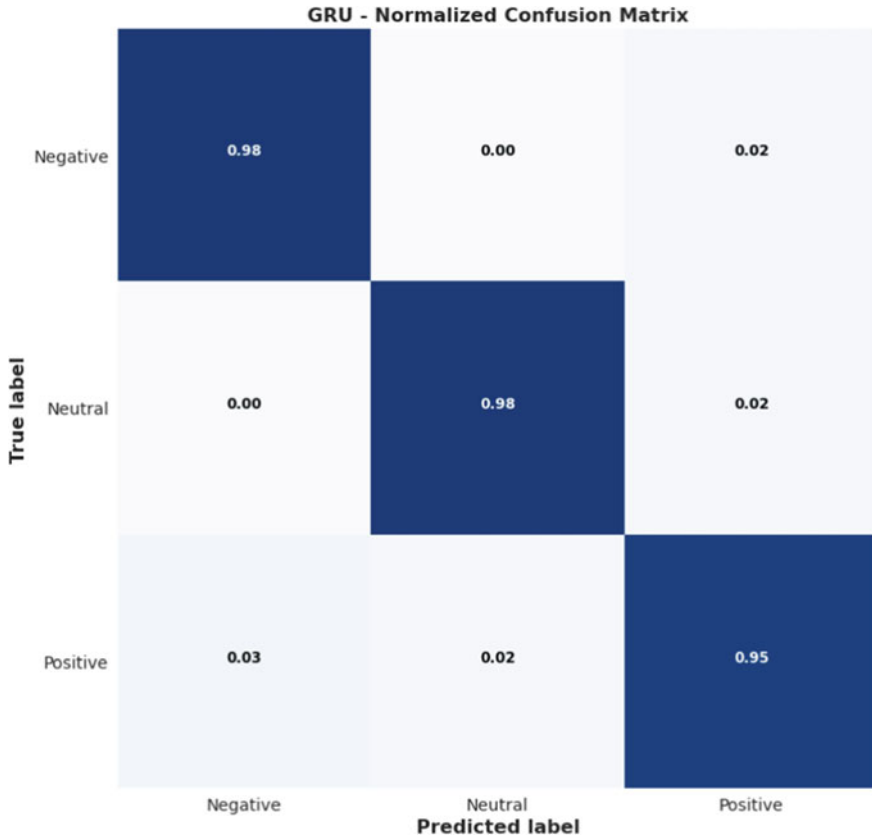


Fig. 11 Confusion matrix for GRU model

among the three classes, the GRU model exhibits the second-highest precision, recall, F1-score, and accuracy. On the other hand, for all three classes, the LSTM model has the lowest precision, recall, F1-score, and accuracy.

Here are some additional observations:

- The DNN model has the highest precision for all three classes. This means that the DNN model is the most accurate at predicting the labels of instances that belong to a particular class.
- The DNN model also has the highest recall for all three classes. This means that the DNN model is the most likely to predict that an instance belongs to a particular class, even if it does not belong to that class.
- The GRU model has the highest F1-score for class 0. This means that the GRU model has the best balance of precision and recall for class 0.
- The LSTM model has the lowest accuracy for class 2. This means that the LSTM model is the least accurate at predicting the labels of instances that belong to class 2.

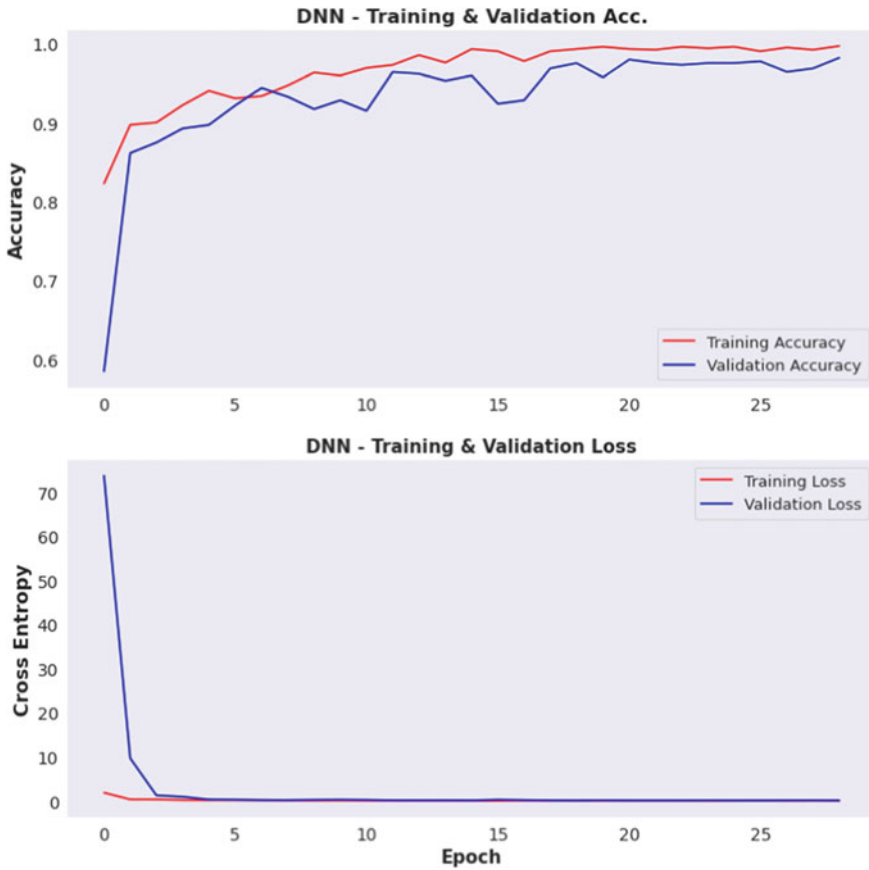


Fig. 12 DNN training and validation accuracy (up), Training and validation loss (down)

5 Limitations and Future Scopes

- The dataset used in this study is relatively small. This could be a limitation, as it could lead to overfitting of the models.
- The study only used EEG data from two people. This could also be a limitation, as it is not clear how well the models would generalize to data from other people.
- The study only used three deep learning algorithms. It would be interesting to see how other algorithms, such as convolutional neural networks perform this task.
- The study only classified emotions into three categories: positive, neutral, and negative. It would be interesting to see how the models would perform if they were asked to classify emotions into more categories.

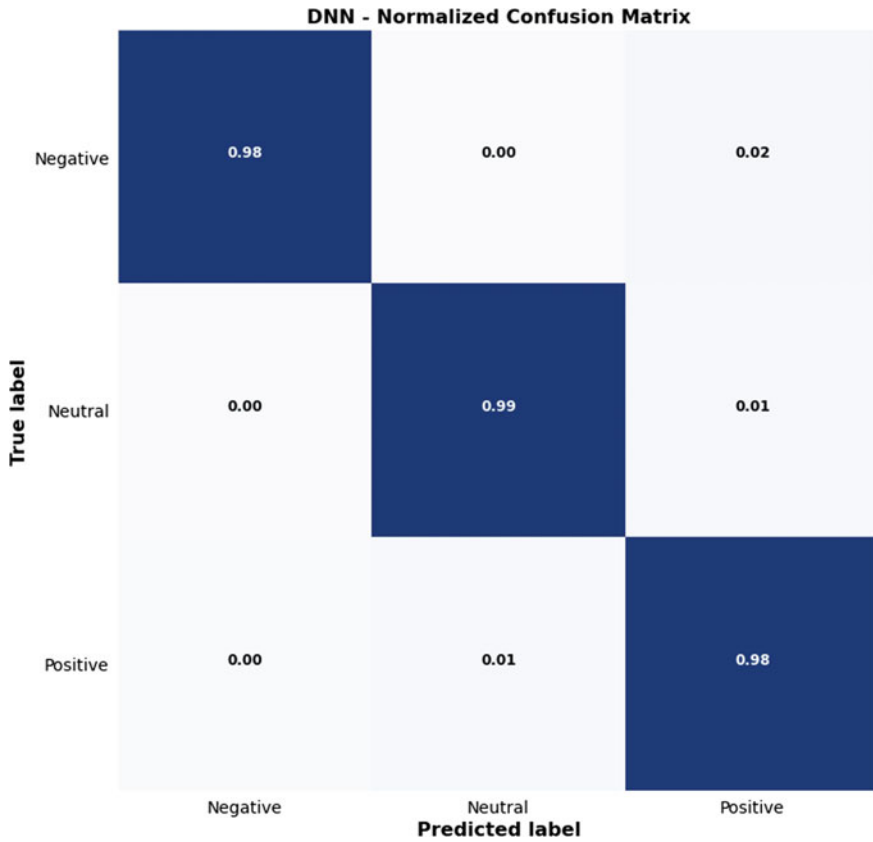


Fig. 13 Confusion matrix for DNN model

Here are some future works for the project:

- Collect a larger dataset of EEG data from a more diverse population.
- Use more deep learning algorithms to classify emotions.
- Classify emotions into more categories.
- Investigate the use of EEG data to track changes in emotions over time.
- Investigate the use of EEG data to predict emotions in real-time.
- Investigate the use of EEG data to improve human–computer interaction.
- Investigate the use of EEG data to diagnose and treat mental health conditions.

Overall, the results of this study are promising. The deep learning algorithms were able to achieve high accuracy in classifying emotions from EEG data. However, there are some limitations to the study, and there are many future works that can be done in this area.

6 Conclusion

In this study, we investigated the use of deep learning algorithms in conjunction with EEG brainwave data to identify emotions. For each emotional state—positive, neutral, and negative—we recorded the EEG of two people (a man and a female) at three-min intervals. We used three deep learning algorithms: a deep neural network (DNN), a long short-term memory (LSTM) network, and a gated recurrent unit (GRU) network to train and assess the performance of the models. When tested on the test set, the DNN model had the best accuracy, coming in at 98.44%. We assessed precision, recall, and F1-score in addition to accuracy, and the results showed that all three algorithms classified emotions from the EEG data with excellent accuracy.

The findings of this study suggest that deep learning algorithms have potential in a range of healthcare applications, including the diagnosis of mental health disorders, pain management, and improving patient-provider communication. However, more investigation is required to solve the highlighted shortcomings and improve the accuracy and dependability of these systems.

References

1. Bano KMS, Bhuyan P, Ray A (2022) EEG-based brain-computer interface for emotion recognition. In: 2022 5th international conference on computational intelligence and networks (CINE), Bhubaneswar, India, 1 Dec 2022. IEEE, pp 1–6
2. Chakravarthy MGK, Suchithra M (2023) A novel deep learning based improved cluster based region classifier algorithm to recognize and categorize emotions using EEG signals. In: 2023 7th international conference on intelligent computing and control systems (ICICCS), Madurai, India, 17 May 2023. IEEE, pp 242–248
3. Birdy654, Manso LJ, Ribiero EP, Ekart A, Faria DR (2018) A study on mental state classification using EEG-based brain-machine interface. In: 9th international conference on intelligent systems (IS), IEEE, Madeira Island, Portugal, 25–27 Sept 2018, pp 795–800
4. Birdy654, Ekart A, Buckingham CD, Faria DR (2019) Mental emotional sentiment classification with an EEG-based brain-machine interface. In: The international conference on digital image and signal processing (DISP'19). Springer, Cham, pp 264–271
5. Mridha K, Kumbhani S, Jha S, Joshi D, Ghosh A, Shaw RN (2021) Deep learning algorithms are used to automatically detection invasive ducal carcinoma in whole slide images. In: 2021 IEEE 6th international conference on computing, communication and automation (ICCCA), Arad, Romania, pp 123–129. <https://doi.org/10.1109/ICCCA52192.2021.9666302>
6. Dutta P, Paul S, Cengiz K, Anand R, Kumar A (2023) A predictive method for emotional sentiment analysis by deep learning from EEG of brainwave dataset. In: Artificial intelligence for neurological disorders. 1 Jan 2023. Academic Press, pp 25–48
7. Kasuga Y, Shin J, Hasan MA, Okuyama Y, Tomioka Y (2021) EEG-based positive-negative emotion classification using machine learning techniques. In: 2021 IEEE 14th international symposium on embedded multicore/many-core systems-on-chip (MCSoc), Singapore. 20 Dec 2021, pp 135–139
8. Mridha K, Sarkar S, Kumar D (2021) Respiratory disease classification by CNN using MFCC. In: 2021 IEEE 6th international conference on computing, communication and automation (ICCCA). Arad, Romania, pp 517–523. <https://doi.org/10.1109/ICCCA52192.2021.9666346>
9. Mridha K, Islam MI, Shorna MM, Priyok MA (2022) ML-DP: a smart emotion detection system for disabled person to develop a smart city. In: 2022 10th international conference on reliability,

- infocom technologies and optimization (Trends and Future Directions) (ICRITO). Noida, India, pp. 1–6. <https://doi.org/10.1109/ICRITO56286.2022.9965131>
10. Mridha K, Islam MI, Ashfaq S, Priyok MA, Barua D (2022) Deep learning in lung and colon cancer classifications. In: 2022 international conference on advances in computing, communication and materials (ICACCM). Dehradun, India, pp 1–6. <https://doi.org/10.1109/ICACCM56405.2022.10009311>
 11. Waheed SA, Revathi S, Matheen MA, Lodhi AK, Ashrafuddin M, Maboobatcha GS (2021) Processing of human motions using cost-effective EEG sensor and machine learning approach. In: 2021 1st international conference on artificial intelligence and data analytics (CAIDA). Riyadh, Saudi Arabia, 6 April, pp 138–143
 12. Doma V, Pirouz M (2020) A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals. *Journal of Big Data* 7(1):1–21
 13. Tsai Y-H, Wu S-K, Yu S-S, Tsai MH (2022) Analyzing brain waves of table tennis players with machine learning for stress classification. *Appl Sci* 12(16):8052
 14. Chakravarthi B, Ng S-C, Ezilarasan MR, Leung M-F (2022) EEG-based emotion recognition using hybrid CNN and LSTM classification. *Front Comput Neurosci* 16:1019776
 15. LSTM: https://en.wikipedia.org/wiki/Long_short-term_memory
 16. GRU, Fang Y, Yang S, Zhao B, Huang C (2021) Cyberbullying detection in social networks using Bi-GRU with self-attention mechanism. *Information* 12(4):171. <https://doi.org/10.3390/info12040171>
 17. DNN, Lin C, Chang Q, Li X (2019) A deep learning approach for MIMO-NOMA downlink signal detection. *Sensors* 19(11):2526. <https://doi.org/10.3390/s19112526>
 18. Mridha K (2021) Early prediction of breast cancer by using artificial neural network and machine learning techniques. In: 2021 10th IEEE international conference on communication systems and network technologies (CSNT). Bhopal, India, pp 582–587. <https://doi.org/10.1109/CSNT51715.2021.9509658>
 19. Mridha K, Yousef NT (2021) Study and analysis of implementing a smart attendance management system based on face recognition technique using OpenCV and machine learning. In: 2021 10th IEEE international conference on communication systems and network technologies (CSNT). Bhopal, India, pp 654–659. <https://doi.org/10.1109/CSNT51715.2021.9509614>

A Comprehensive Review of Conventional to Modern Algorithms of Speech Enhancement



M. Balasubrahmanyam, R. S. Valarmathi , and C. H. Mohan Sai Kumar

Abstract Almost all of today's communication techniques employ speech enhancement. It should go without saying that interference from the environment might reduce the quality of speech during transmission. Digital speech signal processing includes speech enhancement, which seeks to restore the clean speech from the contaminated signal. Different ways for speech enhancement are used depending on the kind of signal deterioration and noise present. The methods discussed in the review vary from single channel speech enhancement methods (statistical model-based approaches, and methods in time and transform domains), multi-channel speech enhancement methods (adaptive noise cancelation and multisensor beamforming), machine learning methods including Convolution Neural Networks, Deep Neural Networks, Residual Convolution Neural Networks, and speech enhancement methods with optimization techniques. The machine learning algorithms have shown considerable improvement of 98% in enhancing the speech quality of signals.

Keywords Speech enhancement · Single channel speech enhancement · Multi-channel speech enhancement · Neural networks · Machine learning

1 Introduction

A key component of the communication equipment is speech enhancement. It improves speech quality and lowers background noise, and it has applications in many fields, including teleconferencing, mobile phones, hearing aids, and voice communication systems. Due to the fact that it recovers and reconstructs the signal during

M. Balasubrahmanyam (✉) · R. S. Valarmathi · C. H. Mohan Sai Kumar
Vel Tech Rangarajan Dr, Sagunthala R&D Institute of Science and Technology, Chennai, India
e-mail: mbalu.422@gmail.com

R. S. Valarmathi
e-mail: drrsvalarmathi@veltech.edu.in

C. H. Mohan Sai Kumar
e-mail: chmohansaikumar@veltech.edu.in

deterioration, speech enhancement and speech restoration are intimately related to one another [1]. The speech enhancement and restoration techniques do differ slightly from one another. To restore speech means to return a noisy signal to its original state, before noise was added. On the other side, speech enhancement assists in bettering the original signal through refinement. Additionally, an original undergirded speech signal can be improved but not restored [2]. These speech enhancement techniques work to strengthen listener-fatiguing perceptual qualities of the degraded voice signal due to added noise, such as overall excellence or comprehension [3]. Enhancement of speech can be utilized in a variety of environments, such as areas where a building's background noise interferes, noisy streets, or roads where cars are passing. The original speech quality is degraded by these interference disturbances to the point that it is no longer clear. Speech bandwidth systems, which are mostly employed in the decoding of digital communication channels, are a context that must be addressed for speech enhancement [4–6]. This method is also required for speech decoding, which involves the integration of voice and data networks as well as speech bandwidth compression systems, which are crucial in systems for communication via speech.

Review work [2] addressed single and multi-channel speech improvement and carried out research on single channel voice augmentation techniques. A discussion of the spectrum subtraction approach and its modification was provided by Amole and Dhonde [7]. Time and transform domain voice enhancement techniques were covered by the authors in [8]. Sunnydayal and Sivaprasad reviewed statistically based voice enhancing strategies in [9]. As far as the authors are aware, there is no prior literature that groups speech enhancement in this manner.

2 Conventional Methods of Speech Enhancement

2.1 Single Channel Speech Enhancement Methods

A single microphone is typically employed for recording purposes in applications like mobile telephony, which are the focus of the single channel speech enhancement challenge. Degraded speech can be improved with the use of these methods. Basically, all of the techniques fall into one of two categories: supervised techniques and unsupervised techniques.

In supervised techniques as NMF, voice and cacophony are modeled, while conditions are generated with instruction examples. While processing is carried out utilizing unsupervised methods, such as the transform domain approaches shown in Fig. 1, the Kalman filter, the Wiener filter, and estimators employing Super-Gaussian without previous knowledge of the speaker's identity or the level of noise [10]. One of the challenging tasks in speech enhancement is the computation of power spectral density of noise, not necessary with supervised algorithms. The classification of these improvements in speech enhancement techniques is displayed in Fig. 2

[10] for better comprehension. Transform domain approaches predominate because frequency domain processing is typically simple and easier to understand in Fig. 2.



Fig. 1 Study’s review for methods of speech enhancement

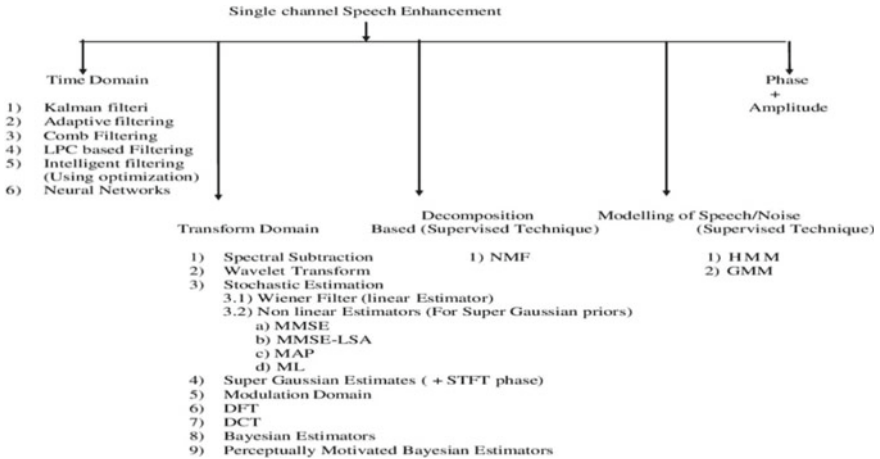


Fig. 2 Single channel speech enhancement methods

2.1.1 Time Domain Approaches

- a. *Kalman Filtering*: The Kalman filtering is broad term for wiener filtering. The Kalman filter is said to mix all the measured data that is currently accessible as well as the system's knowledge and the measurement equipment. One of the key distinctions between the Kalman filter and the wiener filter is the Kalman filter's sensitivity of signals that are not stationary [11].
- b. *Linear Predictive Coding*: When processing audio signals including speech, using information from a linear prediction model, linear predictive coding (LPC) is widely used to compress the spectrum envelop of a spoken digital signal. Inverse filtering is the technique used to remove formants. Remainder is the signal that is still present after the subtraction. It is possible to retain or communicate the numerical representations of the buzz's frequency, intensity, formants, and residue signal. The primary challenge facing the LPC system is how to extract the formants from the original signal. This can be solved by representing each of the n samples as a linear combination of the preceding samples.
- c. *Adaptive Filtering*: Applications for noise cancelation make use of adaptive filtering techniques. For noise cancelation, adaptive algorithms such as Least Mean Square (LMS) and Recursive Least Squares (RLS) are typically utilized. LMS has a minimal computational complexity and is relatively easy to implement. However, it is recognized that the LMS method has convergence problems when reconstructing signals. The convergence is significantly affected by changes in step size.

2.1.2 Transform Domain Approaches

- (a) *Spectral Subtraction Method*: Among the initial and most important methods for enhancing single channel speech is spectral subtraction. The iterative spectral subtraction, which involves reusing the improved output as input, results in a better reduction of the remaining noise after a certain number of times. In the real world, noise has varying effects on speech depending on the frequency. The signal that has been tainted by noise can be described as follows

$$y(\eta) = x(\eta) + d(\eta), \quad (1)$$

where $x(\eta)$ is original speech signal, $y(\eta)$ is output signal, and $d(\eta)$ is noisy signal. Speech signals are non-stationary; hence, the frequency domain is used to analyze the noise component frame-by-frame. The results of the discrete time Fourier transform are as follows:

$$Y(\omega) = X(\omega) + D(\omega). \quad (2)$$

The approach is employed to obtain the spectrum of improved speech.

$$\left| \hat{X}(\omega) \right| = |Y(\omega)| - \left| \hat{D}(\omega) \right|. \quad (3)$$

- (b) *Wavelet Transform*: The general DFT technique cannot localize time and frequency, i.e., it cannot supply the Wavelet employs a range of frequency bands with movable time windows, resulting in higher and lower frequency bands with superior resolution. With speech signals typically benign on-stationary, the wavelet transform is a potent tool for noise reduction in single channel speech improvement. By adding a threshold value to the coefficients after applying the wavelet transform, it is feasible to ignore the noise coefficients and so reduce noise. The sub-band processing strategy is worth-while to improve performance.
- (c) *Stochastic Estimation—Wiener Filtering*: The MSE optimal stationary linear filter for images with additive noise and blur is the wiener filter. It is necessary to make the assumption that the signal and noise process are second order stationary (in the sense of a random process) in order to calculate the Wiener filter. In order to maintain generality, only noise processes with zero mean will be taken into account for this description.
- (d) *Subspace Speech Enhancement based single channel Methods*: When considering speech estimation as a restricted optimizing issue, there is another category of voice enhancement techniques. A signal subspace speech improvement approach was put up by Surendran [12] employing a perceptual feature and the human auditory system's frequency masking or frequency disguising properties. In comparison with various benchmark speech enhancement techniques, the findings of their studies proved that their algorithm worked. Wang et al. [13] suggested a method of the subspace method for noise reduction in various noisy situations relying on the suitable method change and conventional primary factor strategy. It was demonstrated that their method performed better with white noise compared to colored noise. If the SNR of the signal is greater than 10 dB, then it has poor performance. Its decomposition nature sets it apart from the previous subspace techniques (Table 1).

2.2 Multi-channel Speech Enhancement

These systems have a higher level of sophistication than single channel systems. This system makes use of the many signal inputs available and uses noise reference as an adaptive noise canceling mechanism.

By taking into account constraints of single channel systems, compared to single channel systems, these systems can handle non-stationary noises better. Figure 3 demonstrates [14] strategies for improving speech on several channels.

Table 1 Benefits and drawbacks of conventional single channel methods for speech enhancement

Speech enhancement	Benefits	Drawbacks
Spectral subtraction	This approach manages the trade-off among residual noise and speech distortion using a straight forward mechanism and is economical in terms of processing	The negative aspect is the extra rhythmic noise
MSSE estimator	It has less computing tools and resources	There is no mechanism to control the conflict between residual noise and speech distortion
Wiener filtering	<ul style="list-style-type: none"> The algorithm guarantees de-reverberation performance independent of speech source azimuth angle 	In spaces with modest reverberation, there is much possibility for improvement
Kalman filter	<ul style="list-style-type: none"> Under the premise of known statistics of system and measurement errors, the Kalman filter offers a sequential, unbiased, and lowest error variance estimate for linear problems The ability to statistically create flow-dependent error covariance is the Kalman filter’s primary benefit in speech processing applications 	The approaches’ drawback is that they need certain data model characteristics for the particular battery. The resulting inaccuracy will rise if parameters that are uncertain or drifting are not taken into account
Subspace	It offers a method for balancing the effects of voice diffraction and background noise	As a result, there are high computational burdens
Adaptive noise canceller	The traditional ANC wide band algorithms get the greatest outcomes in the lower frequency ranges	Their performance rapidly deteriorates as the noise’s bandwidth and center frequency increase

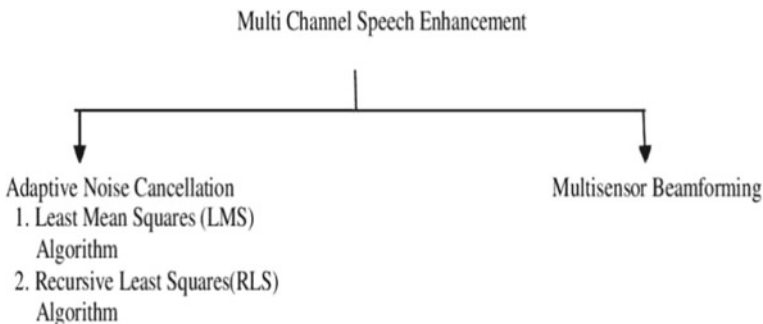


Fig. 3 Multi-channel speech enhancement methods

2.2.1 Adaptive Noise Canceller

An adaptive noise canceller (ANC) is essentially an electromechanical or electro-acoustic approach that disruption in favor of creating a gentler atmosphere, headphones, and hearing protection devices. The generalized beamforming technology known as a multiplexed system can be modified to work with ANC. In terms of adaptive algorithms, Least Mean Square [12] is the most popular. While a significant portion of the energy in the majority of sound parameters are concentrated in the lower frequencies, they also frequently contain not able high frequency components.

Types of Adaptive Noise Cancellation Filters: A computer tool known as an adaptive filter aims to develop and build associations connecting both signals using a sequential method in real time. The phases listed below define an adaptive filter [15]:

- (1) The signal being filtered.
 - (2) The configuration outlining how the filter derives its output signal based on the incoming source signal.
 - (3) Restrictions of framework which could be modified progressively for change how the input and output of the filters are associated.
- (a) *Least Mean Squares (LMS) Algorithm:* Most used methods for adaptable filtration is one of the Least Mean Square algorithm. Windrow and Hoff and Haykin are credited with laying the groundwork for it [16]. It is based on leveraging the mathematical properties of the input signal to approximate the gradient pointing toward the optimal outcome. The LMS algorithm simplicity is a significant trait. With each new sample in this process, filter weights are adjusted as necessary to produce the desired results. Using a transform domain adaptive filter significantly reduces the computational load. They detected and eliminated noise using the spectrogram approach for the cancelation or suppression of the various noises. By using the technique described in this study.
 - (b) *Recursive Least squares (RLS) Algorithm:* An adaptive filter utilizes an adaptive algorithm to change the filter restrictions and structure since it is based on changing the treated signal [17]. The rest of the filter structure is often left unchanged, only the filter coefficients changing. Adjusting the adaptive filters weight's coefficients based on an error signal, the RLS adaptive noise cancellation method produces an estimated interference signal output from the filter. The method then subtracts the filter output from the combined signal with the noise component to obtain the strong signal and produce the desired result of removing the noise signal. Table 2 summarizes the benefits and drawbacks of Recursive Least Square and Least Mean Square methods.

2.2.2 Multisensor Beamforming

Beamforming is a technology that combines several channels with sophisticated enhancing the desired signal and reducing the noise signal via space time domain

Table 2 Benefits and drawbacks of conventional multi-channel speech enhancement approaches

Speech enhancement	Benefits	Drawbacks
LMS	The LMS algorithm is straightforward to use and simple to implement [18]	Slower convergence
RLS	<ul style="list-style-type: none"> • The SNR can rise by up to 10 dB or more • The reduction coefficient may exceed 0.9 • It is possible to obtain good noise suppression compared to LMS, RLS converges more quickly 	Complex and costly

filtering algorithms. When beamforming an array of two or more microphones in a certain geometric arrangement is used. After the sensor outputs have been filtered, the signals are amplified or attenuated by a beamformer depending on the signals' directions of arrival (DOA). This tactic is based on the underlying assumptions because the impact is insignificant and that it is known which way the desired signal will arrive.

3 Machine Learning Approaches

3.1 Neural Networks for Speech Enhancement

Neural Networks-based Speech Enhancement (NNSE) used to help Cochlear Implant (CI) users understand speech in noisy environments [19]. The technique breaks down the noisy speech input into time–frequency bands, extracts a set of auditory characteristics.

In order to maintain speech-dominated components for electrical stimulation, the noise-dominated component is reduced using this approximation. The NNSE algorithm is suitable for use in hearing aids due to its architecture and short processing time. Figure 4 illustrates the framework [20] for speech enhancement using neural network.

3.1.1 Deep Neural Networks for Speech Enhancement

To get the DNN ready, a Restricted Boltzmann Machine pre-training strategy was implemented [21]. A sizable training set is essential to understanding the DNN's complex structure. It has been demonstrated that using additional acoustic framework statistics enhances quality and makes augmented communicating fewer sporadic. Training for several ailments can be used to solve issues including voice enhancement for unfamiliar microphones.

Sensitivity grades within various ambient conditions, along with translation expansion. Based on TIMIT corpus, significant improvements were made when

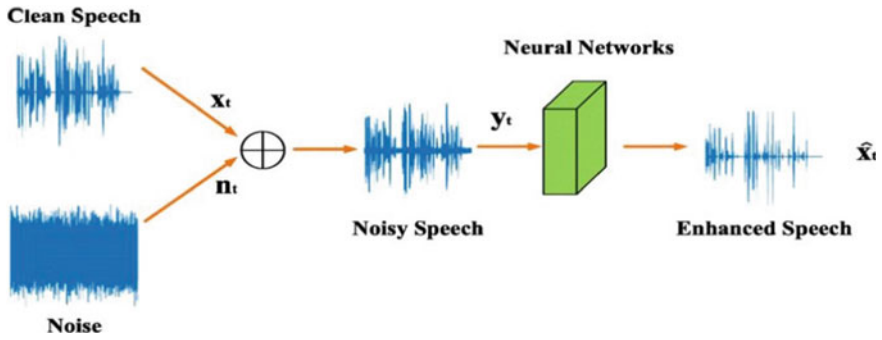


Fig. 4 Framework for speech enhancement using neural network

compared to the SNN-based and Log-MMSE algorithms. Due to the improved speech’s reduced musical noise, an average subjective preference of 76.35% was achieved. This was a method used to enhance voice by using DNN to identify a mapping across stutter-free and clear speech. Figure 5 illustrates a DNN framework [22] with N hidden layers.

The ensuing DNN model, trained with artificially generated data, is also efficient in handling noisy speech data recorded in real-world scenarios without producing the annoying musical artifact typically seen in conventional enhancement techniques. A good generalization skill to environments with concealed noise can be achieved through training with numerous conditions and various noise types. As a result, the suggested DNN architecture has a significant impact on how non-stationary noise is managed in real-world scenarios.

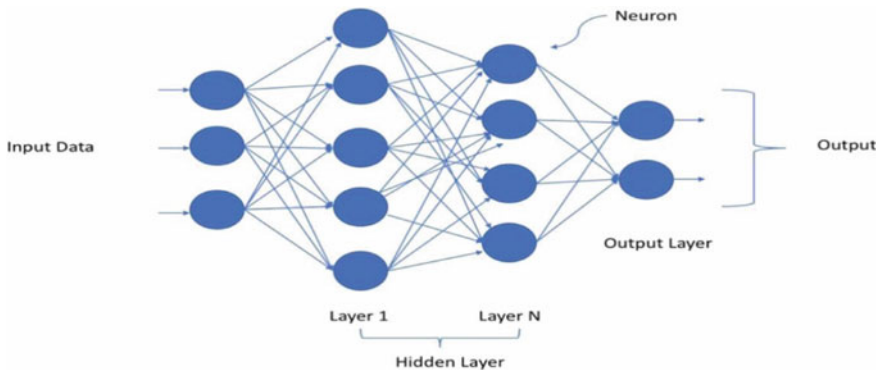


Fig. 5 Deep Neural Network with N hidden layers

3.1.2 Convolution Neural Networks for Speech Enhancement

A model based on a Convolution Neural Network (CNN) that is Signal-to-Noise Ratio (SNR) aware was proposed for Speech Enhancement (SE) in 2016. The local temporal and spectral speech signals are well handled by this CNN model. As a result, the model can efficiently differentiate noise from an input signal from speech signals. To increase these models' ability to generalize and their precision, two SNR-aware methods were put forth utilizing CNN. A Multi-Task Learning (MTL) framework is used in the first algorithm. The model receives as input the noisy speech signal. The technique restores noise-free voice sounds mostly from the input. The SNR level for the cleaned-up speech signals is then estimated.

This conclusion is supported by the results. Additionally, even at unknown SNR levels, the SNR-aware CNN models have improved de-noising capability. This suggests a robust and promising potential for applications in the actual world. A new CNN approach was most recently put forth in 2017 to improve complicated spectrograms and address the challenge of phase estimation.

The suggested model separates noise-free real and imaginary (RI) spectrograms from dirty ones. These waveforms have highly accurate phase information. The Multi-Metric Learning (MML) criterion was used to create the objective function, allowing for the consideration of many metrics. The structure of RI spectrum may be seen regardless of the signal description, which is the core tenet of MML. With the right choice of, MML can simultaneously improve several objective measures.

An input layer, hidden layers, and an output layer make up a convolutional neural network, as shown in Fig. 6 of a typical convolutional neural network.

They consist of (1) MAP estimator for speech coefficients in the ICA domain, which is then used to estimate improved speech in the temporal domain. (2) ICA

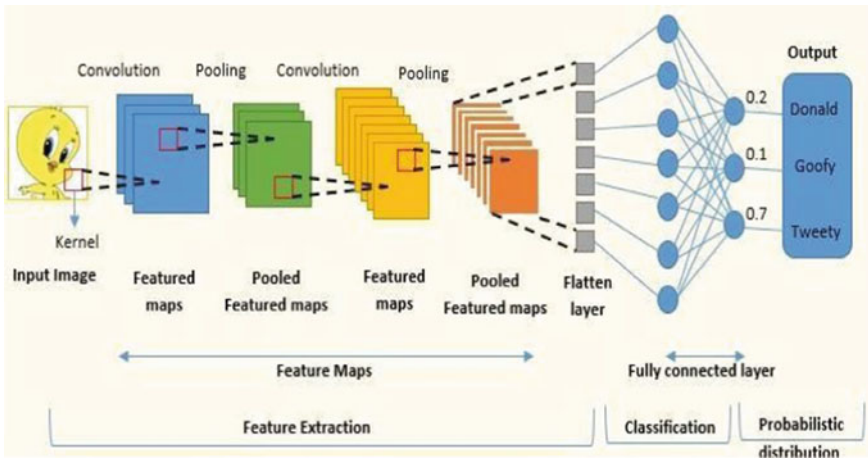


Fig. 6 Typical convolutional neural network

domain data transformation, using lessons learnt from speech training data and used in step one.

The suggested approach was tested and trained using spoken keywords like directions for driving. When compared to real-time noisy speech signals from cars, streets, offices, and factories, our technique significantly improved the de-noising performance with respect to SNR and distortion in boosted signals.

3.1.3 Optimization Techniques for Speech Enhancement

In terms of speech enhancement, this section examines a few well-known and current optimization strategies. All of the optimization methods described here take into account the use of a dual channel enhancement, where one channel is devoted to speech that has been distorted by noise and the other is for pure noise.

1. *BAT Algorithm:* A crew of specific meta-heuristic technique called Bat Algorithm, which was created, was inspired by the hunting techniques of bats. BA is based on how echolocation works in microbats. The overall wide range of alternatives is increased through spectrum adjusting in this technique. Additionally, it incorporates algorithm's automated adjustment capacity with digital enlargement characteristics bats, include their rate of heartbeat and how loud they become as they draw closer to their target. By switching from exploration to exploitation as global optimality becomes closer. This method justifies itself as being an incredible successful improvement method if contrasted comparable meta-heuristic techniques, because it is the first attempt to balance these crucial components. Another attempt to develop dual channel speech enhancement systems employing the Bat Algorithm (BA) was made in. In this method, the weights for the adaptive filter are determined using BA. The input signals are initially divided into frames as part of the approach.

The simulation outcomes using BA were contrasted with those using normal, speeded up PSO, the Gravitational Search Algorithm (GSA), and Hybrid PSO-GSA-based voice improvement methods in 2016 [22].

2. *Modified BAT algorithm:* In 2016, the performance of simulations based on BA was compared to that of normal algorithms. The meta-heuristic BA's advantages over other algorithms for enhancing voice signals are clearly shown by the results. BA has superior performance over the SPSO algorithm in case of enhanced speech's quality and comprehensibility.
3. *PSOGSA algorithm:* The warped noise voice information in twin channel structures could be improved, a hybrid (PSOGSA) Particle Swarm Optimization and Gravitational Search Algorithm was presented. Each member of the swarm, which stands for the filter coefficients, is considered a potential solution. These adaptive filter coefficients are optimized using PSOGSA. The PSOGSA outperformed the GSA and SPSO in terms of performance. Because of this, PSOGSA is better able to minimize the surrounding undesired noise signals of the messy input voices.

3.1.4 Residual Convolutional Neural Network

This study focuses on deep speech enhancement, a technique that virtually eliminates background noise by using Deep Neural Networks for noise reduction and produces a clean speech from a loud one. Using multifaceted convolutional neural networks with residual interactions, this approach takes advantage of both the linearity provided by residual networks and the nonlinear functions acquired by convolution neural networks. These properties assist in the reduction of noise even in low SNR scenarios, where the majority of other speech enhancement models often fail. The proposed model performs a similar function more quickly, accurately, and efficiently than existing deep learning models, and it also requires less training.

4 Multimodal Approaches to Speech Enhancement

In recent years, it was looked into to see if visual speech data could be used to improve the noise-damaged audio speech signals. In 2002, an attempt was made to determine how Automatic Speech Recognition (ASR) systems would be affected by the use of audio-visual speech data for boosting noisy audio qualities. Through the use of linear filters on combined audio and visual data, speech signals are improved. However, the proposed technique performed worse when compared to the audio-visual discriminant feature fusion. The non-stationary non-complex linear filter used for boosting can be blamed for the poor performance.

Independent CNNs are used to process the auditory and visual data to a significant extent (Table 3).

The speech enhancement methods are evaluated on the TIMIT dataset for various noises like BBE, FTRY, STRT, AND PTRT (Babble, Factory, Street and Restaurant). Tables 4, 5, and 6 show the performance of various methods at various SNR in case of various PESQ and Segmented SNR. The training and test sets were created using the TIMIT corpus [24]. Randomly chosen 600 utterances were distorted with different noise types are BBE, FTRY, STRT, AND PTRT (Babble, Factory, Street and Restaurant), and (0, 6, and 12 dB) at six SNR levels. To create the test set, another 150 randomly chosen utterances were combined. The training and test sets' noise types and SNR levels were mismatched to create more realistic experimental conditions.

The input feature vectors were normalized based on mean and variance to increase the stability of the process of training. The Deep Neural Networks used during this examination contained six concealed layers, everyone with 1500 nodes, and activation functions called parametric rectified linear units (PReLU) [24]. CNN had four convolutional layers with padding, each with 50 filters and a filter size of 25×1 , two fully connected layers with 512 nodes each, and two layers with 512 nodes each to create the test set, two 6 dB and 12 dB models were used.

Segmented Signal-to-Noise Ratio (SSNR) and perceptual evaluation of speech quality (PESQ) scores were employed to assess both time domain and frequency

Table 3 Benefits and drawbacks of the machine learning techniques for enhancing speech

Speech enhancement	Benefits	Drawbacks
NN	<ul style="list-style-type: none"> • Simple computational architecture • Reduced processing time 	Accuracy may be better and the generalization performance under uncertain settings isn't particularly strong
DNN	<ul style="list-style-type: none"> • Better performance when compared to L-MMSE and SNN-based algorithms • Really good at managing distorted loud voice recorded under various conditions and several situations that weren't present during DNN training • Effectively reducing highly non-stationary noise, this is typically challenging to deal with [23] 	Improvements are required in the DNN's capacity to generalize to unseen noise, and there is a requirement for a large training set to adequately cover the range of acoustic situations, including speaker and language variances [23]
CNN	<ul style="list-style-type: none"> • Better performance compared to DNN • Capable of handling local spectral and temporal speech signal structures • Effective separation of voice and background noise from noisy input signals • Improved noise reduction with unknown SNR levels • Promising methodology for practical applications • Reliable phase data in increased signal, simultaneous improvement in several metrics [14] 	To train effectively, CNN needs a lot of labeled data. CNN is expensive and time-consuming

Table 4 Various methods' performance at 0 dB SNR level

Method	PESQ				Segmented SNR			
	BBE	FTRY	STRT	RTRT	BBE	FTRY	STRT	RTRT
NN	1.61	1.53	1.71	1.62	- 2.802	- 2.862	- 2.641	- 2.756
DNN	1.942	1.952	2.088	2.011	- 0.465	- 0.486	- 0.369	- 0.417
CNN	2.078	2.056	2.256	2.145	0.399	0.411	0.441	0.321

Table 5 Various methods' performance at 6 dB SNR level

Method	PESQ				Segmented SNR			
	BBE	FTRY	STRT	RTRT	BBE	FTRY	STRT	RTRT
NN	2.354	2.3412	2.369	2.359	1.712	1.621	1.758	1.723
DNN	2.465	2.458	2.478	2.470	1.810	1.792	1.825	1.812
CNN	2.575	2.570	2.596	2.579	1.856	1.836	1.952	1.869

Table 6 Various methods' performance at 12 dB SNR level

Method	PESQ				Segmented SNR			
	BBE	FTRY	STRT	RTRT	BBE	FTRY	STRT	RTRT
NN	2.632	2.622	2.645	2.639	3.210	3.125	3.289	3.262
DNN	2.745	2.665	2.699	2.641	3.356	3.301	3.457	3.378
CNN	2.758	2.712	2.796	2.786	3.511	3.499	3.546	3.522

domain signals have different properties, in order to compare the performance of various models.

5 Conclusion and Future Scope

This study is an overview of the approaches researchers have taken to the problem of speech enhancement over the years. Early research in this field focused on various spectrum enhancement techniques, statistical-based algorithms, and subspace enhancement techniques. These have all shown promise in testing, but each has its own set of limitations in real-world situations. In this regard, adaptive noise cancelation is a further well-liked field. It can be adjusted by using predictive optimization approaches to adjust its parameters, which has made it an ongoing area of study. The scope of machine learning algorithms is fairly broad. Within the confines of this essay, it is impossible to discuss them all. We discussed about some of the more notable ones and mentioned their advantages.

Artificial intelligence developments have had positive effects on improving speech. A powerful tool in this field has been shown to be neural networks. DNN came after simple Neural Network (NN), it provided improved results but still poor real-world performance generalization when it came across noise and speech signals that hadn't been visible to it during the training phase. After that, CNN came into existence, which has since established itself as a trustworthy method for extrapolating noise canceling issues in the actual world. In the future, we will investigate and experimenting with optimization machine learning-based filters for speech enhancement in real-world scenarios. The proposed model Adaptive Residual Convolutional Neural Network, Expected results in better reduction in speech distortion, better

reduction in speech noise, and better model size. This research is a ground-breaking effort to fully utilize neural networks' capabilities without concentrating on growing their size to achieve better performance. More noise types and SNR levels can be taken into account in upcoming work to create a network that is more reliable. Also provide better PESQ and SSSNR. Speech enhancement is the basis for all audio and communication devices.

References

1. Shishir B, Jimish D, Darshana G (2013) GUI based performance analysis of speech enhancement techniques. *Int J Sci Res Publ* 3(9):1–7
2. Ravi Kumar K, Subbaiah PV (2016) A survey on speech enhancement methodologies. *International Journal of Intelligent Systems and Applications* 8(12):37–45
3. Philipos C (2013) *Loizou: speech enhancement: theory and practice*, 2nd edn. CRC Press Inc., Boca Raton
4. Hardik P, Karan G, Dharmesh P (2016) Comparison of different speech enhancement techniques. *Imperial Journal of Interdisciplinary Research* 2(5)
5. Soumasunderaswari D, Prashanthini K (2015) A survey on various multi-channel speech enhancement algorithm. *International Journal of Advanced Research in Computer and Communication Engineering* 4(1):254–255
6. Sunitha D, Yusuf Mugle MD (2015) Review on speech enhancement techniques. *Int J Comput Sci Mob Comput* 3(8):285–290
7. Chaudhari A, Shivraj D (2015) A review on speech enhancement techniques. In: *International conference on pervasive computing (ICPC)*, pp 1–3
8. Kulkarni DS, Deshmukh RR, Shrishrimal PP (2016) Review of speech signal enhancement techniques. *International Journal of Computer Applications* 139(14):23–26
9. Sunnydayal V, Siva Prasad N, Kishore Kumar T (2014) A survey on statistical based single channel speech enhancement techniques. *International Journal of Intelligent Systems and Applications* 12:69–85
10. Nasser M, Timo G, Arne L (2011) A new linear MMSE filter for single channel speech enhancement based on non negative matrix factorization. In: *IEEE workshop on applications of signal processing to audio and acoustics, WASPAA 2011*. IEEE, New Paltz, NY, USA, pp 1–2
11. Berouti B, Schwartz R, Makhoul J (1979) Enhancement of speech corrupted by acoustic noise. In: *Proceedings of international conference on acoustic, speech, and signal processing*. Washington DC, pp 208–211
12. Xu Z, Strake M, Fingscheidt T (2021) Deep noise suppression maximizing non-differentiable PESQ mediated by a non-intrusive PESQNet. *IEEE, Transactions on Audio, Speech and Language Processing*
13. Smita G, Suraj BT, Hari KC, Sharan T (2022) A deep learning approach for noise removal and enhancement of a real time audio signal. *International Research Journal of Modernization in Engineering Technology and Science* 4(3):1066–1072
14. Lakshmikanth S, Natraj KR, Rekha KR (2014) Noise cancellation in speech signal processing— a review. *International Journal of Advanced Research in Computer and Communication Engineering* 3(1):5175–5186
15. Xiang X, Zhang X, Chen H (2021) A convolutional network with multi-scale and attention mechanisms for end-to-end single-channel speech enhancement. *IEEE Signal Processing Letters* 28:1455–1460
16. Hasannezhad M, Ouyang Z, Zhu WP, Champagne B (2021) Speech enhancement with phase sensitive mask estimation using a novel hybrid neural network. *IEEE Open Journal of Signal Processing* 2:136–150

17. Chen Y, Dong B, Zhang X, Gao P, Li S (2021) A hybrid deep-learning approach for single channel HF-SSB speech enhancement. *IEEE Wireless Communications Letters* 10(10):2165–2169
18. Pandey A, Wang D (2021) Dense CNN with self-attention for time-domain speech enhancement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29:1270–1279
19. Bhat GS, Shankar N, Reddy CKA, Panahi IMS (2019) A real-time convolutional neural network based speech enhancement for hearing impaired listeners using smartphone. *IEEE Access* 7:78421–78433
20. Karjol P, Ajay Kumar M, Ghosh PK (2018) Speech enhancement using multiple deep neural networks. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), Calgary, AB, Canada, pp 5049–5052
21. Hou J-C, Wang S-S, Lai Y-H, Tsao Y, Chang H-W, Wang H-M (2018) Audio-visual speech enhancement using multimodal deep convolutional neural networks. *IEEE Trans Emerg Top Comput Intell* 2(2):117–128
22. van Den A et al (2016) WaveNet: a generative modal for raw audio. *Speech synthesis workshop*
23. Garofolo JS et al (1993) TIMIT acoustic-phonetic continuous speech corpus. National Institute of Standards and Technology
24. Rix A, Hekstra A (2001) Perceptual evaluation of speech quality, an objective method for end-to end speech quality assessment of narrow band telephone networks and speech codes. *ITU-T Recommendations*
25. Tsun AH, Wang HM, Lu X, Tsao Y (2020) WaveCRN: an efficient convolutional recurrent neural network for end-to-end speech enhancement. *SPL*, pp 1–5

Gain Improvement of PIFA Antenna Using Glass Substrate in Comparison with FR4 Substrate



G. Kalpana and Bhuvaneshwari Balachander

Abstract The primary intention of this study is to design a novel planar inverted F-antenna (PIFA) using glass substrate in comparison with FR4 dielectric material at a frequency range of 8.5 GHz to optimize the gain in (dB) using a high-frequency structure simulation (HFSS). **Materials and Methods:** The gain of a PIFA antenna using glass substrate ($n = 20$) compared with a PIFA antenna using FR4 substrate ($n = 20$) using the HFSS tool, where n is the sample size. In this study, there are two groups, each with 20 samples and an 80% pretest power (G-power). The PIFA antenna was created using HFSS software utilizing the finite element method. By simulating the PIFA antenna, the bandwidth was calculated. **Results:** The PIFA antenna using glass substrate has an improved gain -14.2430 dB than that of the PIFA antenna using FR4 substrate with a gain of -14.2151 dB. The proposed antenna produced a G-power of 80%. **Conclusion:** The gain of the PIFA antenna using glass substrate shows significantly better performance compared with the FR4 substrate.

Keywords Antenna design · Novel PIFA antenna · HFSS tool · FR4 epoxy substrate · Bandwidth · Glass substrate · Frequency

1 Introduction

This project's objective is to simulate a novel PIFA antenna design using a glass substrate in comparison with FR4 substrate at a frequency range of 8.5 GHz so as to optimize the gain in (dB) by using a high-frequency structure simulation (HFSS) [1]. PIFA antennas play an important role in mobile communication. The design of tiny antennas is critical in the development of portable communication systems. An antenna's resonance frequency, bandwidth, and gain are all directly related to its size and substrate [2]. The novel antenna designs are useful in portable applications including wireless networks, WLAN [3], and Bluetooth [4]. PIFA applications are

G. Kalpana · B. Balachander (✉)

Department of Electronics and Communication Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India
e-mail: bhuvaneshwari@saveetha.com

wireless communication networks and mobile networks [5]. An ultra-compact wide-band modified planar inverted F-antenna for 2.4/5 GHz WLAN applications, PIFA has been developed [6].

A lot of research has been done in this simulated antenna design by analyzing the PIFA antenna using two different substrates. Several articles have been published on this simulated antenna design over the past few years. On this subject, 71 publications were published in Science Direct and approximately 110 were found in Google Scholar. A unique bandwidth augmentation strategy for internal multi-band handset antennas to solve the issue of limited bandwidth for traditional planar inverted F-antennas was proposed [7]. PIFAs have been widely used as embedded antennas in the cell phone industry that can communicate with digital base stations via electromagnetic waves [8]. A parasitic element and two slots can be added to the patch of the basic PIFA to increase the bandwidth on the higher frequency band, enabling it to cover more communication systems [9]. This criterion should be taken into account in the design of the RF transmitter/receiver modules (including antennas) of the mobile devices. The effect of the dimensions of a planar inverted F-antenna (PIFA) on its resonance frequency and the frequency bands that it can support [10]. Our staff has a wealth of knowledge and research expertise, which has resulted in publications of the highest caliber [11–20].

The main problem that was identified by the literature work was relatively low gain value in existing research. Many studies have been carried out using various substrates to understand their performance. The proposed work focused on getting high gain by designing a PIFA antenna by changing the various substrates to efficiently boost the antenna's performance.

2 Materials and Methods

The research was carried out at the Saveetha School of Engineering, SIMATS, Tamil Nadu, India, in the Antenna and Wave Propagation Laboratory from the Department of Electronics and Communication Engineering. For the design and modeling of antennas, HFSS tool was utilized [21]. There were two groups in this study, each with 20 samples, for a total of 40 samples with an 80% pretest power analysis [3].

In the HFSS modeling tool, Group 1 refers to the FR-4 epoxy substrate utilized as a dielectric material medium based on a PIFA antenna. All of the input variables were used in the design process. Select the FR4 substrate in the HFSS workplace. PIFA antenna design using FR4 epoxy substrate and antenna gives the XY plots in which X plot represents frequency and Y plot represents gain in dB. By moving the cursor on the XY plot, 20 samples were taken as a dataset which is given to SPSS software.

Group 2 refers to the glass substrate that was utilized as a dielectric material in the HFSS modeling program to create a PIFA antenna. All the input variables were used in the design process. Select glass substrate in the HFSS tool. PIFA antenna design using glass substrate and antenna gives the XY plots in which X plot represents

frequency and Y plot represents gain in dB. By moving the cursor on the XY plot, 20 samples were taken as a dataset which is given to SPSS software.

The entire work is carried out in windows 10 pro, Intel Core i3, 10th generation, and HFSS software. Ansoft HFSS: Software for 3D EM simulation in RF and Wireless Design is used for designing novel antenna designs and for simulation processes. First, the HFSS tool should be opened. Select the FR4 and glass substrate and dimensions for substrate, dielectric material, feeding, and shorting pins are given. Also, dimensions for substrate and feeding pins should also be given. Finally, validation and design analysis should be done. Thus output can be obtained by selecting the substrates. The data can be exported in the form of excel from the output. This data should be given to SPSS software. Statistical softwares like HFSS and SPSS are used. While SPSS is used for statistical analysis of the data, HFSS is used to build a PIFA antenna and for simulation.

2.1 Statistical Analysis

The statistical software used is SPSS version 2.1 that was used for statistical comparison of parameter gain. SPSS software is used to calculate the mean values, standard deviations that are simulated as graphical representation, and tables. In this research work, the independent variables are length (L), height (H), and width (W) to evaluate the performance of the PIFA antenna and attained significant value ($p < 0.05$), while frequency (dB) and gain (dB) are dependent variables [22].

3 Results

According to the findings, the glass substrate PIFA antenna achieved a high gain compared with the substrate antenna for FR4. With effective output parameters like gain, the suggested solution requires fewer simulation iterations and uses less time overall.

Figure 1 represents the top view of a novel PIFA antenna using FR4 substrate and glass substrate medium in an 8.5 GHz frequency band. The antenna is constructed with the desired variables: length (L) is 80 mm, width (W) is 80 mm, and height (H) is 3.8 mm. Figure 2 represents the side view of a novel simulated model of the PIFA antenna using FR4 substrate and glass substrate medium in an 8.5 GHz frequency band. The antenna is constructed with the desired variables of length (L) is 80 mm, width (W) is 80 mm, and height (H) is 3.8 mm. Figure 3 represents gain improvement of FR-4 substrate on X -axis as frequency in GHz and Y -axis as gain of FR-4 substrate in dB. It is showing the gain of FR4 substrate of PIFA antenna – 14.2151 dB at 8.5 GHz. Figure 4 represents gain improvement of glass substrate on X -axis as frequency in GHz and Y -axis as gain of glass substrate in dB. It is showing the gain of glass substrate of PIFA antenna – 14.2430 dB at 8.5 GHz.

Figure 5 represents comparison of FR4 substrate and glass substrate of PIFA antenna classified as measured by mean gain. The mean gain of the glass substrate of the PIFA antenna is significantly better than the FR4 substrate of the PIFA antenna and the standard deviation of the glass substrate of the PIFA antenna is slightly better than the FR4 substrate of the PIFA antenna. X-Axis: FR4 substrate of the PIFA antenna versus glass substrate of the PIFA antenna algorithms. Y-Axis: Mean gain of detection.

Table 1 represents the comparison of gain improvement between two substrates. The gain improvement of the glass substrate of the PIFA antenna is -14.2430 dB and the FR4 substrate gain is -14.2151 dB at frequency of 8.5 GHz. Table 2 represents

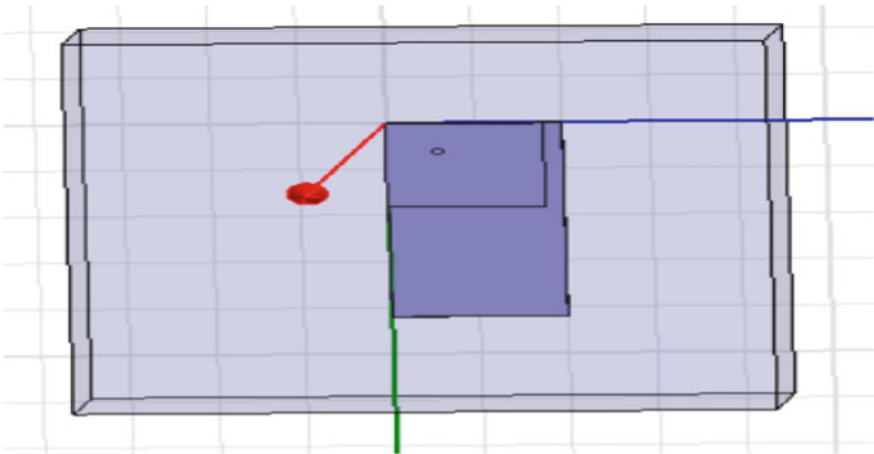


Fig. 1 Represents the top view of a novel PIFA antenna using FR4 substrate and glass substrate medium at a frequency range of 8.5 GHz. The antenna is constructed with the desired variables length (L) is 80 mm, width (W) is 80 mm, and height (H) is 3.8 mm

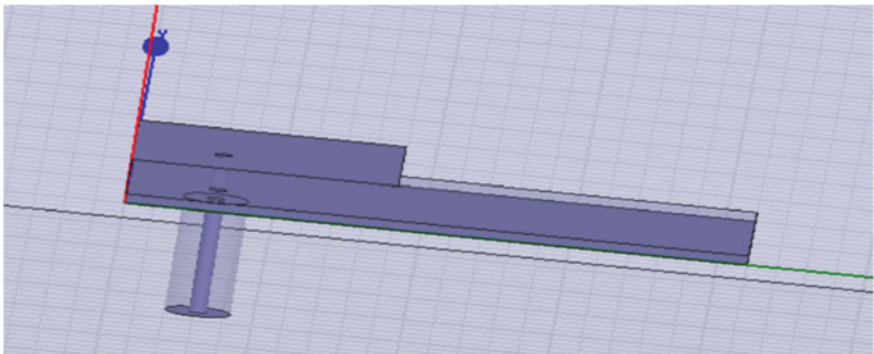


Fig. 2 Represents the side view of a novel simulated model of the PIFA antenna using FR4 substrate and glass substrate medium at a frequency range of 8.5 GHz. The antenna is constructed with the desired variables length (L) is 80 mm, width (W) is 80 mm, and height (H) is 3.8 mm

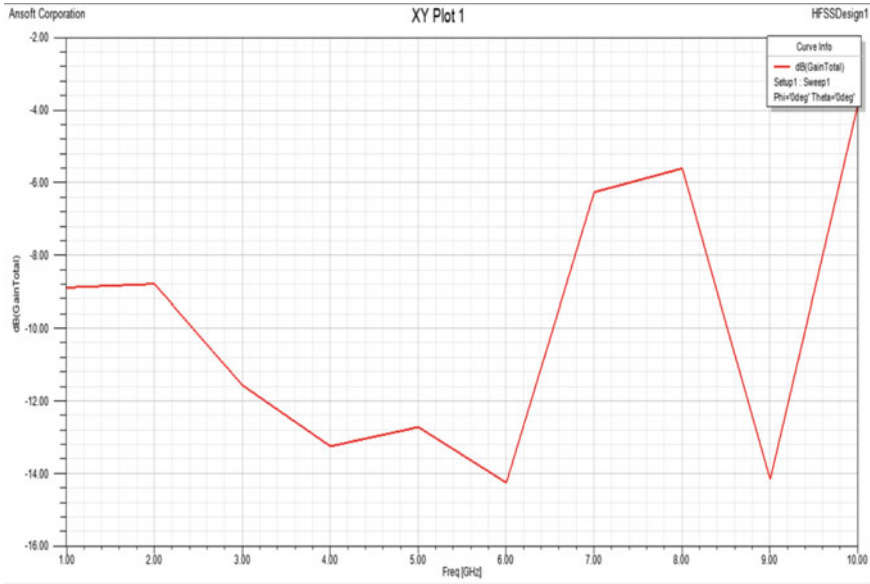


Fig. 3 Gain improvement of FR-4 substrate on X-axis as frequency in GHz and Y-axis as gain of FR-4 substrate in dB. It is showing the gain of FR-4 substrate of PIFA antenna – 14.2151 dB at 8.5 GHz

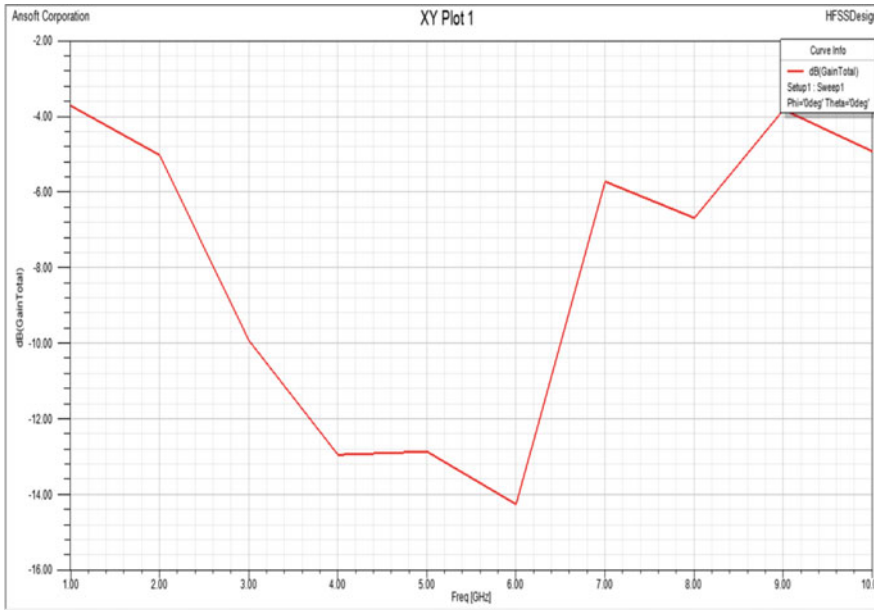


Fig. 4 Gain improvement of glass substrate on X-axis as frequency in GHz and Y-axis as gain of glass substrate in dB. It is showing the gain of glass substrate of PIFA antenna – 14.2430 dB at 8.5 GHz

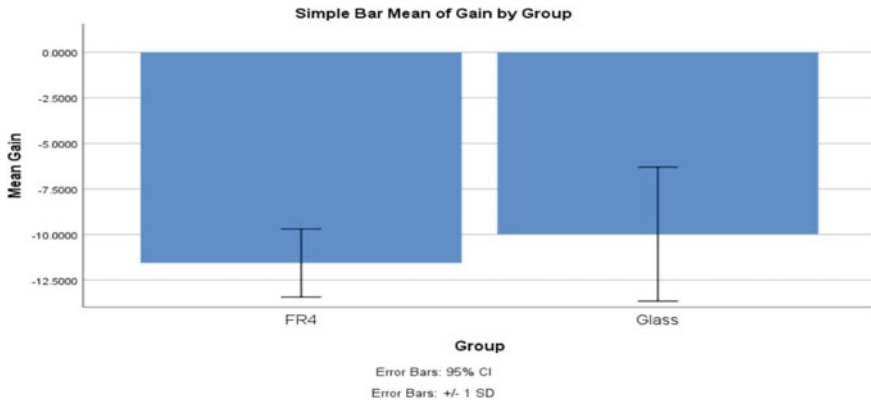


Fig. 5 Comparison of FR4 substrate and glass substrate of PIFA antenna classifier in terms of mean gain. The mean gain of the glass substrate of the PIFA antenna is significantly better than the FR4 substrate of the PIFA antenna and the standard deviation of the glass substrate of the PIFA antenna is slightly better than the FR4 substrate of the PIFA antenna. X-Axis: FR4 substrate of the PIFA antenna versus glass substrate of the PIFA antenna algorithms Y-Axis: Mean gain of detection

the statistical analysis of PIFA antenna used in the proposed work. The mean gain of the glass substrate of the PIFA antenna is $- 9.981675$ dB and the mean gain of FR4 epoxy substrate of PIFA antenna is $- 14.2151$ dB. Table 3 represents the significance, mean difference, and standard error difference of gain of PIFA antenna shown using the independent sample test. The two groups’ standard error difference is 0.9233130. The significance level for gain was determined to be ($p < 0.05$). The significance level of gain is 0.023.

4 Discussion

In this analysis, the FR4 is given a significance value of 0.002 for the independent sample test substrate PIFA antenna in comparison with a graphite substrate. From the simulation, we noticed the gain of the FR4 substrate is 0.89 dB at a resonating frequency of 8.5 GHz.

An inverted F-antenna with two bands technology used for handheld electronics 450 LTE and GSM 900 is displayed. The inexpensive antenna is simple to construct and is designed for 450 LTE and GSM 900, respectively. The suggested small antenna has excellent measured overall efficiency, realized gain at both bands, and good omnidirectional radiation characteristics [23]. A tri-band vertical inversion. The proposed planar inverted-F antenna (PIFA) and design. The antenna can function at three different operating GSM and Bluetooth and uses the following frequencies: 900, 1800, and 2400 MHz. A dual L-shaped slot on the patch was used to realize the main idea of achieving these frequencies (Redzwan 2015). We offer a brand-new, small, single-feed, quad-band planar inverted F-antenna. For a sturdy framework and

Table 1 Shows the comparison of gain improvement between two substrates

Frequency (GHz)	Gain (dB) of glass substrate PIFA antenna	Gain (dB) of FR4 substrate PIFA antenna
1	- 4.0080	- 8.8327
1	- 4.2311	- 8.7769
1	- 4.5378	- 8.7490
1	- 4.7610	- 8.7211
2	- 5.8207	- 9.1673
2	- 6.7968	- 10.2550
2	- 8.1633	- 10.7849
2	- 9.2789	- 11.3984
3	- 10.3386	- 11.5100
3	- 11.0080	- 12.0398
3	- 11.9004	- 12.2629
3	- 12.6534	- 12.9602
4	- 12.9602	- 13.2112
4	- 12.7371	- 12.9602
4	- 12.7928	- 12.7649
4	- 12.9323	- 12.7092
5	- 13.0438	- 12.8486
5	- 13.5179	- 13.3506
5	- 13.9084	- 13.8526
6	- 14.2430	- 14.2151

The gain improvement of the glass substrate of the PIFA antenna is - 14.2430 dB and the FR4 substrate gain is - 14.2151 dB at frequency of 8.5 GHz

Table 2 Represents the statistical analysis of PIFA antenna used in the proposed work

	Group	N	Mean	Std. deviation	Std. error mean
Frequency	FR4	20	3.050	1.5381	0.3439
	Glass	20	3.050	1.5381	0.3439
Gain	FR4	20	- 11.568530	1.8710745	0.4183850
	Glass	20	- 9.981675	3.6809263	0.8230802

The mean gain of the glass substrate of the PIFA antenna is - 9.981675 dB and the mean gain of FR4 epoxy substrate of PIFA antenna is - 11.568530 dB

simple shielding, a foam dielectric substrate is employed. The design is compatible with mobile and wireless communication technologies because of its small size and thin substrate thickness [24]. For mobile devices, for penta-band operation, a dual-feed PIFA has been presented. The suggested antenna takes up a minimal amount of space on the ground plane. It has a volume of roughly 10 35 5 mm³ and doesn't

Table 3 Independent sample test displays the significance, mean difference, and standard error difference of gain of PIFA antenna

	Levene's test for equality of variances		t-test for equality of means							
	F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. error difference	95% confidence interval of the difference		
								Lower	Upper	
<i>Gain</i>	Equal variances assumed	15.558	0.0021	-1.719	38	0.094	-1.5868550	0.9233130	-3.4560045	0.2822945
	Equal variances not assumed			-1.719	28.204	0.097	-1.5868550	0.9233130	-3.4775594	0.3038494
<i>Frequency</i>	Equal variances assumed	0.0015	0.002	1.245	38	1.000	1.2859	0.4864	-0.9847	-0.9847
	Equal variance not assumed			1.862	38.000	1.000	1.2859	0.4864	-0.9847	-0.9847

The standard error difference between the two groups is 0.9233130. The significance level for Gain was determined to be ($p < 0.05$). The significance level of Gain is 0.023

need a clearance area. The PIFA utilizes separate high-band and low-band feeds to achieve penta-band coverage of the GSM900/1800/1900/UMTS2100/LTE2300/2500 communication protocols [25].

The proposed method focused on the optimization of gain by comparing two separate antennas, named novel PIFA antenna using glass substrate and FR4 substrate. The limitation of the novel PIFA antenna with glass substrate is its thick substrate. Future research can further optimize the gain to boost the antenna's functionality.

5 Conclusion

The simulation results for the designed antennas were obtained. The gain of the PIFA antenna with a glass substrate is -14.2403 dB at 8.5 GHz, while the gain of the FR4 substrate is -14.2151 dB at the same frequency. When compared with PIFA antennas employing FR4 substrates, glass substrate PIFA antennas have a noticeably higher gain.

References

1. Elfergani I, Hussaini AS, Rodriguez J, Abd-Alhameed R (2017) *Antenna fundamentals for legacy mobile applications and beyond*. Springer
2. Suvarna K, Murty NR, Vardhan DV (2019) A miniature rectangular patch antenna using defected ground structure for WLAN applications. *Progress in Electromagnetics Research C* 95:131–140. <https://doi.org/10.2528/PIERC19061602>
3. Sabban A (2020) *Wearable systems and antennas technologies for 5G, IOT and medical systems*. CRC Press
4. Tiehong T, Zheng Z (2003) Applications of planar inverted-F antenna for Bluetooth. In: *International conference on communication technology proceedings, 2003. ICCT 2003*. <https://doi.org/10.1109/icct.2003.1209752>
5. Chang D-C (2017) Invited talk 2: various kinds of antenna and their applications. In: *2017 international symposium on antennas and propagation (ISAP)*. <https://doi.org/10.1109/isap.2017.8229076>
6. Sim D-U, Choi J-I (2005) A compact wideband modified planar inverted F antenna (PIFA) for 2.4/5 GHz WLAN applications. In: *VTC-2005-Fall. 2005 IEEE 62nd vehicular technology conference*. <https://doi.org/10.1109/vetecf.2005.1558137>
7. Garg R, Bhartia P, Bahl IJ, Ittipiboon A (2001) *Microstrip antenna design handbook*. Artech House
8. Yao X, Tong L, Cui Y (2020) A planar inverted-F antenna-based chemical sensor for ion concentrations. *Sensors and Actuators Reports* 2(1):100006. <https://doi.org/10.1016/j.snr.2020.100006>
9. Sharawi MS (2014) *Printed MIMO antenna engineering*. Artech House
10. Kaabi AHA, Al Kaabi AH (2017) PIFA antenna design for 4G wireless communications. In: *2017 2nd international conferences on information technology, information systems and electrical engineering (ICITISEE)*. <https://doi.org/10.1109/icitisee.2017.8285492>
11. Bhavikatti SK et al (2021) Investigating the antioxidant and cytocompatibility of *Mimusops elengi* Linn extract over human gingival fibroblast cells. *Int J Environ Res Public Health* 18(13). <https://doi.org/10.3390/ijerph18137162>

12. Karobari MI et al (2021) An in vitro stereomicroscopic evaluation of bioactivity between Neo MTA Plus, Pro Root MTA, BIODENTINE & Glass Ionomer Cement using dye penetration method. *Materials* 14(12). <https://doi.org/10.3390/ma14123159>
13. Shanmugam V et al (2021) Circular economy in biocomposite development: state-of-the-art, challenges and emerging trends. *Composites Part C: Open Access* 5:100138
14. Sawant K et al (2021) Dentinal microcracks after root canal instrumentation using instruments manufactured with different NiTi alloys and the SAF system: a systematic review. *NATO Adv Sci Inst Ser E Appl Sci* 11(11):4984
15. Muthukrishnan L (2021) Nanotechnology for cleaner leather production: a review. *Environ Chem Lett* 19(3):2527–2549
16. Preethi KA, Auxilia Preethi K, Lakshmanan G, Sekar D (2021) Antagomir technology in the treatment of different types of cancer. *Epigenomics* 13(7):481–484. <https://doi.org/10.2217/epi-2020-0439>
17. Karthigadevi G et al (2021) Chemico-nanotreatment methods for the removal of persistent organic pollutants and xenobiotics in water—a review. *Bioresour Technol* 324:124678
18. Bhanu Teja N, Devarajan Y, Mishra R, Sivasaravanan S, Thanikaivel Murugan D (2021) Detailed analysis on sterculia foetida kernel oil as renewable fuel in compression ignition engine. *Biomass Conversion and Biorefinery*. <https://doi.org/10.1007/s13399-021-01328-w>
19. Veerasimman A et al (2021) Thermal properties of natural fiber sisal based hybrid composites—a brief review. *J Nat Fibers* 1–11
20. Baskar M, Renuka Devi R, Ramkumar J, Kalyanasundaram P, Suchithra M, Amutha B (2021) Region centric minutiae propagation measure orient forgery detection with finger print analysis in health care systems. *Neural Process Letters*. <https://doi.org/10.1007/s11063-020-10407-4>
21. Kang T-W, Yeh S-H (2016) Multi-band LTE-advanced antenna design for mobile phone. In: 2016 IEEE 5th Asia-Pacific conference on antennas and propagation (APCAP). <https://doi.org/10.1109/apcap.2016.7843163>
22. Belhadef Y, Boukli Hacene N (2011) PIFAS antennas design for mobile communications. In: International workshop on systems, signal processing and their applications, WOSSPA. <https://doi.org/10.1109/wosspa.2011.5931429>
23. Loutridis A, Yang K, John M, Ammann MJ (2016) A dual band 450LTE/GSM900 PIFA for portable devices. In: 2016 Loughborough antennas & propagation conference (LAPC). <https://doi.org/10.1109/lapc.2016.7807613>
24. Nashaat DM, Elsadek HA, Ghali H (2005) Single feed compact quad-band PIFA antenna for wireless communication applications. *IEEE Trans Antennas Propag* 53(8):2631–2635. <https://doi.org/10.1109/tap.2005.851872>
25. Liu M, Zhao B (2014) Dual-feed small-size penta-band PIFA for LTE/WWAN mobile handset applications. *International Journal of Antennas and Propagation* 2014:1–7. <https://doi.org/10.1155/2014/104614>

Author Index

A

Abul Kalam, Md., 73
Afzal, Ahmad Bin, 317
Ahuja, Stuti, 599
Ajmeri, Moina, 367
Alrajhi, Hasan, 317
Amosedinakaran, S., 15, 201
Azim, Ahmar, 317

B

Baburao, Pasupulati, 101
Balachander, Bhuvanewari, 649
Balamurugan, G., 101
Balasubrahmanyam, M., 633
Bharti, Prem Prakash, 219
Bhogal, Aman Singh, 439
Butchi Raju, Y., 101

C

Cao, David, 425

D

Daigavane, Prema, 27, 177
Ding, Kun, 163
Dubey, Smriti, 489
Dwivedi, Divyanshi, 49

E

Eswaramoorthy, Muthusamy, 219

F

Feng, Changsen, 119

Francis, Febin, 89

G

Gautam, Ruchi, 287
Gedam, Shubham, 27
Ghosh, Ankush, 613
Goyal, Devendra, 219
Guhe, Tushar, 27, 333
Gupta, Prakash Chandra, 37
Gupta, Ridhi, 507

H

Halloum, M.-Ramez, 89
Hameed, Salman, 317
Harinkhede, Somesh, 151
Hiware, Rutuja, 137
Hooda, Manish Kumar, 457
Hussain, Md Tahmid, 317

J

Jawale, Pratiksha Prakash, 403
Jiang, Wei, 1

K

Kalpana, G., 649
Kanojia, Sarika S., 253
Kapoor, Kartik, 219
Kattepogu, Priyanka, 559
Kaur, Harpreet, 457
Khalid, Mohd Rizwan, 317
Khandelwal, Anant, 467
Khan, Rashid Ahmed, 317

Kim, Ryan, 425
 Kiran Kumar Reddy, B., 89
 Kisson, Nikhil, 343
 Kolhatkar, Chinmay, 381
 Kumar, Deepak, 583
 Kumar, Harish, 393
 Kumari, Nisha, 61
 Kumar Naidu, Hari, 177
 Kumar, Prakash, 73
 Kumar, Priyanshu, 367

L

Lin, Zheming, 119
 Liu, Liyang, 1
 Li, Zihao, 119
 Lu, Steven, 425

M

Maddipatla, Jagadeepam, 425, 467
 Ma, Junwei, 1
 Mehra, Neelesh, 489
 Meng, Xuyao, 163
 Meshram, Pratiksha, 439
 Mishra, Ankit, 293
 Mishra, Suchismita, 415
 Mohan Sai Kumar, C. H., 633
 Mridha, Krishna, 613
 Murdan, Anshu Prakash, 273, 343

N

Nigam, Manoj Kumar, 293
 Nikhil, P. G., 219
 Nithin Reddy, G., 89, 535

O

Ohol, Shantipal S., 403

P

Pachkawade, Vaishnavi, 137
 Palanisamy, Rajakumar, 15
 Pal, Mayukha, 49
 Pandit, Vishram, 535
 Pasupulati, Baburao, 15, 201
 Patil, Megha, 357
 Patil, Sonali, 599
 Paul, Kaushik, 61, 73
 Pradeep Kumar, Ch. S. K. B., 101

R

Rai, Satayush, 545
 Rajakumar, P., 201
 Rajderkar, V. P., 333
 Ramasamy, Jagadeshwaran, 573
 Ramaswamy, Parthasarathy, 535
 Ramez Halloum, M., 535
 Rao, Pranil Ashok, 545
 Reddy, D. Maneesh, 49
 Redrouthu, Sathvik, 425, 467

S

Sabnis, Divya, 357
 Sagar, Anil Kumar, 415
 Sakthidasan, A., 201
 Samiappan, Dhandapani, 573
 Sankardoss, V., 189
 Sarker, Tasnim, 613
 Sathyaraj, J., 189
 Saxena, Richa, 393
 Senthil Kumar, M., 201
 Shah, Aagam, 253
 Shaikh, Aman, 545
 Shakya, Devendra Kumar, 489
 Sharma, Abhilasha, 591
 Sharma, Dileep Kumar, 591
 Sharma, Kushagra, 583
 Shaw, Rabindra Nath, 613
 Shelke, Maya, 545
 Shinde, Ganesh, 545
 Shinde, Yashraj Mohan, 305
 Shukla, Madhu, 613
 Sibbala, Bhargava Reddy, 15
 Singh, Lakhwinder, 241
 Singh, Manjeet, 287
 Singh, Manpreet, 241
 Singh, Piyush Pratap, 37
 Sinha, Mausmi, 439
 Sinha, Pampa, 61
 Sivakumar, S., 15
 Subba Reddy, B., 89, 535
 Sugesh, M. S., 517
 Sulakhe, Prakansha, 333
 Sundar, R., 201
 Sun, Yalu, 163

T

Talke, Sagar Appaso, 305
 Tamboli, Arshad Mansur, 305
 Thakurai, Altaf Hussain, 317
 Tripathi, Atul S. M., 559

UUmanand, L., [535](#)**V**Vadde, Pranav, [467](#)Vaidya, Prajakta, [151](#)Vairavel, G., [517](#)Valarmathi, R. S., [633](#)Vashist, Apurva, [415](#)Virkhare, Roshni M., [177](#)**W**Wagle, Kranti, [381](#)Wang, Kun, [1](#)Wang, Xiaohui, [1](#)Wankhede, Sagar, [357](#)Wen, Fushuan, [1](#), [119](#), [163](#)**X**Xue, Honglin, [1](#)**Y**Yadav, Manish Varun, [507](#)Yadav, Swati Varun, [507](#)Yang, Changhai, [163](#)Yemula, Pradeep Kumar, [49](#)Yeole, Dipak Suresh, [305](#)**Z**Zaki, Zoya, [89](#), [535](#)Zaman, Rawnoquz, [613](#)Zhan, Bochun, [119](#)